AMENDED/UPDATED DEVELOPMENT GUIDELINES FOR DATA INTEGRATION AND PROCESSING - VERSION 3

D4.1.3 Ammendment following M24 review

Circulation:	RE - Restricted
Lead partner:	CNR-IMATI-GE
Contributing partners:	CNR-IMATI-GE, UCL, SINTEF, TUDelft, UBO, IGN, Fraunhofer
Authors:	Giuseppe Patanè & Michela Spagnuolo (Editors, CNR-IMATI-GE), Vibeke Skytt (SINTEF), Jan Boehm (UCL), Roderik Lindenbergh (TUDelft), Andrè Stumpf (UBO), Mathieu Bredif (IGN), Michel Krämer (Fraunhofer)
Quality Controllers:	Michel Krämer (Fraunhofer), Ewald Quak (SINTEF)
Version:	4.0 (Amendment/Update to D4.1.3)
Date:	~~01.03.2015~~01.03.2015

DOCUMENT HISTORY

Version[1]	Issue Date	Stage	Content and Changes
3.0	31 October 2014		D4.1.3 Deliverable submitted to the PO
3.1.0	15 December 2014		Preparation of preliminary material
3.1.1	12 January 2015		Update of preliminary material
3.2	2-4 February 2015		Update of D4.1.3 discussed at IQmulus meeting
3.2.1	13 February 2015		Update of D4.1.3 with Partners' material
3.2.2	15 February 2015		Update of D4.1.3 at in a TELCO
3.3	18 February 2015		Update of D4.1.3 distributed to Partners
3.4	23 February 2015		Revision of D4.1.3
3.5	24 February 20142015		Update of D4.1.3 at in a TELCO
3.6	27 February 2015		Update of D4.1.3 at in a TELCO
4.0	27 February 01 March 2015		Updated version to be submitted to the PO

[1] Integers correspond to submitted versions

CONTENTS

formázott: Bekezdés alapbetűtípusa, Nyelvhelyesség ellenőrzése

formázott: Bekezdés alapbetűtípusa, Nyelvhelyesség ellenőrzése

INTRODUCTION

This document contains the decisions made by the Consortium in reply to the rec~~c~~ommendations received during the second year project review, concerning in particular the following:

(excerpt from the Technical Review Report - Ref. Ares(2015)463495 - 04/02/2015)

As at the previous review, the partners are strongly encouraged to focus on big data aspects during the platform development, showcase definition, and data testing. For at least some well-chosen proposed services, one or more big data aspects (e.g., volume and variety) should be supported consistently with what is envisioned in the Description of Work. The advantage of using IQmulus over other solutions (or as-is solutions) should be clearly stated and assessed, in terms of aspects such as performance, scalability, and accessibility. An updated version of D4.1.3 due by 01/03/2015 should contain this information:

- *Which specific services will address at least one big data aspect, and the specific aspects to be addressed for such service;*
- *For each service and big data aspect to be addressed, the solutions to be pursued to improve with respect to the as-is status; indicate also the targeted improvement and how the improvement is to be assessed and validated;*
- *A time plan for these activities;*
- *A description of how the IQmulus system will excel over competing systems in terms of its support for big data*

The review report has been, once again, a source of important suggestions that helped the Consortium to get a better perspective on the ongoing activities and crucial objectives to aim ~~to~~ at in the next project phase.

First of all, we would like to mention that some of the issues raised above w~~h~~ere already ~~matter~~ topics of discussion and actions in WP4: previous versions of the ~~D4~~task 4.1 deliverables clearly indicated the guidelines adopted for the selection, development, and release of WP4 processing services, and their inclusion in the IQmulus Functional Prototype (see D4.1.2 and D4.1.3). Most importantly, the release of toolkits entails ~~that,~~ that for each service:

- **proper identifications and descriptions of performance criteria** ~~is~~ are ~~made~~ provided (see Annex B 'Service testing within WP4' of D4.1.3), as mentioned in the technical annex and as reported in the service metadata table;
- **proper motivations for software and libraries** selected for implementation are given;
- **clear identification of example input/output data sets** is provided, chosen among IQmulus data and selected on the basis of the challenges addressed by the ~~Project~~project, in particular, service metadata should contain information on (i) volume; (ii) heterogeneity; (iii) processing speed[2];
- a **description of the data partitioning strategy**, if applicable, and the computing scenario is given;
- the **significance and quality** of the service metadata is ensured by the task leader.

All t~~T~~ask leaders, with the support of the WP4 leader and the Quality Assurance Coordinator, are committed to reinforce the quality check of the toolkits with respect to the issues above.

[2] *see the minutes of the technical meetings in Darmstadt, February 2014~~,~~; Delft, March 2014~~,~~; Brest, September 2014*

Concerning the suggestion to ~~focus~~ better focus the efforts in service development, we ~~remind~~ approval ~~to~~ of service development was always guided by the significance and importance of the service with respect to the showcases described in deliverable D1.2.3. These guidelines also apply to the third and fourth project years and, as such, are in line with the requests of the Project Officer and the Reviewers as mentioned in the review report.

In addition to that, and in order to provide ~~a~~ specific feed~~-~~back to the review report and its recommendations, the Partners involved in WP4, ~~with the~~ coordinat~~ion~~ed ~~of~~ by the WP4 t~~T~~ask contributed to the collection of the following information for each service:

- a short statement characterizing the service with respect to big data: ~~(~~does the service address big data? ~~I~~if yes, in what sense: