STATE OF THE ART ANALYSIS

Deliverable D1.1

Circulation:	PU: Public
Lead partner:	FOMI
Contributing partners:	SINTEF, Fraunhofer, CNR-IMATI-GE, UCL, TUDelft
Authors:	Nguyen Thai Binh, Jan Böhm, André Brodtkorb, Tor Dokken, István Fekete, Roberto Giachetta, Michel Krämer, Dániel Kristóf (editor), István László, Roderik Lindenbergh, Angéla Olasz, Giuseppe Patané, Michela Spagnuolo
Quality controller:	Ewald Quak
Version:	1.1
Date:	11.02.2013

Consisting of

SINTEF	STIFTELSEN SINTEF, Department of Applied Mathematics, Oslo, Norway
Fraunhofer	Fraunhofer Institute for Computer Graphics Research (IGD), Darmstadt, Germany
CNR-IMATI-GE	Institute for Applied Mathematics and Information Technologies of the National Research Council (CNR-IMATI), Genova, Italy
MOSS	M.O.S.S. Computer Grafik Systeme GmbH (MOSS), Munich, Germany
HRW	HR Wallingford Ltd (HRW), Wallingford, UK
FOMI	Hungarian National Mapping and Cadastral Agency (FOMI), Institute of Geodesy, Cartography and Remote Sensing, Budapest, Hungary
UCL	University College London (UCL), Research centre for Photogrammetry, 3D Imaging and Metrology, London, UK
TUDelft	Delft University of Technology (TUDelft), Department of Earth and Climate Sciences & Man-Machine Interaction Group, Delft, The Netherlands
IGN	Institut National de l'Information Géographique et Forestière (IGN), Paris, France
UBO	Université de Bretagne Occidentale (UBO), European Institute for Marine Studies, Brest, France
Ifremer	L'Institut Francais de Recherche pour l´Exploitation de la Mer (Ifremer), Brest, France
Liguria	Regione Liguria, Genova, Italy

DOCUMENT HISTORY

Version[1]	Issue Date	Stage	Content and Changes
1.0	15.01.2013		First release submitted to the Project Officer
1.1	11.02.2013		Included IQmulus logo

[1] Integers correspond to submitted versions

EXECUTIVE SUMMARY

This deliverable is the documentation of the "State of the Art" at the beginning of the project. As such, it is an important reference, a "landmark" in relation to which the future progress of IQmulus can be measured. It also identifies high-level requirements designating the initial directions taken.

For an efficient and meaningful documentation of the above, a dual approach was chosen, corresponding to the two different major perspectives and hence two possible definitions of "State of the Art" to be definitely taken into account.

First, scientific and development aspects corresponding to the understanding of "State of the Art" as the landscape of currently available cutting-edge scientific and technical possibilities and trends are elaborated. This viewpoint represents a "top-down" approach to define the main directions for the project, and is mainly influenced by project-internal scientists and developers. It contains an updated and enhanced version of the scientific and development aspects already included in the Description of Work. It was considered important, also for further circulation of relevant information beyond the consortium, to have a framework for orientation without referring back to the DoW all the time. Main ideas were reorganised according to the work package and task structure of the project, and revised by the leaders of the respective scientific and development Work Packages and tasks, principally WPs 2 (Infrastructure Design), 3 (Heterogeneous Data Integration Platform), 4 (Processing Services) and 5 (Interactive Visual Decision Support). Structurally, this is the content of "Part ONE". Each scientific-technical topic is assessed through a structure fitting the distribution of Work Packages and tasks, describing "Baseline", "Issues", "Development directions" and (wherever already possible) "Indicators".

Second, user aspects corresponding to "the other" viewpoint are analysed. This includes the landscape of current solutions, data, problems and expectations of the (possible) users of the system that IQmulus aims at developing. This viewpoint is much more practical and represents the "bottom-up" approach to determine what are the main development areas where users are waiting for new solutions to be developed. It is mainly influenced by project-internal users at this stage, with externals being involved progressively during the project evolution via numerous different activities in the user-related Work Packages 1 (Requirements), 7 (Assessment and Evaluation) and 8 (Dissemination and Outreach). This is the content of "Part TWO", which is subdivided into application areas determined by the analysis of an initial set of high-level user requirements submitted by the core group of consortium-internal IQmulus users. Each application area is assessed through (high-level) "Requirements", "Constraints", "Relations" (describing the logical links to Part ONE) and relevant "Sample data" (already available at the IQmulus Data Repository server).

Additional chapters provide a more detailed introduction to the objectives of the project and the document (Chapter 1), and describe the methodologies applied (Chapter 2). The two major parts are logically linked (besides the above-mentioned "Relations" subchapters in Part TWO) through the global "Conclusions" (Chapter 13). All user stories gathered this far and included in the analysis are contained in Annex 1, and the metadata table describing all sample datasets already uploaded to the IQmulus Data Repository can be found in Annex 2.

This document is the starting point. The inevitable and necessary duality of "developer" and "user" aspects is reflected both in its structure and in its content, and initial logical links are established. One of the important outcomes of further project evolution will be a deeper mutual understanding of the two main areas through intensive dialogue, and a deepening and consolidation of requirements to be described in future Deliverables dealing with User Requirements (D1.2.1, D1.2.2, D1.2.3, D1.2.4).

TABLE OF CONTENTS

1 INTRODUCTION

1.1 OVERALL OBJECTIVES OF IQMULUS

The “big data” challenge addressed by IQmulus is well synthesized by the following quotation “*The computing times for a typical large-scale problem in scientific computing can today be solved a million times faster than twenty years ago. Here a factor of less than 1000 is due to the hardware development and a factor of more than 1000 is due to the algorithmic progress*”[2].

The main point here is that “big data” cannot be addressed solely by increasing the computing power of our systems: there is a need for increasing the smartness of our software in terms development of new “intelligent” computational solutions able to exploit all information carried by the data. IQmulus will work on this aspect of the big data challenges for geospatial data focusing on the following issues[3]:

- **Scalability:** given the size of datasets, and their pace of production, it is necessary to devise computational solutions which can exploit the benefits of a high-performance computing infrastructure, including enhancements in multiple aspects including architectures, interfaces, algorithms and processing. This also has strong implications on the speed of processing perceived by users.
- **Automation:** the ultimate goal of IQmulus is to provide a high degree of automation in its service provision; in practice, full automation is hard to achieve for all the services envisioned by the project. Global processes are often based on local conditions, and local processes often depend on global estimates. IQmulus will exploit context dependent and context free approaches to take advantage of both according to the particular cases;
- **Modeling/representation:** given the size of the datasets, it is important to organize the important semantic elements they contain and to provide an effective access data structure, with indexing mechanisms that allow for a quick retrieval of the relevant data to be fed to processing tasks; currently, indexing is mainly addressed at the level of geographical coordinates and a semantic access-key to datasets is necessary;
- **Quality:** in current solutions, being able to process large data sets quickly and in a fully automated manner implies accepting a loss of quality; quality has to become a mandatory documentation item both for data and processing services so that the users are fully supported in their workflows and may decide what solution to use according to the needs of the specific context.

These issues will be addressed taking into consideration that in the geospatial domain cloud-based software use and cloud-based data distribution are becoming more and more important. Both the concepts of public and private clouds are relevant to this project. In particular, the Public Cloud will affect the project in the distribution of data and in data as a service with respect to the related commercial exploitation strategies. The aim is that the number of stakeholders with access to the project results becomes as large as possible,

[2] Ulrich Trottenberg and Han La Poutre. Simulation and Modeling for Research and Industry, ERCIM NEWS n. 81, April
2010

[3] Przemyslaw Musialski, Peter Wonka, Daniel G. Aliaga, Michael Wimmer, Luc van Gool, Werner Purgathofer A Survey of Urban Reconstruction. In *EUROGRAPHICS 2012 State of the Art Reports*, pages 1-28. May 2012.

providing the data for further usage to as broad a community as possible. Opportunities to build services on top of the data will be considered and the related exploitation strategies will be more connected to the Private Cloud concept, where the software use should be rather restricted.

From a technical point of view the two cloud concepts do not result in different requirements for the project. The focus is on the use of the cloud concept to the application domain, adjusting the geo-data processing technology to be developed to run in cloud environments, e.g. by supporting cloud monitoring environments, service awaking and down-shutting, cloud service metadata, and service execution standards. Whether the execution of the processing is performed in public or private cloud environments is then rather a question of the cloud setup and configuration and not of the project's processing techniques.

1.2 AIM AND SCOPE OF THIS DOCUMENT

The start of development planning requires the definition of basic high-level requirements based on the state of the art, identification of key issues to be resolved, and determination of the progress to be acquired over the state of the art.

Two aspects representing different viewpoints are to be taken into account:

- **Scientific and development aspects** corresponding to the understanding of "state of the art" as the landscape of currently available cutting-edge scientific and technical possibilities and trends. This viewpoint represents a "top-down" approach to define the main directions for the project. It is mainly influenced by project-internal scientists and developers.
- **User aspects** corresponding to another view on "state of the art": the landscape of current solutions, data, problems and expectations of the (possible) users of the system that IQmulus aims at developing. This viewpoint is much more practical and represents the "bottom-up" approach to determine what are the main development areas where users are eagerly waiting for new solutions to be developed. It is mainly influenced by users – project-internal at the beginning, with externals being involved progressively during the project evolution.

The current deliverable is the documentation of the "state of the art" including both of the above-mentioned aspects, as seen at the beginning of the project.

It contains an updated version of the scientific and development aspects already included in the Description of Work.

It also reports on the preliminary survey of the requirements that we collected during the early phase of user engagement, which has started and will continue in the next months. Mainly consortium-internal testbed users have been involved at this stage to make them aware of the foreseen IQmulus results and to start comparing their expectations and desiderata with respect to the state-of-art solutions for heterogeneous data handling and processing. This analysis phase is particularly important for the IQmulus platform development and, as such, it will be subject of further activities in the next month. In particular, we mention: the update of this deliverable after the completion of the planned user workshops (see D1.2.1 and its revisions from D1.2.2 to D1.2.4); the survey activities focusing on algorithmic aspects of processing services (see D4.1.1).

Given the early stage of user engagement, the development directions over the state-of-the-art have been primarily discussed by the project partners considering areas where IQmulus is

expected to generate significant progress, as delineated already in the Description of Work. This was decided during the project kick-off, during which testbed users had the opportunity to discuss with the partners research outcomes, demos of relevant software and research challenges in areas relevant to IQmulus.

2 METHODOLOGY

Scientific and development aspects corresponding to the understanding of “State of the Art” as the landscape of currently available cutting-edge scientific and technical possibilities and trends were to be elaborated. This corresponds to a “top-down” approach to define the main directions for the project, and was mainly influenced by project-internal scientists and developers.

As for the user side representing the “bottom-up” approach, the requirements gathering methodology in IQmulus is partly based on what is known as AGILE methodologies for software development.[4] The common process for gathering requirements for software systems is to gather as many requirements as possible at the beginning of the project, assuming that the more requirements there are, the better the developed software will be. However, “*extensive upfront requirements gathering and documentation can kill a project in many ways*” [5]. Capturing requirements in a few hundred pages documents without a clear prioritization of requirements, references to the tasks and use cases where the requirements are derived from, and explanations of ambiguous terms might result in a situation where it is almost impossible for the developers to start developing the software. Therefore, the AGILE methodologies, focusing on user tasks at the beginning and refining the high-level requirements in a communication process with the users and stakeholders afterwards seemed to fit quite well to the project. From our point of view, capturing all requirements at the beginning of a project is impossible. Requirements evolve during project runtime, especially when first prototypes are available and given to users.

By using Agile methodologies and following the tasks specified for the project’s WP1 (requirements), our first phase was to gather user requests in a form that we could easily transform later on into functional requirements based on the widely known VOLERE template. This initial work has been successfully carried out resulting in the user stories listed in Annex 1.

2.1 SCIENTIFIC AND DEVELOPMENT ASPECTS

Elaboration of scientific and development aspects was carried out by restructuring, updating and enhancing the scientific and development aspects already included in the Description of Work. Intensive discussions led to a substantial refinement and improvement of the raw concepts presented there. Basically, the main ideas were taken from the DoW, reorganised according to the work package and task structure of the project, and revised by the leaders of the respective scientific and development Work Packages and tasks, principally WPs 2 (Infrastructure Design), 3 (Heterogeneous Data Integration Platform), 4 (Processing Services) and 5 (Interactive Visual Decision Support).. Elaboration of each topic follows the structure below:

[4]http://agilemanifesto.org/

[5] Cohn, Mike: User Stories Applied: For Agile Software Development. Addison-Wesley, 2004.

- **Baseline:** The landscape of currently available cutting-edge scientific and technical possibilities and trends.
- **Issues:** Problems, gaps and caveats of the current solutions.
- **Development directions:** Initial directions taken by IQmulus to progress beyond the state of the art.
- **Indicators:** A list of indicators to make the progress measurable over the project lifetime (where already possible).

2.2 USER ASPECTS

User roles

It is a common mistake in the process of capturing the functional requirements for a software system to lose sight of the type of user that issued a requirement. Often, the assumption is that there is a single type of user of a system and all requirement specifications (let them be IEEE-styled requirements a la “The SYSTEM should…” or user stories) are written for this single type of user, sometimes referred to as *“the user”*. However, in most cases, a software system has multiple types of users with different experiences, backgrounds and goals while using the software. Delivering requirement specifications without a clear reference to the type of user that issued the requirement leads to problems in grouping and especially prioritizing them. Therefore, it is of huge importance to clearly identify the types of users (User Roles) and to unambiguously link them to the requirement specifications.

To define possible User Roles, a preliminary classification matrix was agreed on, in which the vertical structure corresponds to the *application area* or *work field* (e.g. civil protection, spatial planning, etc.), while the vertical structure (*user level*) corresponds to the *level of decision making* of the user (e.g. field agent, GIS expert, decision maker, etc.). User roles may also include technical personnel of the IQmulus system management and IT staff. The user roles are placed in the above-mentioned two dimensional system based on user work field and user level. Work fields and user levels are determined based on the initial stakeholders and expanded as new stakeholders are introduced. In the later phases of the project (especially during User Requirement Survey), user roles are identified based on the stakeholders invited to the workshops and the gathered user stories. All work fields will have named local stakeholders that are gathered and who will be part of the evaluation and assessment team in WP7. Thus, our User Inventory will be more defined and concrete when Task 1.4 (User Requirement Survey) adds more “raw material” for this.

In this initial phase of the project, however, a complete classification or inventory of possible user roles cannot be obtained but is not even necessary, as our goal is to get an overall knowledge on potential roles. Information is therefore being collected from the user stories by extracting and analysing relevant roles and populating the inventory based on this. We have divided user roles into basic <user levels> and <work fields> like:

- <GIS Expert> at a <Civil Protection Authority>,

- <Decision maker> at a <Territorial Planning Office>.

Also, user roles must be ranked for their importance between 1 and 5 (with 1 the most important role, 5 the least) based on

- the IQmulus project goals and aims defined in Part 1,
- the number and importance of user stories gathered with the specific role.

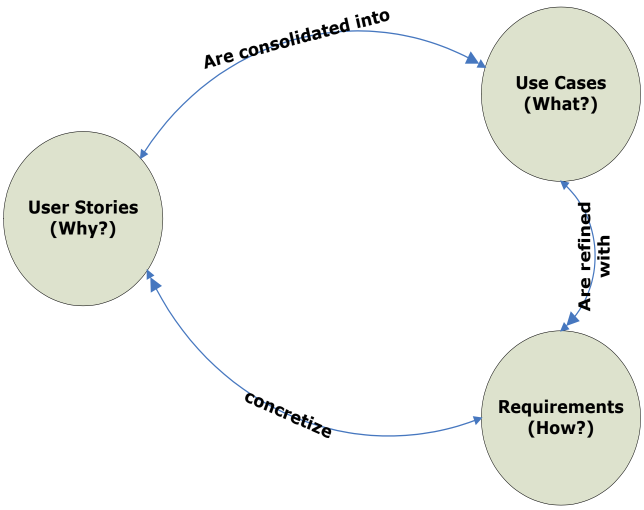
This latter step will be carried out in the next phases of the project, when a “critical mass” of user input is already available.

User Stories – Collection

„User Stories“ are known from the AGILE software methodologies. User stories serve as a basis for further steps to derive (both functional and some non-functional) requirements in the form of Use Cases.

Functional requirements describe the system capabilities required by a user for solving a functional problem. The requirements are primarily defined by use cases. A use case describes a concrete, functionally self-contained sub-process. The entirety of the use cases then defines the system behaviour. The functional requirements are the central system for the development of specifications. They will be integrated into the overall system specification and concretized as required (V-Model). The steps for this are the following:

1. Identify the various user roles
2. Gather their user stories
3. Derive respective use cases *(this will be done in later phases)*



A user story can be defined as "a software system requirement formulated as one or more sentences in the everyday or business language of the user". They are short natural language sentences of the intended functionality of a software system, written in the language of the user or stakeholder of the system. User stories help in ensuring a common understanding between the different stakeholder groups (developers, end users). As described in the AGILE methodology[6], user stories are composed of three aspects:

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

A user story should follow the template:

As a <user>, I want <something> so that <benefit>.

Stories following that template usually contain some relevant information that might sometimes be missed, namely:

i. Which user role / actor issued the requirement (<user>) and

ii. What is the rationale behind the requirement (<benefit>).

[6] http://agilemanifesto.org/

An example for an actual User Story gathered within IQmulus:

As a <GIS expert> at a <civil protection authority>, I want to <create and visualize> a <3D city model> by the <automatic classification> of <LiDAR data> so that I can <support and accelerate> the disaster <decision making process> (e.g., flood modelling).

Note that in this case the user role is "GIS expert" (vertical structure) at a “civil protection authority” (horizontal structure). High level information related to data involved, processes and algorithms to be applied and expected outputs can be interpreted from the descriptive parts of the user story.

It is important to note here that there can be different levels of details for user stories. Although, in general, user stories should be kept small in terms of the functionality to implement, there can be generic user stories that capture whole use cases in a single sentence. Such stories are often called *epics*. For example, a generic user story or epic derived from the IQmulus use cases is *“As an IQmulus user I want to interactively explore data that are large”*. Obviously, this is a rather generic user story capturing a lot of use cases and therefore, a huge set of details is missing such as what are the search criteria (thematic, spatial or temporal or all of them), is it possible to store the queries, etc. Such details can be captured in additional low-level detailed user stories derived from the epic, and will be part of the transformation of User Stories into valid Use Cases and Requirements.

In this first phase of the project, project-internal users were targeted. The basic idea behind this approach was to put together initial material representing the “user side”, already present within the IQmulus Consortium. Internal Users (i.e., consortium members participating in IQmulus trials, and also working with relevant data and applications during their everyday activities) have been involved and they provided User Stories describing their views on the system functionality, and provided sample datasets to have some “real” data available for the project already at this early stage. The User Stories collected in this phase can be found in Annex 1, while Annex 2 contains the metadata table describing what is already available in the IQmulus Data Repository.

User Stories - Analysis

The user stories collected in this first phase have been analysed by taking into account the following aspects:

- Available and potential input data
- Application areas
- Overall goals

Available and potential input data have been gathered by analyzing user stories. In most of the cases terrain model, remote sensing and vector spatial database inputs are dominating. Sample data are available on the Fraunhofer SFTP server. This will give an early opportunity for WP2, WP3 and WP4 members to inspect the quality and format of input data.

Application areas have been determined by analyzing the user stories. Each application area corresponds to multiple user stories. The number of relevant user stories can be used as an indicator to determine their standing. This means that we can see which areas are more important to the users. In Part two, sections have been distinguished and studied thoroughly according to application areas.

Overall goals are abstract high-level definitions gathered from user stories for further requirement analysis. This could give the basic skeleton for functional and non-functional requirement analysis.

PART ONE: SCIENTIFIC AND DEVELOPMENT ASPECTS

3 STATE OF THE ART IN ARCHITECTURE DESIGN

3.1 SPATIAL DATA PROCESSING ARCHITECTURES AND INTERFACES

Processing is the core of the IQmulus project and will cover different analysis, fusion, integration, interpolation or extrapolation, merging or correlation tasks. So far, these different tasks have been executed on different architectures and platforms, based on the data topology (unstructured point clouds, regular grids, irregular grids-triangulation, contours …) and on the task itself.

Baseline

Distributed processing of spatial data is – after 2D data provision ("download" and "view" services) has now been handled sufficiently – a topic actively discussed in the community, as indicated by the recent establishment of Geoprocessing conferences[7] and workshops. Spatial data processing is still primarily done on single machines using desktop or server software such as ESRI ArcGIS[8] or Intergraph GeoMedia[9] in batch mode, creating long-running jobs. These data processing applications are now being grid- or cloud-enabled. Examples for this include ESRI ArcGIS Server[10], Safe's FME[11], the open source geoprocessing framework 52° North[12] WPS, and novaFACTORY. While this approach works reasonably well for data storage and provision tasks, the situation is more complex with processing: the algorithms used internally still remain the same and are only deployed in a different manner, reducing the potential gain from distributed or manycore environments in many situations. With its Earth Engine (EE)[13], Google has opted for a different approach based on a reimplementation of algorithms on the basis of their MapReduce[14] architecture. While this works tremendously well for problems that can be transposed well to this architecture, it obviously excludes other problems, such as when full dataset understanding is required, as it is the case in spatial join operations, because of differences in semantics, and also because of differences in locational accuracy and resolution, e.g., data resulting from exploiting aerial surveys and from mobile vehicle mapping. Furthermore, EE requires data to be available in Google's infrastructure and is currently limited to processing of raster data.

Issues

A hurdle to fast migration to distributed environments and service oriented architectures is the fact that the geospatial community has developed their own set of standards, such as the Web Feature Service (WFS), Web Map Service (WMS) and Web Processing Service (WPS). Commonly required functionality such as security and authorization, workflow management and service/deployment control thus require additional implementation effort, and for many

[7] Example: http://www.iaria.org/conferences2012/GEOProcessing12.html (Fourth GeoProcessing conference)
[8] Website: http://www.esri.com – Last Visited: 10.01.2012
[9] Website: http://www.intergraph.com/ - Last Visited: 10.01.2012
[10] Website: http://www.esri.com/software/arcgis/arcgisserver/ - Last Visited: 10.01.2012
[11] Website: http://www.safe.com/ - Last Visited: 10.01.2012
[12] Website: http://52north.org/ - Last Visited: 10.01.2012
[13] Website: http://earthengine.google.org – Last Visited: 10.01.1012
[14] Jeffrey Dean and Sanjay Ghemawat. 2008. MapReduce: simplified data processing on large clusters. *Commun. ACM* 51, 1 (January 2008), 107-113. DOI=10.1145/1327452.1327492 http://doi.acm.org/10.1145/1327452.1327492

of these aspects there are no compatible standards to the aforementioned ones. Also, there is less tooling support available – as an example, while it is trivial to generate a SOAP- or JSON-based web-service, there are no similar tools available for generating OGC Web Processing Services and clients.

Development Directions

In IQmulus, progress beyond the state-of-the-art in architectures for spatial data processing will be made in several areas:
- Intelligent Execution Architecture:
Connect state-of-the-art workflow execution with distributed environment management software and allow the system to decide how to execute the processing for a request based on a service and algorithm annotation model.

- Coping with heterogeneous input sources:
Ability to cope with heterogeneous input sources, and to take advantage of these in a way that best exploit their characteristics by providing an architecture that is not optimized only for one data type.

- Scalability and Performance:
Ability to scale up to huge data sets and wide areas by implementing a novel architecture based on functional (stateless) basic (performing atomic and generic steps) processing services, built on established W3C/OMG/OASIS/OGC standards that have proven their usefulness in other domains and on the corresponding tools.

- Semantic enrichment of data and combined reasoning:
The architecture will be designed to take into account processing semantics and the semantic enrichment of input data in general. For this purpose, an annotation model has to be developed that can be used with the heterogeneous topologies of input and output data.

- Provenance and lineage information:
Quality information and impacts of each processing steps have to be documented. This provides the ability to perform a quality assessment of resulting data, both qualitatively and quantitatively, and modelling of this quality assessment so as to be able to make use of it in further processing. Related to this is how to incorporate methods of interpolation (like Kriging) and extrapolation into the workflow, in order to match different attributes sampled at different spatial and/or temporal resolution.

- Iterative processing approach:
Such an approach allows coping with:

- Situations which can be characterised by the availability of pre-existing data which already brings in some semantics, which in turn is useful to best perform a merge or a characterization;
- Combined signal and data processing to perform a first level of extraction which helps further processing.

Big data is an essential aspect within the project as the consortium aims to identify, integrate and evaluate the most suitable and efficient existing solutions for its purposes. Consequently this aspect will strongly influence the IQmulus architecture decisions. The storage layer in the architecture will be addressed in WP2 and be reflected in the deliverable D2.1, with impact on D2.3.

3.2 DOMAIN-SPECIFIC LANGUAGES (DSLS)

Developing for distributed processing environments brings a set of difficult problems with it. Especially in programming for distributed environments, a non-obvious mistake in the implementation of an algorithm can decrease its performance massively by requiring waiting and synchronisation times. This often means that experts for spatial data also need to become experts for parallel processing problems.

Baseline

There are several different approaches being employed to the definition of spatial processing workflows, including procedural and object-oriented programming, declarative approaches[15] and rule-based approaches.
Recent work has shown that rule-based systems can be used to solve problems from the geospatial area such as data harmonization and data integration as well as application-specific querying and visualization[16]. Other applications such as data validation and quality evaluation are also possible[17]. They can also be used to automatize processing—business rule management systems (BRMS) are an example for that.

Issues

Current implementations of rule-based systems suffer from the problem that they are rather complex and generic. This makes them hardly usable for non-IT personnel. The user must have a deep understanding of internal matters (i.e. data models, algorithms, etc.). Handling all these details keeps users from focusing on the actual problem.
However, the largest part of processing algorithms is still implemented using conventional programming languages such as C++ (e.g. GRASS) or Java (e.g. JTS/GeoTools), where—in part due to the lack of suitable frameworks—developers have only limited insulation from the complexity of multithreading, memory access and other issues.
Domain-specific languages (DSLs) can alleviate this issue since they are tailored to a very specific application domain. They use a vocabulary that is well-known to the user (domain expert) and they completely hide the implementation-specific details. Under the hood, rules can be used to map DSL language elements to parallel processes.
Krämer, Ludlow and Khan[18] summarize the state of the art in applying domain-specific languages to parallel computing as follows: "In recent times, domain-specific languages are also used to simplify the development of distributed programs and algorithms. For example, Lee et al. present the Delite Compiler Framework which allows users to describe parallel code execution using an abstract DSL.[19] The framework is able to use the abstract description to generate optimized code for several, heterogeneous platforms such as different operating systems, mobile devices or GPUs via CUDA or OpenCL. Their work aims for hiding the specifics of each target platform from the user. In order to develop their languages, Lee et al. use the so-called language virtualization technique.[20] Components of a high level language (in

---

[15] Reitz, T.: A Mismatch Description Language for Conceptual Schema Mapping and its Cartographic Representation. Proceedings of the 6th GIScience Conference, Zürich (2010).
[16] Thum, S. and Krämer, M.: Reducing Maintenance Complexity of User-centric Web Portrayal Services. Proceedings Web3D 2011: 16th International Conference on 3D Web Technology, ACM Press (2011), 165- 172.
[17] van Oosterom, P.: Research and development in geo-information generalisation and multiple representation. Computers, Environment and Urban Systems 33, 5 (2009), 303-310.
[18] Krämer, M., Ludlow, D., & Khan, Z.: Domain-specific rule languages for urban policy modelling (to be submmitted). European conference on modelling and simulation (2013)
[19] Lee, H., Brown, K., Sujeeth, A., & al.: Domain-specific Languages for Heterogeneous Parallel Computing. IEEE Micro 31, vol. 5 (2011), 42-53.
[20] Chafi, H., & al.: Language Virtualization for Heterogeneous Parallel Computing. Proc of ACM Int'l Conf. Object Oriented Programming Systems Languages and Applications. ACM Press (2010), 835-847

this case Scala) such as lexer, parser and type checker will be reused. Based on that, components for optimization and code generation will be implemented. Sujeeth et al. use the Delite Compiler Framework and this technique to implement the OptiML language which is a DSL for distributed machine learning.[21]”

Development Directions

We propose to develop a set of domain-specific languages that make the definition of several types of spatial data processing and analysis tasks simpler. These domain-specific languages will have several properties:

- They hide the complexity of the underlying execution architecture by ensuring that the developer does not have to care about distribution and parallelization;
- They enable developers to annotate an algorithm with semantic information and with crossparameter constraints, i.e. with relationship information between input and output of the algorithm such as data dimension, data topology, reference system or other data properties;
- They can be compiled to run on parallel processing environments, such as GPGPUs, Cloud stacks, or MapReduce implementations such as Hadoop. The compiled algorithms can be packaged as declarative services to run on the respective environment they have been compiled for.

To be able to effectively work with the DSL, we propose to develop an Integrated Development Environment, together with the required compiler infrastructure. This IDE will include features such as data-driven testing and debugging, which are essential for effective development work.

4 STATE OF THE ART IN SPATIAL DATA PROCESSING

An important step is the identification of the requirements for the development of the functional and domain processing services in order to (i) maximize the use of heterogeneous and large data sets, (ii) support data quality evaluation, and (iii) provide support for the analysis of quickly changing environmental conditions. The requirements’ identification is also useful for the development of new algorithms for data processing, their comparison with respect to previous work, and the development of those parts of the infrastructure that will include and integrate the processing tools. For the development of processing services, “data heterogeneity” mainly refers to data dimensionality and representation. Another important issue is the access to data and tools through web interfaces, which are tailored to different users such as data providers, GIS experts, end users or decision makers, IT administrators or system operators, and researchers.

To measure the qualitative and quantitative improvements of processing tools with respect to existing methods, we can classify processing algorithms into two main categories: (i) basic processing algorithms for spatio-temporal data fusion and multivariate surface generation; (ii) analysis algorithms for feature extraction, classification, correlation, change detection and dynamics. Both categories will also take into account uncertainty processing, which will be used to represent and analyze how uncertain or partial data affect the processing pipeline, and knowledge management, which includes both data enrichment with metadata and semantic annotation.

[21] Sujeeth, A. K., & al.: OptiML: An Implicitly Parallel Domain-Specific Language for Machine Learning. Proc of the 28th Int’l Conf. Machine Learning (ICML 11). ACM Press (2011), 609-616

Based on the end-user stories and through an analysis of the state-of-the-art of processing tools for geographic data, the requirements analysis will allow us to identify the main guidelines for developing the processing services. The first step includes (i) the set-up of an inventory of the main data type and processing tools, which will provide an initial analysis of previous work, the main input for the design and implementation of the processing services; (ii) the survey of the main algorithms for data processing, integration, and analysis with respect to computational complexity and robustness. The intermediate step includes the development of specific methods that address the four main aspects of the processing pipeline; i.e., the spatio-temporal data fusion; the multivariate surface generation; the feature extraction, classification, and correlation; the detection of data changes in time-varying data. The final step is to define guidelines for enriching data with metadata about processing history, data quality, quality propagation, result reliability at all stages of the processing pipeline; and to propose a validation framework for the developed algorithms.

The aforementioned steps will allow us to produce a prioritized list of processing services that can be adapted, developed, and integrated. Finally, the quality, scalability, and degree of user interaction of each algorithm will be analyzed with respect to the application fields and state-of-the-art methods.

4.1 HETEROGENEOUS DATA INTEGRATION

Data fusion aims at obtaining information of greater quality by combining information from multiple sources, compared to using information from only one source. In the IQmulus project, we are looking at fusion of heterogeneous spatial data sets, which are heterogeneous in one or many of the following aspects:

- Spatial and Temporal Coverage
- Resolution, completeness, density, positional accuracy (internal/external)
- Method of acquisition and method of processing
- Topological correctness
- Structure and topology
- Dimensionality
- File Formats and Conceptual Schemas

Coregistration plays an important role as part of the fusion process when coping with the heterogeneity issues here mentioned here.

Baseline

While there are already several methods proposed for fusion of heterogeneous spatial data, these are often focused at data that is heterogeneous in only one or two of the aforementioned aspects. Furthermore they usually do not explicitly handle metadata of the used input dataset like quality of dataset, acquisition methods or modifications. This limits the use of complex data fusion processes to the initial input datasets in terms of data model and often data format too. Input data changes often lead to extensions of the data fusion workflows to transform the new input dataset into the structures of the initial input dataset even if information that a later process step can benefit from gets lost.

Example methods for spatio-temporal data fusion include notably least squares methodology like spatio-temporal Kriging and spatial Kalman filtering. Although such methods are optimal in a specific mathematical sense (variance is minimized) there applicability is certainly not straight-forward because of:

a) the risk of large computational efforts,
b) the challenge to incorporate thematic information, which is a particular aim of this project, and
c) the need to systematically incorporate a sound quality description.

Registration is the process of aligning a – in our case – topographic dataset in a common coordinate system. There are several techniques available for registration. On the other hand there are also several well-known problems associated to registration, (see for example EuroSDR[22], 2011). Registration of satellite and aerial images is a well assessed problem, as long as resolution discrepancy is not too wide. Registration of 3D vector data and other terrestrial data, or of terrestrial data and aerial data, however, is still a matter for further research.

In modern acquisition geolocation of the acquired dataset is mostly directly obtained using a combination of IMU (Inertial Measurement Unit) and GNSS (Global Navigation Satellite System). By combining GNSS and IMU with the, say, range data of the main sensor (like LIDAR or MBES), a directly geo-referenced dataset is obtained. For most acquisition systems a second registration step is needed to obtain a correct registration at least in the sense that data from overlapping ship or plane tracks matches correctly. For better results this second step may take advantage of in-situ data (geodetically measured features, previously captured data…).

Issues

Still, each topographic dataset may have some locational errors, in general a combination of unresolved systematic errors (affecting the accuracy) and random errors (described by the precision of the measurements or data set). When combining different topographic datasets, as in this project, it is therefore essential to take the geo-location errors into account in each processing step. Thus we propose to consider two alternative but not mutually exclusive approaches: a *stochastic approach* and a *geometrical* approach.
In the stochastic approach the error budget of each entry in the topographic dataset is explicitly described and incorporated in each processing step. By rigorous error propagation the mathematical framework of testing theory can be applied to decide with a given reliability whether a certain difference between two data points at the same location corresponds to a real change (difference is significant) or is due to the data errors[23]. Semantics may play a role here too, in that differences between characterized features which ought to be the same may indicate a registration error as opposed to an update to consider[24].
In the geometric approach first explicit matches (for example between two representations of the same tree, or two representations of the same building facade) are established before conclusions on possible changes are drawn. In this second, geometric approach the first stochastic approach can be integrated (For example, given the quality of the many individual points representing together a facade, how certain is the location of this facade). Advantages of such integration are that a) the redundancy of the data is exploited, and b) the computational efficiency is increased by considering derived models instead of individual points.

[22] http://www.eurosdr.net/publications/59.pdf
[23] Teunissen, P. J. G., 2000. Adjustment theory. Delft: Delft University Press.
[24] Champion, N., Boldo, D., Pierrot-Deseilligny, M. and Stamon, G., 2010. 2D building change detection from high resolution satellite imagery : A two-step hierarchical method based on 3D invariant primitives. Pattern Recognition Letters 31, pp. 1138–1147.

Development Directions

The research questions in the proposed fusion work are:

1. Should the points simply be merged together, leaving the individual original points intact?
2. If not, how should interpolation be performed?
3. How to best combine semantics based reasoning and geometrical merging?
4. How can non-geometric attributes be merged?
5. How can quality / accuracy be assessed?

Combining different topographic datasets enriched with semantic information and moreover acquired at different moments, obtained with different sensors at different spatial resolutions, introduces a new set of challenges. First, the semantic information pieces added to the different topographic may not directly match. For instance, a point may be classified `leaf` in one data set, `tree` in a second, and `garden` in a third dataset. Moreover, classifications may be completely different, like `grass` in one data set compared to `car` in another data set, either due to classification errors or due to real changes.
These new challenges call for new methods to cope with multiple enriched point cloud data sets. One approach is to explicitly store the classification method together with the topographic data set rather than the result of the classification. In addition it will be beneficial to introduce a fixed semantic, scale dependent hierarchy of objects. In both approaches mechanisms will be built in to incorporate both data and classification uncertainty.

4.2 SPATIO-TEMPORAL DATA FUSION

Data fusion aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application. The task will focus on processing methods aiming at creating a single data set containing information of the same type (univariate data set) out of different data sources (multi-source fusion). The processing will deal with data of different spatial and temporal resolution, with data of different quality, data that is more or less out of date, and with data describing different but often correlated attributes. Quality assessment of the resulting data (localized, at best at point level) will be provided both qualitatively and quantitatively. Various knowledge sources will be considered to perform fusion: sensor characteristics, characteristics of the measurement process and known characteristics of the entities measured. This task will propose new ways to incorporate methods of interpolation and extrapolation into the workflow, in order to match different attributes sampled at different spatial and/or temporal resolution. A key issue in this task is also the feasibility of the proposed algorithms to systematically handle huge data sets.

Baseline

Probably the best-known method for data fusion is so-called Kalman filtering. In this method the state of an attribute value of a point in spatial-temporal space is updated whenever new information becomes available. The contribution of the new information is weighted with respect to both the quality of the old information and the new information. Using Kalman filtering it is also possible to predict the state of the point in future (that is, some moment after the latest information was incorporated). Kalman filtering incorporates the quality of and correlation between data in terms of variance-covariance matrices using the technique of least-squares. Compare also Task 4.4 below.

Typically, Kalman filtering aims at estimating and predicting the state of an attribute value in a continuous setting. When dealing with spatial-temporal data often instantaneous large

changes occur, e.g., because of the influence of a high energy event (flooding, earthquake) or due to human intervention (chopping down a tree). A different set of methodologies exist to support such kind of instantaneous changes with a larger spatial support. For such changes both a threshold on the change in attribute value, and spatial correlation of the change is taken into account, see also Task 4.5.

Issues

For developing a successful set of methods aiming at spatial-temporal data fusion within the context of IQmulus several issues have to be taken into account. A first issue is the quality of individual measurements at the moment of measurement acquisition. This quality depends on the sensors used, on the ambient conditions, on the measurement geometry and on the properties of the measured object. In general, the better the quality of the input data is known, the smaller the error bandwidth of the resulting estimated parameters can be. Unfortunately, state-of-the-art quality descriptions are often only global and do not address influencing factors at an individual measurement basis. A second issue is the aging of the input measurements and therefore of the derived estimated parameters. A third issue is to take into account what measurements actually represent, compare Task 4.3, that is, adding semantic knowledge may result in better results at the cost of making the processing more complex. A final issue to be mentioned here is the feasibility of current and new methods to work in an automatic way on huge data sets.

Development Directions

Some methods already naturally incorporate the quality of the individual measurements, but for many methods it is not at all obvious how this should be done. One of the opportunities of this task is to directly cooperate together with the other tasks by considering how an explicit and localized quality assessment of the results can be obtained for each of the proposed processing methods. This means that in parallel to an often deterministic processing flow also a stochastic procedure is developed to propagate the quality of the input measurements through the processing chain into a quality description of the estimated output parameters.

Another development direction will aim at making the methodology computationally feasible, while still maintaining the information needed for a proper quality assessment. This goal will be aimed at in two different ways. First it will be investigated how neighbourhood searches can be localized as much as possible by incorporating only those points that have an effective influence on the result. We call this approach range limitation. It will be also investigated how to set up algorithms in a scalable way, by using only that part of the abundantly available measurements that is needed to obtain a certain performance goal. We call this approach *resolution limitation*. Together these two approaches should limit the computational efforts as much as possible (as described by, e.g., a pre-set quality target).

Indicators

Performance indicators

Quality:

1. Registration error after alignment, i.e., point distances in-between registered point clouds
2. Shape fitting (e.g., plane) for known geometries (façade, street)
3. Developed methods should be clear, intuitive, understandable and implementable for relative outsiders
4. Results of the fusion should be delivered together with a quality indicator

5. Quality indicators should be tested in situ by field validation
6. Quality indicators should be tested by cross validation

Scalability:

7. Processing time
8. Numbers of parameters to be set should be low
9. Computational and storage complexity
10. Temporal update should probably only require the new information + the last state (so that not all old data needs to be reprocessed)

Degree of human intervention foreseen:

11. aim for 0% manual intervention during processing, but expect manual checking (10% of overall effort)

4.3 FEATURE EXTRACTION, CLASSIFICATION AND CORRELATION

The processing here is targeted at a semantic enrichment of the data sets. By semantic enrichment we mean automatic extraction and annotation of high-level information by segmentation or classification processes. To improve feature identification and classification, correlations between different data sets will be explored systematically. Geometrical, statistical and combined reasoning will be explored to cope with various kinds of data and features. In particular, random forest classifiers will be studied for highly automated classification.

Baseline

Quantities extracted from sensor readings are used to classify regions according to different surface types. Such classifications are well-established in the remote sensing and photogrammetry community. Often schemes of land-usage are employed for classification. For LiDAR readings the LAS 1.2 format specifies a classification standard (6 land-use classes: Ground, Low Vegetation, High Vegetation, Medium Vegetation, Building, Water and further flags for processing[25]). Several classification algorithms are reported in the literature for Geo-data. Probably still the most widely used algorithm is the k-means algorithm. However, most recent research has provided us with tools that are more efficient for large datasets, such as random forest classifiers.

Issues

In current commercial practice feature extraction is still a process heavily dependent on manual interaction. In day-to-day practice classification proves to be a bottleneck in processing as it involves massive amounts of manual work in checking and re-classification. We aim for 0% manual intervention during processing, but expect manual checking (10% overall effort). Manual corrections give feedback to the training of classifiers, which improves the classification accuracy.

[25] http://asprs.org/a/society/committees/standards/asprs_las_format_v12.pdf, page 8

Development Directions

We propose a scheme that uses unsupervised classifiers for highly automated classification. For example random forest classification has shown promising results in early experimental studies of 94% accuracy[26]. They are a flexible and adaptive scheme for classification that has proven to perform well on huge data sets and thus are ideally suited for the mass data processing aimed at within this project.

Supervised classifiers are trained using pre-classified training data. This allows for flexible classification even outside standard schemes. Once regions are classified into rough categories geometric reasoning can be applied for individual feature recognition, e.g., building blocks for points classified as 'building'.

Indicators

Quality:

- Scientific indicators are false positives, false negatives, ROC curves.
- For segmentation, comparison to manual segmentations is standard. For, e.g., building extraction we can use hand-generated building models available from national mapping agencies, e.g., Ordnance Survey.
- User's accuracy, producer's accuracy, Kappa coefficient.
- Methods should be robust in time and space: dealing with data errors, updating misclassifications.
- Classification performance (using e.g., a contingency matrix), which will score the reliability of the semantic information extracted from the data.

Scalability:

- Processing time per unit of volume data.
- Classification time per unit of volume data.
- Methods could be hierarchical, working with hierarchies of features and classes.
- Methods could be flexible, allowing the introduction of new classes and features.

Degree of human intervention foreseen:

- Aim for 0% manual intervention during processing, but expect manual checking (10% overall effort).
- Manual corrections feed-back to training of classifiers, which keeps the classification accuracy better than 90%.
- Some human intervention needed, as features and classes may need to be defined on the spot, depending on the requirements.

[26] Nesrine Chehata, Li Guo, Clément Mallet, AIRBORNE LIDAR FEATURE SELECTION FOR URBAN CLASSIFICATION USING RANDOM FORESTS, In: Bretar F, Pierrot-Deseilligny M, Vosselman G (Eds) Laser scanning 2009, IAPRS, Vol. XXXVIII

4.4 MULTIVARIATE SURFACE GENERATION

The task will focus on the integration of multivariate data sets, with the objective of providing the correct geospatial and temporal alignment of the different datasets, and the definition of a new data set where various entities are represented (multivariate case). As an example, we describe the service that could be activated to generate surface models for water flooding simulations in urban areas. The surface model should be complete, detailed and accurate: we propose to integrate point clouds and dual surfaces available in IQmulus, from all sources, to fulfil the needs of these simulations. Instead of filtering point clouds (e.g., removing and all raised objects removed), the integrated data set will still contain all of them: all of these objects are important because they are obstacles to water expansion.

We will thus develop methods to integrate data fusion and feature extraction (see Task 4.2 and Task 4.3) to properly present all the surfaces that could be relevant for the simulation process (e.g., obstacles that are potentially anchored in the ground, and the ones that are not) and identify non-anchored objects (cars, buses, etc.) that could be carried away by the flood and that could act as corks enhancing locally the effects and the danger of the flood. The methods and algorithms will be designed in order to be robust to scaling up. The surface generation is also a fundamental step to support further steps such as high quality and reliable spatial data visualization; to provide computational efficiency and feasibility, and to guarantee quality description of all results.

By using the novel Locally Refined B-spline technology from Partner SINTEF anchored objects can be integrated into one large compact LR B-spline surface from which the simulation can be run. Non-anchored objects are treated as independent objects in the model.

The locally refined B-spline representation is well suited for making a compact validated terrain model that represents terrain shape. This will allow efficient registration of new point clouds with respect to the master model, and allows local remodelling of the master model when this is to be changed.

Baseline

Concerning the LR B-spline technology, a first version of the LR B-splines has been integrated into the SINTEF GoTools Library focused on spline technology. The first research paper where the LR B-spline basics are described was accepted for publication in the journal Computer Aided Geometric Design in December 2012 and will appear in 2013, preprints are already available. The first terrain models using LR B-splines have been generated in the SESAR JU Project 12.4.9 *3D airports for remotely operated towers* based on GIS-data from Avinor in Norway.

Issues

The development of algorithms for generating LR B-spline surfaces is just starting, and the algorithms to be developed in IQmulus have to be able to adapt to the characteristics of the geographical shape. Most terrain shapes not modified by humans, such as hills and mountains in Norway, with a few exceptions, were smoothed by glaciers during the ice age. In Italy the mountains were shaped by erosion giving a much more uneven shape, Hungary on the other side is dominated by smooth slowly-varying terrain shapes.

Development Directions

With respect to LR B-splines we foresee a development of methods for:

- Generating the master model adapted to the surface characteristics of the region modelled.
- Registration of data sets with respect to the master model.
- Updating the master model based on shape changes.
- Validating the master model with available datasets and pointing out discrepancies
- Prepare LR B-spline simulation models where anchored objects are integrated into LR B-spline surfaces to allow for simplified simulation.

4.5 CHANGE DETECTION AND DYNAMICS

Detecting changes in geospatial data sets is particularly useful in land management and decision making processes. The activities under this task will mainly concentrate on changes that are relevant for the demonstration scenarios and that can be generalized to morphological changes, taking place either on the seabed (e.g., sand dunes and sandbanks evolution), or on the land (e.g., landslides evolution). We will concentrate mainly on two approaches: the raster approach, where cross correlation can provide information on surface displacement and height differential can give the vertical motion; the vectorial approach, where surface models are extracted first and relevant morphological features are indexed so as to evaluate similarities/dissimilarities between their 3D shape and location. The latter approach can be particularly useful when deformable shapes are to be monitored and characterized.

A third approach, related to the raster approach, is to use the validated locally refined B-spline representation described in 4.4. In this setting the new clouds of data points first have to be registered with respect to the LR B-spline model. Then the difference between the new point cloud and the LR B-spline surface can be calculated directly. The approach is well suited for parallelization, as the LR B-spline surface can simply be split into subsurfaces covering specific, possibly overlapping areas, which allows splitting the points to be checked into subsets well inside a subsurface. Sufficient overlapping of the subsurfaces ensures that these subsets of points only have to be compared with one subsurface.

Baseline

As the LR B-spline technology is integrated into the SINTEF GoTools library, the functionality of GoTools can be extended with efficient closest point calculation for LR B-splines. Currently GoTools has an efficient closest point calculation for B-spline surfaces based on iteration. The closest point technology for B-splines was developed for comparing measured points to Computer Aided Design models, a calculation that generally is more complex compared to a terrain LR B-spline surface.

Issues

While tensor product B-splines surfaces are built form a regular mesh of rectangular polynomial surfaces, LR B-splines surfaces are built from a patchwork of rectangular polynomial surfaces of varying sizes. The sizes are determined during the building of the surface to introduce just the sufficient amount of degrees of freedom to model the shape. This partition of rectangles inherently has to be supported by more complex data structures than the regular structure of tensor product B-splines. Thus to properly support the iterative closest point calculations it is important to efficiently find the rectangular region entered when the iteration moves outside the current rectangle.

Development Directions

- Devise efficient navigation structures for LR B-splines.
- Implement efficient closest point iterators for LR B-spline surfaces.
- Representing the change as an LR B-splines surface overlaying the master LR B-spline surface, with further refinement than the master if necessary.

Indicators

Performance indicators

Quality

– Change detection accuracy (vs. reference such as ancillary punctual).

5 STATE OF THE ART IN INTERACTIVE VISUAL DECISION SUPPORT

A preliminary analysis of the workflow processes of the stakeholders within IQmulus revealed that today, decision making is often a two-staged process involving the following users and roles:

– **experts** prepare the decision-making using sophisticated tools to collect and combine information in an appropriate way and generate visualizations where the essence of the analysis is rendered in an obvious way for
– **decision-makers** who actually take the decisions.

The latter form of visualisations is of limited interactivity – if interactive at all, whereas the preparation requires – more than ever – the possibility to interactively analyse digital information in a visually revealing form. Efficient visual interaction/interactive visualisation is key for human beings in generating hypotheses about causes and effects and their inter-relationships, thus supporting the visual thinking of human beings.

The IQmulus stakeholders strongly believe that this splitting of roles and responsibilities will continue as a typical work habit. Therefore we foresee two dedicated components on the visualization layer in our conceptual IQmulus system architecture:

– Interactive visual decision support component based on leading-edge GPU technology to facilitate new levels of interactivity when handling large heterogeneous data sets;

- Web-application deployment component based on an innovative declarative 3D Internet approach allowing for deployment of dedicated 3D Web applications to decision makers.

Our main **objectives for interactive visual decision support** are:

- gaining new insights in complex formerly uncorrelated data-sets (in collaboration with the services provided by the processing layer)
- achieving interactivity in the mapping, rendering and navigation stages of the visualization pipeline[27] for large amounts of data of different content, structure and semantics
- leveraging the power of modern graphics processing units (GPU) technology
- providing more precise and accurate display of data and results
- achieving higher visual fidelity
- providing new user experiences while considering scalability and security in open environments;
- automating and optimizing the data preparation processes for web-specific environments through web-server infrastructures.

5.1 COMPOSITE VISUALIZATION METHODS

A central part of IQmulus is the processing and extraction of knowledge hidden in very large data sets. One way of gaining new knowledge is by combining different types of data, and draw forth new meaning from their composition. The visualization of such data-sets into a single common view is essential to get the overview required in decision making: nothing can beat the eyes of a trained expert in certain situations. A naïve example is the combination of digital elevation models, satellite imagery and flood simulation results. A flood simulation model will not always be able to correctly capture "small" features such as train tracks or highways. One reason for this is that the simulation model can have one elevation representing a 15 by 15 meter area or more. These features, however, act as levees and often have a strong impact on floods. An expert user will quickly discover if the flood simulation takes roads and train tracks into account, whilst it can be difficult to do so automatically for a computer. A second example is that human experience and intuition can give a skilled operator insight in new and unforeseen ways.

This task will focus on high-quality interactive visualization of previously uncorrelated data-sets, based on a modern hardware paradigm.

Baseline

There is a plethora of techniques for scientific visualization of different types of scientific data today. However, many techniques are designed for an out-dated hardware paradigm (e.g., fixed function OpenGL), and they are most often tailored for a specific use and data representation. The efficient combination of existing visualization techniques for uncorrelated data-sets in a modern hardware paradigm is therefore an open research question. For example, the rendering technique for different data-sets might be “incompatible”, in the sense that their composition is highly non-trivial.

[27] Matthew Ward, Georges G. Grinstein, Daniel Keim; Interactive Data Visualization; Taylor & Francis Ltd; 2011

Issues

Whilst there exist a range of techniques for scientific visualization of different data today, the composition of such visualizations is often non-trivial without loss of interactivity or fidelity. Furthermore, many existing techniques are based on an outdated hardware paradigm, and are only able to handle relatively small data sets.

Development Directions

This task will focus on high-quality interactive visualization of previously uncorrelated data-sets. Example types of data-sets include

- Vector based cartographic information (e.g., hydrography, cadastre, roads, utility lines),
- Raster based terrain data (e.g., elevation data, land use), and
- Vector and tensor based simulation data (e.g. , wind, pollution, flooding).

Interactivity is a key concept in this task, as exploring a dataset in real-time gives the user more insight than still images are capable of. The development of high-quality visualization for interactive data exploration will involve

- Development of new composite visualization algorithms
- The use of modern hardware features
- The use of multi-resolution and level-of-detail methods

Indicators

Important performance metrics for this task include interactivity and scalability. We therefore foresee that the developed methods execute at interactive frame-rates[28], and show good scaling with respect to data-set size.

5.2 VISUALIZATION TECHNIQUES UTILIZING NEWEST GPU FEATURES

Visualisation for supporting decisions can range from generating diagrams over visualizing simulation results in different dimensions and domains (2D, 3D, 4D, computational fluid dynamics, computational structural dynamics, etc.) and spanning up to non-photorealistic illustrative rendering. The state of the art in visualisation is way too broad to by fully discussed here[29] and only partially relevant to IQmulus and our visualization objectives.
The heterogeneity of the more non-traditional data sets in GIS, namely coverage data and simulation data create new challenges to the processing and visualization stages of heterogeneous information management and visual decision support systems.
The key **enabling technology** of this visualization layer component are **graphics processing units** (GPUs). Over the last years, GPUs have shown much better performance increase factors than CPUs. Interestingly, even Intel published a study and committed that the actual speedup of GPUs in comparison to CPUs is around a factor of fourteen[30]. The study focused

[28] We define interactive frame-rates in this context to be <u>at least</u> 10 frames per second for the most demanding visualization techniques for large data-sets. Motion pictures are today often 24 frames per second, and interactive games run at 30 frames per second or more.
[29] The yearly IEEE VisWeek conferences (Vis, VAST and InfoVis) provide a good source for the current STA in the field. Additionally, several books in the field give overviews on information and or scientific visualization.
[30] Debunking the 100X GPU vs. CPU myth: an evaluation of throughput computing on CPU and GPU, 2010

solely on general purpose computations on the GPU. Although, this is an impressive factor "still", our experience is that for dedicated interactive visualization techniques the speed-up of GPU-based versus traditional visualisation approaches can even be considerably higher[31]. This clearly reveals that GPUs are the tool of choice for high performance interaction visualization solutions. Thus, we limit our discussion of the state-of-the-art in this section on GPU-based visualization techniques.

Baseline

GPU-based visualization techniques are an increasing research topic since the introduction of freely programmable graphics hardware pipelines in 2004. Extensions to the functionality and capabilities of modern graphics hardware made this a continuous research field since then.
Daniel Weiskopf summarizes the methodology and technologies in his book on GPU-Based Interactive Visualization Techniques[32]. He outlines how smart combinations of computer graphics features such as shader programs, 3D textures, and color buffers can be used to convey data from many different domains into visual representations.
More recently – to just give two examples - Bürger et al. showed a fast extraction and visualization of surface representations of flow-fields[33]. In the field of medical imaging especially volume rendering techniques are accelerated lately with the utilization of new GPU features[34]. Beside these two papers, there are more singular research activities which report their results on conferences such as VisWeek (http://www.visweek.org/).

Issues

As Dan R. Lipsa et al.[35] outline in their current State-of-the-Art report on Visualization for the Physical Sciences , there was only (comparably) little research on the application of GPUbased techniques for visualization in the area of earth sciences and physics over the past years.
This very recent statement has motivated our contribution to IQmulus and its research programme. We assume the statement in the report can be attributed to the fact that a large part of the scientific visualization research community is currently focusing on server-side visualisation for the high performance computing area, the Ultra-Scale[36]. Those approaches do not concentrate on leveraging the local GPU resources typically contained in engineering desktop workstations as used by experts in the decision preparation stage or the end users (decision-makers) who in some cases only access via web interfaces (for GPU-based web viewing technology see section 5.3).
Other reasons mainly include data amount issues. The GPU memory amount is typically a factor of ten smaller than compared to workstation CPU RAM (6 GB to 64 GB). In IQmulus, we foresee that filtering mechanisms of the processing layer will boil down data amounts to a size which is manageable on a GPU workstation.
We do not consider the transfer of GPU based visualization techniques to the geospatial domain to be easy. Therefore, this has to be evaluated in this research project.

---

[31] According to internal benchmarking, GPU-based visualization approaches can show a speedup factors in the order of 30 compared to CPU approaches.
[32] Daniel Weiskopf: GPU-Based Interactive Visualization Techniques, Springer, 2007
[33] Interactive Streak Surface Visualization on the GPU, Kai Bürger, Florian Ferstl, Holger Theisel, and Rüdiger Westermann, IEEE VisWeek 2009
[34] Mapping High-Fidelity Volume Rendering for Medical Imaging to CPU, GPU and Many-Core Architectures, IEEE VisWeek 2009
[35] Dan R. Lipşa, Robert S. Laramee, Richard Walker, Jonathan C. Roberts, and Simon Cox, Visualization for the Physical Sciences, EUROGRAPHICS 2011, State of the Art Reports, pages 49-73, 11-15 April 2011, Llandudno, Wales
[36] Visualizing Ultra-Scale Data, Siggraph 2008

Development Directions

A very important graphics technology mainly used by games, but with an excellent potential for interactive 3D scientific visualization, is deferred shading[37]. It is an approach to defer complex shading computations to the screen space by creating a series of screen image buffers containing scene data as seen from the current view point. Computations, such as shading, are then run in parallel on the image buffers by the GPU, both increasing the overall performance, and providing new effects through subsequent image processing operations. Deferred shading is nowadays indispensable in the gaming industry. However, researchers have just begun to generalize and apply this technique for scientific visualization[38]. We will research new techniques to further adopt the idea of deferred shading for scientific purposes, especially tailored towards physics and geo data.
Therefore we need to find ways to store abstract data properties such as result values, metadata, and identifiers in screen space image buffers. Depending on the nature of the heterogeneous geo-information data input, this will require novel methods for conveying such data into GPU-friendly representations for rendering. Once the screen space image buffer data is available, it acts as input for the deferred data processing.
Scientific visualization needs mapping techniques, for example false contouring of result values using transfer functions. Thus, we will research ***deferred mapping*** techniques in the first place. In the second place, we will research how the given abstract screen space data buffers can serve as basis for novel ***deferred fetching*** techniques. Depending on the source data properties, it can be beneficial to first identify the visible portions of data geometry, and fetch other parts of the data set based on this information in a staged fashion. We expect performance improvements over conventional approaches especially if fetching operations are costly, or if one explores only subparts of a complete data set at a time.
We will research how the deferred mapping and deferred fetching techniques can be combined with interaction methods. Ben Shneiderman describes interaction methods in his well-known Readings in Information Visualization[39]:

- brushing/linking, focus/context, detail on demand and
- direct manipulation, continuous level of detail, etc.

We will research combinations of deferred mapping with user interaction techniques for direct manipulation, as well as combinations of deferred fetching with brushing/linking and focus/context.

These interaction techniques are tailored for interactive use. Thus, they will greatly benefit from performance improvements by deferred visual data processing.
The combination of techniques shall go hand-in-hand with the GPU-based visualization techniques mentioned beforehand, so that the whole interactive visualization system remains logically on the GPU and harnesses the graphics hardware most efficiently.
Many data sources (sensors, simulation tools, etc.) produce data in formats which are mainly determined by the characteristics, needs, requirements and restrictions of the source itself, instead of being tailored for purposes of interactive visualization. This is especially problematic if the data does not fit into the machine memory and becomes even harder when direct interaction and manipulation is needed.

---

[37] Mark Harris, Shawn Hargreaves. Deferred Shading. http://http.download.nvidia.com/developer/presentations/2004/6800_Leagues/6800_Leagues_Deferred_Shading.pdf, 2004

[38] B. McDonnel, N. Elmqvist. Towards Utilizing GPUs in Information Visualization: A Model and Implementation of Image-Space Operations. In IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2009), 15(6):1105-1112, 2009.
Niklas Elmqvist, Pierre Dragicevic, Jean-Daniel Fekete, "Color Lens: Adaptive Color Scale Optimization for Visual Exploration," IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 6, pp. 795-807, June 2011, doi:10.1109/TVCG.2010.94

[39] Ben Shneiderman, with Stuart K. Card and Jock D. Mackinlay; *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999. ISBN 1-55860-533-9

New graphics hardware features make it essential to re-think the usage of data structures and formats.
Thus, we are addressing the following research questions:

- Which are the data properties that permit or impede the utilization of modern graphics hardware features?
  This will be done for different rendering techniques (ray tracing vs. deferred shading), taking different typical post-processing algorithms into account (iso-surfaces vs stream-lines).
- Which existing data structures and file formats cope with those requirements and where do they fail?

Within this research, we analyze the actual difficulties encountered when applying GPU-based visualization techniques within IQmulus and suggest extensions for existing data formats, or new data formats. Based on this analysis, we will formulate, develop and select requirements for data structures and file formats. Such requirements may comprise the following characteristics: streamable, progressive, seekable (jumping to certain position, i.e., time steps), etc. The goal is finding ways to tighten data transfer chains down from processing layers to graphics hardware memory, effectively minimizing the "time to screen" for an improved user experience.
The application of these findings in IQmulus is not foreseen since most layers would be affected by data format changes, requiring additional adoption and development efforts by the affected project partners.The basis for these developments is our interactive visualization framework (iFX: www.i-fx.net) which IGD brings into IQmulus as background. iFX has been designed as an extensible framework having the following aims in mind: joy of use , efficient visual analysis and goal orientation through in-scene interaction elements, direct manipulation and immediate feedback.
iFX leverages GPU features to a certain extent according to the OpenGL 2.0 and Shader Model 3.0 standards and shall be extended in IQmulus to exploit OpenGL 4.2 and Shader Model 5.0 capabilities. In the past few years iFX has successfully been tailored towards a base 3D scientific visualization component of a commercial simulation tool which has a user basis of 10.000 seats world-wide.

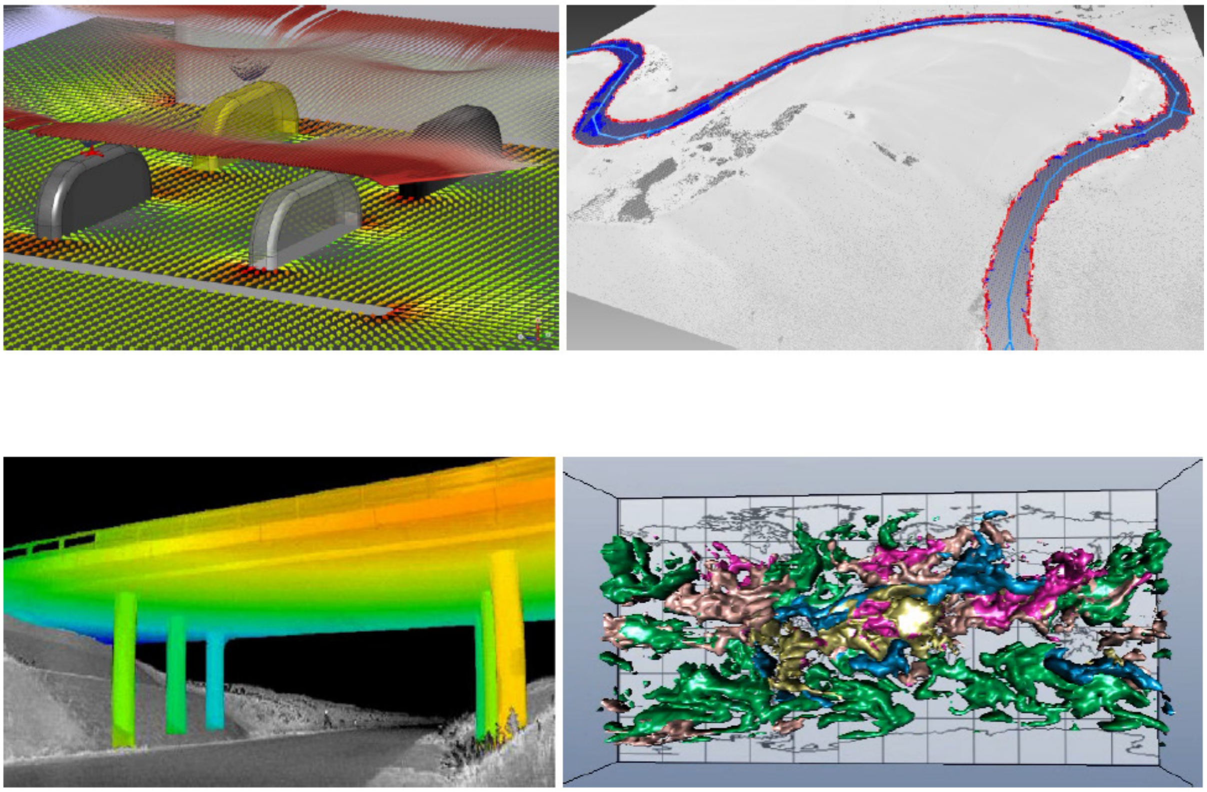


Figure 2: top left: scientific visualization framework (iFX); top right: integration of ATKIS and LIDAR data for semi-automatic classification of land / water masses and extraction of land-water borderlines; bottom left: combination of simulation data with coverage data; bottom right: visualization of a time step of a global wind field

In summary, the following results will go beyond the state of the art:

- Designs, concepts and technologies for utilizing modern GPU features for the purpose of scientific visualization
- Extensions to the concept of deferred shading beyond the application in games
- New screen space rendering algorithms to the scientific visualization community
- Algorithms for fast interaction which will exploit the additional information given in screen space, due to our deferred shading approaches
- Requirements for visualization aware data formats with focus on new algorithms for modern GPUs
- Discussion of these requirements w.r.t. existing formats especially from the geo visualization domain
- Guidelines for file format selection
- Basis for designing new file formats and extending standards.

5.3 3D-WEB-BASED VISUALIZATION

Distributed data-centred applications are one of the common implementation concepts for scientific visualization solutions today to process huge data sets, and therefore web-service architectures or even cloud-based solutions are just the next logic step. Recent developments in the area of high performance web-service architectures and the requirement to provide applications not just for a small expert group lead to new approaches in the field of web-based scientific visualization. The just emerging support for GPU-supported and therefore high-performance graphics in modern web-client implementations and standards provide a new application environment, which is especially interesting for the demands of scientific

visualization solutions. Therefore, some of the scientific visualization packages (e.g. ParaView with ParaViewWeb from Kitware) already utilize this opportunity to move the established application model to a Web and cloud-based solution. Deploying the current application model to a new environment can just be the first step. However, this field is not yet explored to utilize its full potential. Tailored web apps with specific UIs and interaction methods allow advanced user experiences that lead towards more efficient decision making processes that enable the user to directly grasp the problem by interacting with it.

Baseline

The goal of the proposed approach is to provide important features for scientific visualization[40] in a specialized web application, which therefore must be able to handle visualization-specific representations that consist of registered and merged point, surface, and even volume data. In this regard, not only the choice of an appropriate visualization technique but also the possibility to interactively explore the data set (e.g., by changing the value range or transfer function) is of great importance for grasping and analyzing the information and therefore requires real-time capable mapping and rendering methods, which nowadays is often achieved using GPU-based methods[41]. This is especially of interest, since some 3D scanning devices such as 3D scanners, LiDAR equipment, and sonar scanners all deliver their data as point clouds. The GPU abstraction in WebGL already includes basic point and polygonal primitives. A recent framework to additionally simplify streaming and rendering of point clouds based on JavaScript and WebGL is e.g. "XB PointStream"[42].
In Congote et al.[43] the implementation of a direct volume rendering system for the Web[44] is discussed. In this publication the authors discuss both, performance and scalability, of the volume rendering by WebGL ray-casting in two different but challenging application domains: medical imaging and radar meteorology. Currently, this work is integrated into X3DOM[45] according to the new X3D[46] volume rendering component extension. In Schwarz et al.[47], the authors propose a WebGL-based framework for representing reflectance information via Bidirectional Texture Functions, which allows for the progressive transmission and interactive rendering of digitized artifacts consisting of 3D geometry and reflectance information. This is achieved by employing a novel progressive streaming approach for the huge BTF data set that allows the smooth interactive inspection of a steadily improving model during download. Analogously, in X3DOM we are currently following a similar approach for lightweight geometry compression and transmission with our image geometries that likewise utilize image compression techniques.

[40] [LLS10] Dan R. Lipsa, Robert S. Laramee, R. Daniel Bergeron, and Ted M. Sparr, Data Representation for Scientific Visualization: An Introduction. Ch. in Computer Graphics. Nova Science, 2010.
[41] Tony McLoughlin and Robert S. Laramee and Ronald Peikert and Frits H. Post and Min Chen, Over Two Decades of Integration-Based, Geometric Flow Visualization, in Eurographics 2009 - State of the Art Reports, Eurographics Association, 2009, pp. 73-92.
[42] XB PointStream. 2011. http://scotland.proximity.on.ca/asalga/demos/freecam/
[43] John Congote, Alvaro Segura, Luis Kabongo, Aitor Moreno, Jorge Posada, and Oscar Ruiz. 2011. Interactive visualization of volumetric data with WebGL in real-time. In Proceedings Web3D '11. ACM, New York, USA, 137-146.
[44] Chris Marrin. WebGL Spec. 2011. https://www.khronos.org/registry/webgl/specs/1.0/
[45] Johannes Behr, Yvonne Jung, Jens Keil, Timm Drevensek, Peter Eschler, Michael Zöllner, Dieter Fellner. A scalable architecture for the HTML5/ X3D integration model X3DOM. In Proceedings Web3D 2010, New York, USA, ACM Press, pages 185—193, 2010.
[46] Web3D Consortium, Extensible 3D (X3D), 2011. http://www.web3d.org/x3d/specifications/
[47] Christopher Schwartz, Roland Ruiters, Michael Weinmann und Reinhard Klein: WebGL-based Streaming and Presentation Framework for Bidirectional Texture Functions. In Proceedings VAST 2011, Eurographics Association, pages 113-120, 2011.

Issues

Currently, the open ISO standard X3D is the only standardized 3D deployment format and differs from other 3D formats (like the interchange format COLLADA) in that it also includes the scene's runtime behaviour. The standard already provides point and surface primitives, and volume rendering will be part of the next minor revision. In addition, since most geo-referenced data is provided in a geodetic or projective spatial reference frame, the X3D Geospatial component[48] supports the Spatial Reference Model, high-precision geospatial modelling, and terrain rendering. Thereby, X3D already provides a solid foundation for scientific visualization tasks and enables an automated connection of existing data with atmospheric, oceanographic, or geological data to be visualized in the Web.

However, these X3D components only extend X3D for various application and visualization scenarios but do not address one of the essential issues of X3D, namely that X3D is still bound to a plugin integration model that has major usability and performance issues especially for large data sets [49]. WebGL (an open JavaScript binding for OpenGL ES 2.0 in the Web Browser[50]), Adobe's Flash11 with Stage3D, and Microsoft's Silverlight 5 now all provide access to the native GPU layer, but the issue of the missing DOM integration still exists. In this regard, with the X3DOM project recently a DOMbased integration model for declarative (X)3D in HTML5 was proposed, that allows a seamless integration of 3D contents in the HTML document model.

A service-oriented architecture (SOA) is a software architecture paradigm for structuring and using distributed functionalities that are managed by various owners. Here, web services help supporting the collaboration between different applications running on different platforms by utilizing REST, JSON, and XML-based standards like UDDI, WSDL, and SOAP as web-service protocol [51]. In this regard, one prominent application of web services are the standardized geo-services (WCS, etc.[52]) provided by the OGC to make spatial geo-data accessible in a structured way. Hence, some rendering systems already make prototypical use of REST and JSON for scene-graph manipulation and distributed rendering [53], build a workflow pipeline from data acquisition up to interactive web-based visualization in the Cultural Heritage domain based on web-services or build smart device networks for mobile application[54] .

Development Directions

Currently, there is a strong trend towards 3D data and 3D documents (like 3D scanners and printers, geo-referenced data, and Augmented Reality), as well as a convergence of application platforms (like W3C WebApps, HTML5, etc.) towards web-based content. However, there are still various system platforms with very different soft- and hardware

[48] Michael McCann, Richard Puk, Alan Hudson, Rex Melton, Donald Brutzman. Proposed enhancements to the X3D geospatial component. Web3D 2009, pp. 155-158.
[49] Johannes Behr, Yvonne Jung, Jens Keil, Timm Drevensek, Peter Eschler, Michael Zöllner, Dieter Fellner. A scalable architecture for the HTML5/ X3D integration model X3DOM. In Proceedings Web3D 2010, New York, USA, ACM Press, pages 185—193, 2010.
[50] Chris Marrin. WebGL Spec. 2011. https://www.khronos.org/registry/webgl/specs/1.0/
[51] Thomas Erl. Service-Oriented Architecture: Concepts, Technology, and Design. 2005, Prentice Hall Intl.
[52] Peter Baumann. Flexible, Open, Free: The New OGC WCS 2.0 and Its Reference Implementation. In Proc. FOSS4G 2010.
[53] Schiffer, Thomas; Schiefer, Andreas; Berndt, Rene; Ullrich, Torsten; Settgast, Volker; Fellner, Dieter W. Enlightened by the Web: A Service-oriented Architecture for Real-time Photorealistic Rendering. 5. Kongress Multimediatechnik Wismar, 2010.
[54] Conrad Thiede, Christian Tominski, Heidrun Schumann. Service-Oriented Information Visualization for Smart Environments. 2009.

requirements such as smart phones (iOS, Android, RIM, WindowsPhone), different desktop systems and cloud bases and so on, which typically make cross-platform development rather complicated and time consuming.
Thus, the goal is to have web services that automatically convert multi-dimensional data from different sources (e.g. geo-spatial data and architectural models) into interactive 3D visualizations for the Web that can be **delivered as web application on any device**, thereby allowing on-site visualization for everyone. For example, value ranges that can be interactively explored like a temperature or density distribution could automatically be tagged like, e.g., it is often done in modern IDEs (such as the CG FX Composer framework) to provide all data that is necessary for an appropriate use interface (e.g., a slider with a corresponding value range).
Therefore, the visualization needs to be scalable in such a way that mobile and desktop machines with rather different computing and 3D capabilities are supported alike (for example via a hybrid approach that provides both streaming for (low-end) clients and direct web-based 3D rendering for high-end desktop computers). Finally, security aspects are of high importance. Here again, streaming technologies can help with information hiding in that only image streams but no 3D models are transferred to the client for display.
Obviously, a unified standard for shared services and communication is essential, too. Therefore, declarative, XML-based languages (such as the X3D standard) are suited well for SOAs in that content (such as geo-spatial data) can be directly transformed, e.g., via XSLT, from one representation to another.

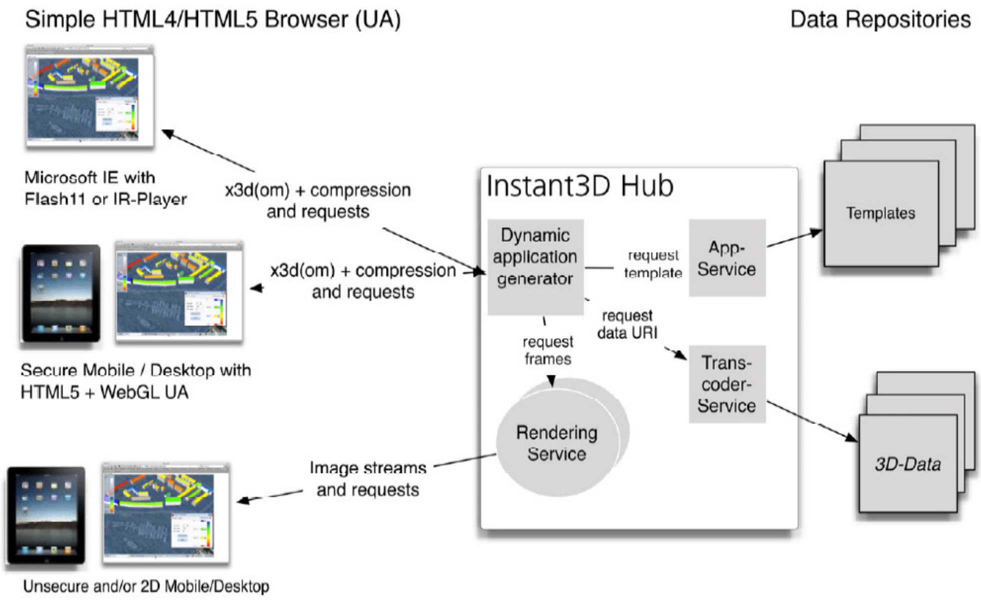


Figure 3: Instant3D Hub: scalable, secure, and efficient SOA for 3D visualizations on the Web.

Figure 3 broadly shows the concept of the planned service-oriented architecture for the web-based visualization deployment layer, which will mainly address the decision makers' needs but possibly even those of on-site users. The core component is the "Instant3D Hub" that acts as application provider, whereas the application identification is resolved via a URI. Thereto, the app service takes a request from the Hub and extracts an application template, which at least contains the HTML(5) pages, the necessary meta-information of the respective application, as well as descriptions of all required data containers, e.g. in X3D format. For data preparation, the template specifies the data and application characteristics.
After that, the Hub invokes the appropriate transcoder service for providing the concrete data (e.g., an X3D scene with certain camera positions). Thereto, the transcoder service translates and, where required, caches the 3D data from already existing descriptions by converting them from various formats, which are not yet suitable for real-time web

presentation, to, e.g., X3D as declarative deployment format. The result is then provided as URI.

Depending on data security aspects and the respective web browser capabilities, the Hub generates a concrete visualization application. In contrast to the expert tool, which is based on IFX and explained in the previous section, the generated web app will have a dedicated functionality and user interface. On the one hand, this will be realized via client-side rendering by utilizing X3DOM, which uses WebGL or Flash 11/ Stage 3D for real-time rendering. One open research issue is scalable methods for binary compression of big geometric data – e.g. by utilizing our proposed image geometry (IG) approach – and of material data as well as suitable caching strategies. The advantages of client-side rendering are a very simple server infrastructure and highly interactive apps since everything is rendered on the client. Disadvantages are that the data-load can easily overburden the client, that lots of 3D data needs to be transferred, and security or IPR issues.

On the other hand, server-side rendering will be used, too. The approach includes a scene-graph manipulation model which allows to control client and server side data through a single unified API. This allows the environment to select client, server or even hybrid visualization very late during the last deployment state. The final selection depends on security, device/browser capability, user and preference aspects. In this case, for instance an X3D runtime environment is utilized for rendering, which streams the rendered image frames via MJPEG to the client. Since interactions are handled via WebSockets or XMLHttpRequests, here the latency is much higher than for client-side rendering. Another disadvantage is the fact that a complex server infrastructure is required. However, since no 3D data but only images are transferred over the network, there are no IPR issues concerning the 3D data and data security is inherent. Also, as another advantage of server-side rendering, the visualization can be executed on arbitrary clients.

The simplified SOA architecture includes three specific REST-based resource interfaces which are responsible for the application and UI retrieval, transcoding and server-based rendering. All those resource providers are always known to the centralized HUB and therefore do not rely on an explicit discovery action.

Instead of simply delivering data to a rigid viewer, as the traditional model does, this architecture delivers a user experience that is tailored for specific user scenarios and the actual context. The final environment will be able to derive and deliver a dynamic application from this context with regards to user, system, security, and data requirements. This allows the system to provide dynamic technical solutions and automatically choose a client, server, or even hybrid visualization for an unchanged application at runtime, while also considering the device capabilities as well as IPR and security concerns. What might be even more important, besides the technical aspects, is the ability to provide very use-case-specific applications for interactive web-based visualization, since the system is not built as a rigid application model for a small expert group. Thus, this enables new application scenarios, including scientific visualization for decision support applications for decision-makers, as well as applications for public information and on-site scenarios, which completes the full-fledged desktop application envisaged in the previous section.

In summary, the proposed approach focuses on two important aspects:

- it will implement an established application model in the new web-client and cloud-based environment, and
- it will also introduce a web-service architecture as core of the application distribution and deployment infrastructure.

PART TWO: USER ASPECTS

6 USER STORY ANALYSIS RESULTS

Using our user story analysis methodology we identified the possible application areas and input data from the currently available stories. The following application areas have been identified:

- Modeling
- Data collection and transformation
- Data classification
- Visualization
- Simulation and prediction.

All application areas can be linked to multiple project goals, as can be seen in the following sections. It can be seen that most user stories contain data management and transformation related issues, while some focus on data visualization techniques. Several user stories are complex and fall into multiple categories. Also, most user stories require the handling of heterogeneous, multi-source and multi-format input data, mainly LIDAR, point-clouds and digital terrain models (DTM).

Table 1 contains the summary of initial user story analysis results.

User stories	Application area							Input data							
	Modelling	Data collection and transformation	Data analysis and monitoring	Data classification	Visualization	Simulation and prediction	Other	Orthophoto	LIDAR	Terrain model	Remotely sensed imagery	Point cloud	Road network	Vector database	Terrestrial Laser Scanner
1		X	X	X	X			X	X						
2		X			X					X					
3		X	X						X	X					
4		X	X							X					
5				X											X
6			X	X	X			X		X		X			
7			X	X	X							X			
8		X	X	X						X					
9				X	X	X				X					
10			X							X					
11		X	X						X						
12		X	X							X		X			
13	X		X		X	X				X					
14					X										
15					X	X									
16			X			X						X			
17					X	X									
18					X										
19					X										
20					X										
21		X			X			X							
22			X		X			X			X			X	
23	X			X	X				X						
24			X	X	X					X				X	
25	X				X	X			X			X			
26	X		X								X				
27			X		X								X	X	
28		X	X	X	X			X					X	X	
29		X								X				X	
30			X								X			X	
31		X								X	X			X	

User stories	Application area							Input data							
	Modelling	Data collection and transformation	Data analysis and monitoring	Data classification	Visualization	Simulation and prediction	Other	Orthophoto	LIDAR	Terrain model	Remotely sensed imagery	Point cloud	Road network	Vector database	Terrestrial Laser Scanner
32	x					x			x	x		x	x	x	
33	x		x	x	x			x	x	x	x	x	x	x	
34		x						x	x		x	x	x	x	
35			x	x							x			x	
36			x	x				x		x	x			x	
37		x		x					x						
38		x		x								x			
39				x				x		x	x				
40				x										x	
41		x		x										x	
42		x						x			x				
43			x											x	
44		x			x			x		x	x				
45	x	x							x						
46				x					x						
47		x							x						
48		x										x		x	
49		x	x						x					x	
50		x		x					x						
51			x	x		x				x		x			
52			x			x				x					
53			x			x								x	

The detailed analysis of the user stories in the application areas can be found in chapters 7 to 12.

7 MODELLING

Requirements

- 3D city model generation including buildings.
- Measuring multiple aspects of buildings (area, height).
- Detecting effects of environmental or industrial disasters (e.g., flood) to buildings.
- Detecting effects of environmental or industrial disasters to calculate an area of risk (polygon).

Constraints

- The model must be as accurate as possible.
- The model must handle temporal aspects as simulation requires it.
- Operations must perform as fast as possible even with large datasets.
- Calculations must handle the complete 3D model (including all building models and elevation data) of any city or equal area.

Relations

- Related User Stories: 13, 23, 25, 26, 32, 33, 45
- Related Scientific and Development topics:
  - 4.1 Spatial Data Processing Architectures and Interfaces
  - 5.1 Heterogeneous Data Integration
  - 5.2 Spatio-Temporal Data Fusion
  - 5.4 Multivariate Surface Generation
  - 5.5 Change Detection and Dynamics

Sample data

Sample data available:

Dataset/Service ID	Organization name	Dataset/service title
4	6 FOMI	Land Parcel database
6	6 FOMI	Digital Surface Model (DSM)
14	6 FOMI	Building boundary database
15	9 IGN	Paris 3D Model
22	12 Liguria	LIDAR (XYZ)
23	12 Liguria	LIDAR (LAS)

8 DATA COLLECTION AND TRANSFORMATION

Requirements

- Accessing data from WMS/WFS services.
- Creation of multi-resolution DTM based on heterogeneous point clouds to have the best available resolution at any location.
- Merging terrain models of different resolutions to have the best available resolution at any location.
- Assessing the quality of a newly acquired point cloud or DTM in comparison to reference data.
- Updating vector datasets in a particular area by using newly available vector data.
- Verifying building cadastral record adequacy in the cadastral database by comparing the building contours with orthophoto and DTM.
- Computing displacement maps from DTM.
- Combining 2D/3D polygons with LIDAR data to extract building elevation.
- Combining Topographic data and LIDAR data to detect evolutions of a landscape to improve the database update process.
- Superposing LIDAR data on top of a terrain model to evaluate the quality of the terrain model.
- Superposing rater data onto the terrain model.
- Real-time update of road status attributes.
- Extracting submarine dune shapes and breaklines from digital terrain models, point clouds.
- Extracting coastline from LIDAR and DTM.
- Extracting forest data from LIDAR dataset (such as vertical distribution of wood/leaves, tree position, height, size or canopy diameter).
- Extracting values from one to several raster layers at a given location and export them to spreadsheets.
- Statistically summarizing values from raster and point cloud data within a given polygon.
- Raster to vector conversion of classified layers.
- Generating land cover categories in a given area using orthophotos and satellite imagery.
- Generating NDVI maps based on remotely sensed data from different sources.
- Generating crop cover maps for crop parcels from remotely sensed images taken at different dates.
- Generating waterlog maps (with at least the following categories: open water surface, flooded vegetation, wet soil) in a particular area.

Constraints

- The amount of data is usually large from 1GB to several TBs.
- Operations must perform as fast as possible even with large datasets.
- Operations can be customized with parameters.

- Several data formats must be supported (including 3D building models) and mobile/remote import of data must be available.
- Record all changes in the dataset to quickly review and finalize them.

Relations

- Related User Stories: 13, 23, 25, 26, 32, 33, 45
- Related Scientific and Development topics:
  - 4.1 Spatial Data Processing Architectures and Interfaces
  - 5.1 Heterogeneous Data Integration
  - 4.2 Domain-Specific Languages (DSLs)
  - 5.2 Spatio-Temporal Data Fusion
  - 5.3 Feature Extraction, Classification and Correlation
  - 5.4 Multivariate Surface Generation

Sample data

Sample data available:

Dataset/Service ID	Organization name	Dataset/service title
4	6 FOMI	Land Parcel database
5	6 FOMI	Orthophotos
6	6 FOMI	Digital Surface Model (DSM)
7	6 FOMI	Digital Elevation Model (DEM)
8	10 UBO	Digital Terrain Model (DTM)
9	10 UBO	Digital Terrain Model (DTM)
10	10 UBO	Point Cloud from TLS
11	10 UBO	Point Cloud from TLS
12	10 UBO	Digital Terrain Model (DTM)
13	10 UBO	Digital Terrain Model (DTM)
14	6 FOMI	Building boundary database
15	9 IGN	Paris 3D Model
17	12 Liguria	Bathymetry ponits
18	12 Liguria	Bathymetry grid
19	6 FOMI	Road Network
20	12 Liguria	Regione Liguria Forestry map
21	12 Liguria	Regione Liguria Land Cover map
22	12 Liguria	LIDAR (XYZ)
23	12 Liguria	LIDAR (LAS)

9 DATA ANALYSIS AND MONITORING

Requirements

- Statistically summarizing values from either a raster or a point cloud within a given polygon (e.g., total surface of habitat type within designated area).
- Evaluating quality and accuracy of data. Comparing newly acquired data to reference.
- Verifying the adequacy of cadastral data using orthophoto and DTM.
- Detecting and quantifying displacements.
- Automatically detecting building contours.
- Extracting significant breaklines from high resolution DTMs or point clouds, such as dikes, dams, banks, ditches.

- Monitoring the evolution of topography, costal zones, dunes, etc., using data acquired at different dates.
- Monitoring the agriculture-related characteristics (plant species, breed, biological condition) of land (to provide more reliable agricultural damage assessment and yield loss estimation, etc.).
- Delineating the extent and the maximum possible extent of the area potentially flooded in case of an environmental or industrial disaster.
- Delineating slopes steeper than a given threshold (with DTM, with or without LIDAR, etc.) to support detection of erosion risk areas.
- Calculating the difference between two datasets of point clouds.
- Route planning by taking flooded areas or other barriers into account.

Constraints

- Record all changes in the dataset to quickly review and finalize them.
- Operations must handle large amount of data (from 1 GB to several TBs).
- Operations can be customized with parameters.
- Results are interactively explorable.

Relations

- Related User Stories: 1, 3, 4, 6, 7, 8, 10, 11, 12, 13, 16, 22, 24, 26, 27, 28, 30, 33, 35, 36, 43, 49, 51, 52, 53
- Related Scientific and Development topics:
  - 4.1 Spatial Data Processing Architectures and Interfaces
  - 4.2 Domain-Specific Languages (DSLs)
  - 5.2 Spatio-Temporal Data Fusion
  - 5.3 Feature Extraction, Classification and Correlation
  - 5.5 Change Detection and Dynamics

Sample data

Sample data available:

Dataset/Service ID	Organization name	Dataset/service title
8	10 UBO	Digital Terrain Model (DTM)
9	10 UBO	Digital Terrain Model (DTM)
10	10 UBO	Point Cloud from TLS
11	10 UBO	Point Cloud from TLS
12	10 UBO	Digital Terrain Model (DTM)
13	10 UBO	Digital Terrain Model (DTM)
17	12 Liguria	Bathymetry ponits
18	12 Liguria	Bathymetry grid
20	12 Liguria	Regione Liguria Forestry map
21	12 Liguria	Regione Liguria Land Cover map
22	12 Liguria	LIDAR (XYZ)
23	12 Liguria	LIDAR (LAS)

10 DATA CLASSIFICATION

Requirements

- The classification of measurement data (multispectral satellite images, aerial photographs, RADAR, LiDAR) in the updating of geographical inventory systems
- The classification of different zones of coastline via the identification of rocks, vegetation, constructions and sand (using Terrestrial Laser Scanner dataset)
- The classification of changes in the evolution of topography
- The classification of changes in surface elements, e.g., rock slides on a cliff, in order to calculate the associated volumes of debris
- The classification of seabed to extract dune shapes and the evolution of dunes along time
- Visual evaluation of shaded DTM in order to detect rifts and other geological structures
- Creation and visualization of a 3D city model by the automatic classification of LiDAR data
- The classification of a slope map derived from DTM (and possibly from LiDAR) in order to find steep areas
- Real-time updating of road status attributes
- The classification of areas regarding floods (areas constantly and temporarily under water)
- The classification of crop cover of agricultural parcels
- The classification of areas regarding eligibility for agricultural subsidies and temporal stability

Constraints

- Inclusion of several data sets (orthophotos, DTMs or point clouds) in the classification
- Inclusion of data sets of the same kind collected at different dates (e.g., time series of multispectral satellite images) in order to correctly determine the land cover or land use
- Inclusion of data sets of the same kind collected at different dates in order to monitor temporal changes
- Exact co-registration of different data sets before their combined use
- In case of visual evaluation, interactive change of viewing parameters
- Integrated usage of height data derived from different sources (aerial photography, LiDAR, very high resolution satellites)
- The usage of mobile devices in field measurements in an emergency situation
- The prediction of water inundation using height data
- Classification using interpolated data (i.e., at a higher nominal density than the actual density of measured data set)
- Restriction of the area to be processed (Area of Interest) within the whole area of input data sets

Relations

- Related User Stories: 1, 5, 6, 7, 8, 9, 23, 24, 28, 33, 35, 36, 37, 38, 39, 40, 41, 46, 50, 51
- Related Scientific and Development topics:
  - 4.1 Spatial Data Processing Architectures and Interfaces
  - 4.2 Domain-Specific Languages (DSLs)
  - 5.1 Heterogeneous Data Integration
  - 5.2 Spatio-Temporal Data Fusion
  - 5.3 Feature Extraction, Classification and Correlation
  - 5.5 Change Detection and Dynamics
  - 6.1 Composite Visualization Methods
  - 6.2 Visualization techniques utilizing newest GPU features

Sample data

Sample data available:

Dataset/Service ID	Organization name	Dataset/service title
5	6 FOMI	Orthophotos
8	10 UBO	Digital Terrain Model (DTM)
9	10 UBO	Digital Terrain Model (DTM)
10	10 UBO	Point Cloud from TLS
11	10 UBO	Point Cloud from TLS
12	10 UBO	Digital Terrain Model (DTM)
13	10 UBO	Digital Terrain Model (DTM)
17	12 Liguria	Bathymetry ponits
18	12 Liguria	Bathymetry grid
20	12 Liguria	Regione Liguria Forestry map
21	12 Liguria	Regione Liguria Land Cover map
22	12 Liguria	LIDAR (XYZ)
23	12 Liguria	LIDAR (LAS)

11 VISUALIZATION

Requirements

- Interactive visualization of voxel-based 3D simulation results.
- Visualizing heterogeneous data in best available resolution in any location.
- Visualizing heterogeneous data (orthophoto, DTM, point clouds) acquired at different dates by swapping between the dates. Visual differences between the dates.
- Visualizing shaded DTM and interactively change shade parameters.
- Visualizing and exploring 3D models of a city, simulations (e.g., flood) and other environments.
- Interactively explore datasets with the visualization being a true (not approximate) representation of the data.
- Interactively modifying and creating derived data from the displayed dataset.
- Interactively exploring results of a flood simulation with photo-realistic effects in 3D with textures (from satellite images).
- Interactively exploring a multitude of uncorrelated datasets simultaneously (DTM, simulation results, cadastral data, etc.).

- Displaying data in stereo view.

Constraints

- Visualization must enable the view of large datasets.
- Visualization must support real-time updating.
- Operations must perform as fast as possible even with large datasets.
- Visualization must be web browser compatible and support several resolutions.

Relations

- Related User Stories: 1, 2, 6, 7, 9, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 33, 44
- Related Scientific and Development topics:
    - 4.1 Spatial Data Processing Architectures and Interfaces
    - 6.1 Composite Visualization Methods
    - 6.2 Visualization techniques utilizing newest GPU features
    - 6.3 3D-Web-based Visualization

Sample data

Sample data available:

Dataset/Service ID	Organization name	Dataset/service title
4	6 FOMI	Land Parcel database
5	6 FOMI	Orthophotos
6	6 FOMI	Digital Surface Model (DSM)
7	6 FOMI	Digital Elevation Model (DEM)
8	10 UBO	Digital Terrain Model (DTM)
9	10 UBO	Digital Terrain Model (DTM)
10	10 UBO	Point Cloud from TLS
11	10 UBO	Point Cloud from TLS
12	10 UBO	Digital Terrain Model (DTM)
13	10 UBO	Digital Terrain Model (DTM)
14	6 FOMI	Building boundary database
15	9 IGN	Paris 3D Model
17	12 Liguria	Bathymetry ponits
18	12 Liguria	Bathymetry grid
19	6 FOMI	Road Network
20	12 Liguria	Regione Liguria Forestry map
21	12 Liguria	Regione Liguria Land Cover map
22	12 Liguria	LIDAR (XYZ)
23	12 Liguria	LIDAR (LAS)

12 SIMULATION AND PREDICTION

Requirements

- 3D environmental and industrial disaster effect simulation on lands and cities including the calculation of risk areas.
- Flood simulation with photo-realistic effects in 3D including calculation of flood maps, impacted human and economic stakes.

- Environmental damage simulation in land parcels (natural: water, drought, hail, disease; damage caused by game/ animals; agrotechnical errors).
- Hydraulic model based simulations.
- Custom point-cloud based simulations.

Constraints

- Simulations must be interactive and can be parameterized.
- Generated simulation results must be savable, and exportable as spatial data.

Relations

- 4.1 Spatial Data Processing Architectures and Interfaces
- 4.2 Domain-Specific Languages (DSLs)
- 5.2 Spatio-Temporal Data Fusion
- 5.5 Change Detection and Dynamics
- 6.1 Composite Visualization Methods

Sample data

Sample data available:

Dataset/Service ID	Organization name	Dataset/service title
4	6 FOMI	Land Parcel database
5	6 FOMI	Orthophotos
9	10 UBO	Digital Terrain Model (DTM)
10	10 UBO	Point Cloud from TLS
11	10 UBO	Point Cloud from TLS
12	10 UBO	Digital Terrain Model (DTM)
13	10 UBO	Digital Terrain Model (DTM)
17	12 Liguria	Bathymetry ponits
18	12 Liguria	Bathymetry grid
21	12 Liguria	Regione Liguria Land Cover map
22	12 Liguria	LIDAR (XYZ)
23	12 Liguria	LIDAR (LAS)

13 CONCLUSIONS

This document has set the baseline and analyzed the state of the art concerning scientific and development aspects (top-down approach) as well as user aspects (bottom-up approach).

An initial set of user stories (54 stories) and sample data (23 datasets) have been collected, which have been analyzed and yielded user roles and high-level user aspects including application area data requirements.

As this preliminary analysis shows, these 54 stories have been collected in several application areas that cover most of the IQmulus project goals. The application areas are all related to multiple IQmulus project goals and the requirements that could already be extracted are mostly within the scope of the project.

At the moment, the user story collection process has just started, including mainly consortium-internal users in this first round. Therefore, the user aspects section does not represent a complete requirement collection. Additional user stories will be gathered from local stakeholders and a full analysis and user requirement specification will only commence in the next phases of the project. Other application areas may also emerge from these future user stories.

This document is the starting point. The inevitable and necessary duality of “developer” and “user” aspects is reflected both in its structure and in its content, and initial logical links are established. One of the important outcomes of the further project evolution will be a deeper mutual understanding of the two main areas through intensive dialogue, and a deepening and consolidation of requirements to be described in future Deliverables dealing with User Requirements (D1.2.1, D1.2.2, D1.2.3, D1.2.4).

ANNEX 1: USER STORIES

ID	Organization	User Story	List of related dataset/ service IDs	Comments (author)
1	10 UBO	As a scientist in a research institution (works also for a survey consultant), I want to import, visualize and extract rapidly a big dataset of LIDAR and orthophotos so that I can work with the data in my zone of interest (at least on a regional scale)		
2	10 UBO	As a scientist in a research institution, I want to merge a number of DTMs of different resolutions and display the best resolution available at any location so that I can work in a optimised manner		
3	10 UBO	As a surveyor in a research institution, I want to superposed a point cloud of LIDAR data on top of a DTM so that I can evaluate quality of my DTM compared to the original dataset		
4	10 UBO	As a data processor in a survey company I want to create a multi-resolution DTM from a heterogeneously spaced point cloud so that I use the data at its best resolution		
5	10 UBO	As a GIS expert in a research institution I want to identify areas of rocks, vegetation, constructions and sand on a Terrestrial Laser Scanner dataset so that I can classify the different zones of the coastline		
6	10 UBO	As a scientist in a survey consultant company I want to visualize a couple of datasets (orthophotos, DTMs or point clouds) acquired at different dates so that I can assess rapidly the evolution of topography in the area		
7	10 UBO	As a data processing expert in a Risk Management Agency I want to compute and visualize differences between 2 datasets of point clouds (acquired at different dates) so that I can identify rock slides on a cliff and know the associated volumes of debris		
8	10 UBO	As a sedimentologist in a Hydrographic Service I want to extract submarine dune shapes from a DTM so that I can see the evolution of dunes along time		
9	10 UBO	As a geologist in a research institution I want to visualize a shaded DTM and interactively change shade parameters so that I can detect rapidly rifts and other geological structures		
10	10 UBO	As an expert in a survey consultant company I want to compute displacement maps from DTMs acquired at multiple dates so that I can detect and quantify displacements		
11	10 UBO	As a GIS expert in a Coastal Observatory I want to realize automatic extraction of coastline from a LIDAR DTM so that I can monitor the coastal zone		
12	10 UBO	As a surveyor in a research institution I want to assess the quality of a new acquired dataset (point cloud or DTM) in comparison to reference datasets (existing and validated ones) so that I can see if the data acquisition and processing were okAs a surveyor in a research institution		
13	1 SINTEF	As an expert user, I would like to interactively explore the results of a flood simulation with photo-realistic effects. I start iQumulus, and I will be able to view the DEM in 3D with sattelite textures and the simulated flood.	16	Technical user story. Key concepts: * 3D visualization of DEMs * Interactivity in data exploration * Photorealistic effects Datasets: * 3d_flood_visualization.wmv * photorealistic_effects1.avi * photorealistic_effects2.mov
14	1 SINTEF	As an iQumulus user I would like to interactively explore the dataset, knowing that the visualization I am presented with is not an approximation, but a true representation of the underlying data-set.	16	Technical user story. Key concepts: * Guaranteed precision on visualization * Interactivity in data exploration * Exact representation of vector/cadastral data Datasets: * guaranteed_precision.mov

ID	Organization	User Story	List of related dataset/ service IDs	Comments (author)
15	1 SINTEF	As an iQumulus user I would like to view voxel-based 3D simulation and processing results visualized as iso-surfaces interactively.	16	Technical user story. Key concepts: * Visualization of 3D voxel-based data sets (Voxels are like sugar cubes, and stacked together in e.g., a 512x512x512 grid, this is non-trivial to visualize) Datasets: * 3d_datasets.avi * 3d_voxels1.avi * 3d_voxels2.avi
16	1 SINTEF	As an iQumulus user I would like to explore point-cloud simulation and processing results both as points (the original underlying data), and as a surface or volume.	16	Technical user story. Key concepts: * Interactive visualization of 3D point-based data sets. * Point clouds may represent volumes or surfaces. Datasets: * 3d_datasets.avi * 3d_point_cloud.avi
17	1 SINTEF	As an iQumulus user I want to visualize 3D simulation results interactively as volume data.	16	Technical user story. Key concepts: * Ray-casting * Alpha Blending (transparency) * Interactivity Datasets: * volume_visualization.mpeg
18	1 SINTEF	As an iQumulus user I want to interactively explore data that are large, and not only limited to a small spatial/temporal extent.	16	Technical user story. Key concepts: * Large cityscape scenes ("out-of-core") * Large datasets ("out-of-core") * Multi-resolution Dataset: * large_scene.mov * large_datasets.mov
19	1 SINTEF	As an iQumulus user I want to interact with the data-set being visualized, for example by modifying it interactively, or displaying derived data.	16	Technical user story. Key concepts: * Data interaction * Derived data generation (i.e., time-series of water elevations for a single point, etc.) Datasets: * data_interaction.mp4
20	1 SINTEF	As an iQumulus user I want to see a multitude of uncorrelated data sets simultaneously: raw DEMs, simulation results, cadastral data, etc. I want to interactively explore these datasets to gain new insight.	16	Technical user story. Key concepts: * Real-time data fusion / data compositing Datasets: * data_fusion.pdf
21	6 FOMI	As a GIS expert I want to access remote sensing data via WMS (Web Map Service) automatically displayed in stereo view in order to support the interpretation task.	5	
22	6 FOMI	As a RS expert I want to compare the automatically classified building contours based on orthophoto, Digital Elevation Model and Digital Surface Model with the building boundaries in the cadastral database, so that I can verify the adequacy of the cadastral records.	4,5,6,7,14	
23	6 FOMI	As a GIS expert I want to create and visualize a 3D city model by the automatic classification of LiDAR data so that I can support and accelerate the disaster decision making process (e.g.: flood modeling).	6,7,14	
24	6 FOMI	As a GIS expert I want to delineate slopes steeper than a given threshold (with DDM, without LiDAR, with LiDAR, etc.) so that I can support the definition of erosion risk areas.	6,7	
25	6 FOMI	As a GIS expert I want to produce a Digital Surface Model from 3D point cloud, which is appropriate to generate true orthophotos so that I can support the area measurements during the building detection.	5,6,14	
26	6 FOMI	As an agricultural expert I want to generate NDVI maps in a particular area and time period from different source of remotely sensed data, so that I can detect changes in vegetation conditions.	Later will be provided	
27	6 FOMI	As an agricultural expert, I want to monitor the agriculture-related characteristics (plant species, breed, biological condition) of a land parcel to support the planning of agrotechnical interventions.	4, more to be provided	

ID	Organization	User Story	List of related dataset/ service IDs	Comments (author)
28	6 FOMI	As a field expert in disaster management, I want to perform real-time updating of road status attributes in an emergency situation (flooding or high-level inland excess water) on my mobile device so that I can make route planning by takinf into account the flooded areas or other barriers.	5,19	
29	6 FOMI	As a GIS expert I want to update a vector dataset (e.g. road network) in a particular area by using newly available vector data arising from external sources, so that I can provide up-to-date information.	19	
30	6 FOMI	As a GIS expert I want to generate a categorized waterlog map (with at least the following categories: open water surface, flooded vegetation, wet soil) in a particular area in order to provide data for a quicker and more reliable agricultural damage assessment and yield loss estimation.	4,5,6, more to be provided	
31	6 FOMI	As a Remote Sensing expert I want to know if a particular land parcel was harvested/mowed, and if so, when did it exactly happen, so that I can reliably and quickly justify the intervention.	4, more to be provided	
32	6 FOMI	As a Remote Sensing expert I want to receive answers about the damage of a land parcel (natural: water, drought, hail, disease; damage caused by game/ animals; agrotechnical errors), in order to be able to estimate the scope and the extent of the damage.	4, more to be provided	
33	6 FOMI	As a GIS expert, I want to delineate the extent and the maximum possible extent of the area potentially flooded in case of an industrial disaster (e.g. dam burst of a waste reservoir) so that I can provide the polygons of the risk area to the disaster management authorities.	4, more to be provided	
34	6 FOMI	As a GIS expert I want to carry out partial updating of a vector dataset having predefined topological ruleset using external data sources, by respecting the topological rules, in order to obtain a complete and updated database.	Later will be provided	
35	6 FOMI	As an agricultural expert, I want to determine crop cover for crop parcels of a given area using color composites of multispectral satellite images taken at different dates for documenting the status at given dates and to consider the justness of agricultural subsidies for this area.	4, more to be provided	
36	6 FOMI	As an agricultural expert, I would like to delineate stable, steady limits in a given area and land cover categories within these limits using orthophotos, color composites of multispectral satellite images taken at different dates and digital elevation model in order to keep the reference geographical information system of agricultural subsidies up to date.	5,6,7, more to be provided	
39	11 Ifremer	As an expert seabed mapper, I would like to process large quantities of Lidar waveforms to retrieve several predefined form attributes (centroid, Peak, Front slope angle, Ground return ratio, ….) and classify with thresholds or statistically.		
40	11 Ifremer	As an expert seabed mapper, I would like to interpolate large quantities of point cloud data (soundings, wave form attributes, …) to nominal density or specific resolution		
41	11 Ifremer	As an expert seabed mapper, I would like to ensure coregistration (planimetric and altimetric) of several data (aerial photograph, hyperspectral image, DTM) in order to perform layer stacking.		
42	11 Ifremer	As an expert seabed mapper, I would like to perform unsupervised or supervised classifications using all selected attributes on a specified area defined by a polygon.		
43	11 Ifremer	As an expert seabed mapper or coastal manager, I would like to make a raster to vector conversion of a classified layer and be able to annotate polygons.		

ID	Organization	User Story	List of related dataset/ service IDs	Comments (author)
44	11 Ifremer	As an expert seabed mapper or coastal manager, I would like a.to extract values from one to several raster layers at a set of given locations and save them in a spreadsheet (e.g. change in seaweed cover throughout time) b.to statistically summerise values from either a raster or a point cloud within a given polygon (e.g. total surface of habitat type within designated area)		
45	11 Ifremer	As an expert coastal manager, I would like drape the imagery onto the DTM to highlight specific features and display it in 3D from various perspectives.		
46	9 IGN	As a GIS expert in a data production institute, I want to combine 2D/3D building polygons from a topographic DB with LiDAR data in order to extract building elevation so that I can enhance geometric description of buildings in my DB.		
47	9 IGN	As a GIS expert in a data production institute, I want to register topographic LiDAR data and bathymetric LiDAR data acquired at different times so that I can perform sea/land classification.		
48	9 IGN	As a GIS expert in a data production institute, I want to perform accurate sea/land classification of combined LiDAR datasets so that I can produce a continuous sea/land DTM.		
49	9 IGN	As a GIS expert in a data production institute, I want to merge altimetric datasets from various sensors at different resolution and acquired at different times so that I can produce a continuous DTM at an optimized resolution and accuracy.		
50	9 IGN	As a GIS expert in a data production institute, I want to detect evolutions of a landscape by combining topographic database and a newly acquired LiDAR dataset so that I can improve DB updating process.		
51	9 IGN	As a forest expert, I want to extract information from LiDAR datasets, such as vertical distribution of wood/leaves, tree position, height, size or canopy diameter so that I can populate a forest inventory DB.		
52	9 IGN	As a hydrologist, I want to extract significant breaklines from high resolution DTMs or point clouds, such as dikes, dams, banks, ditches so that I can improve my hydraulic model and make more accurate simulations.		
53	9 IGN	As a hydrologist, I want to combine flood simulation model with high resolution DTM so that I can publish a water height map on the flooded area.		
54	9 IGN	As a hydrologist, I want to combine flood simulation models with all kind of topographic data in order to know the flooded area and all the human and economic stakes impacted so that I can provided decision makers with accurate and reliable informations.		

ANNEX 2: SAMPLE DATA TABLE

Dataset/ Service ID	Organization name	Dataset/service title	Dataset/service short description	Country(/ies)	Date	Spatial resolution: Resolution distance (m)	Sensor type Platform category	Data characteristics: Dimensionality	Data characteristics:File storage type	Data characteristics :File format	List of related User Story IDs
4	6 FOMI	Land Parcel database	Cadastral map or land registration map it contains cadastral parcels.	Hungary, Eger	2. okt. 2012	0.01	NA	2D	Relational spatial database	ESRI shp	22,24,27,28, 29,30,31,32, 33,34,35,36
5	6 FOMI	Orthophotos	Multichannel Imagery (RGB+NIR) Ortophotos produced from aerial photographs taken 2000, 2005 2007, 2008, 2009, 2010, 2011, 2012 with native resolution of 0,5m/px.	Hungary, Eger	15 júl. 2010	0.5	Airborne	2.5D (2D + height/elevation)	Binary (all kinds of file formats)	geotiff	21,22,27,28, 29,34,36
6	6 FOMI	Digital Surface Model (DSM)	Represents the elevation of the earth's surface and includes all objects on it.	Hungary, Eger	18 okt. 2012	0.4	Airborne	3D	Binary (all kinds of file formats)	las	23,25,33,34
7	6 FOMI	Digital Elevation Model (DEM)	Represents the elevation of the ground surface.	Hungary, Eger		5	NA	3D	Binary (all kinds of file formats)	grid	24,27,30,32, 33,34
8	10 UBO	Digital Terrain Model (DTM)	Bathymetry of a small pocket beach in Brittany (Porsmilin)	France	19 ápr. 2011	1	Mobile (vehicle)	2.5D (2D + height/elevation)	ASCII (column/tabular)	asc	
9	10 UBO	Digital Terrain Model (DTM)	Bathymetry of a small pocket beach in Brittany (Porsmilin)	France	27 máj. 2010	1	Mobile (vehicle)	2.5D (2D + height/elevation)	ASCII (column/tabular)	asc	
10	10 UBO	Point Cloud from TLS	Topography of a small pocket beach in Brittany (Porsmilin)	France	3 jún. 2010		Static	3D	ASCII (column/tabular)	xyz	
11	10 UBO	Point Cloud from TLS	Topography of a small pocket beach in Brittany (Porsmilin)	France	16 dec. 2009		Static	3D	ASCII (column/tabular)	xyz	
12	10 UBO	Digital Terrain Model (DTM)	Bathymetry of submarine sand dunes in Brittany (Four)	France		2	Mobile (vehicle)	2.5D (2D + height/elevation)	ASCII (column/tabular)	asc	
13	10 UBO	Digital Terrain Model (DTM)	Bathymetry of submarine sand dunes in Brittany (Four)	France	1 szept. 2011	2	Mobile (vehicle)	2.5D (2D + height/elevation)	ASCII (column/tabular)	asc	
14	6 FOMI	Building boundary database	Contains the boundary of the private buildings, public buildings, commercial buildings, agricultural buildings, transportation buildings and other kind of bulidings.	Hungary, Eger	1 jan. 2000	0.01	NA	2D	Binary (all kinds of file formats)	ESRI shp	22,24,27,28, 29,30,31,32, 33,34,35,36
15	9 IGN	Paris 3D Model	3D model made from 10cm aerial images on the city of Paris	France	20 júl. 2011	0.1	Airborne	3D	Structured data modeling / data interchange languages (e.g. GML, KML)	cityGML	

Dataset/ Service ID	Organization name	Dataset/service title	Dataset/service short description	Country(/ies)	Date	Spatial resolution: Resolution distance (m)	Sensor type Platform category	Data characteristics: Dimensionality	Data characteristics:File storage type	Data characteristics :File format	List of related User Story IDs
16	1 SINTEF	SINTEF Visualization	This dataset contains technical user stories together with visualization demos. This is not a data-set per se, but it contains visualization demos that can be used to inspire users for what the visualization can offer				NA	NA	Binary (all kinds of file formats)		
17	12 Liguria	Bathymetry ponits	this dataset contains multibeam points for the study area of Riva Trigoso, Liguria (italy)				Mobile (vehicle)	3D	NA (no file uploaded - I provide link to a service)	txt	
18	12 Liguria	Bathymetry grid	this dataset contains a depth grid obtained by multibeam points for the study area of Riva Trigoso, Liguria (italy)				Mobile (vehicle)	3D	NA (no file uploaded - I provide link to a service)		
19	6 FOMI	Road Network	The dataset contains the road network for a particular area.	Hungary	2000-		NA	2D	Binary (all kinds of file formats)	ESRI shp	28,29
20	12 Liguria	Regione Liguria Forestry map	this dataset contains the Regione Liguria Forestry map at scale 1:25000. It was carried out by remote sensing using satellite images (2004-2007).	Italy	1. jan. 2008		NA	2D	Relational spatial database	.shp, .mdb (Geomedia Access)	
21	12 Liguria	Regione Liguria Land Cover map	this dataset contains the Regione Liguria Land Cover map at scale 1:10000. It was carried out by remote sensing using aerial orthophotos (2010) and satellite images (2004-2009).	Italy	31. dec. 2012		NA	2D	NA (no file uploaded - I provide link to a service)	.shp, .mdb (Geomedia Access)	
22	12 Liguria	LIDAR (XYZ)	this dataset contains LIDAR dataset for the study area of Vernazza, Liguria (italy)				NA	3D	NA (no file uploaded - I provide link to a service)		
23	12 Liguria	LIDAR (LAS)	this dataset contains LIDAR dataset (.LAS) for the study area of Vernazza, Liguria (italy)				NA	3D	NA (no file uploaded - I provide link to a service)		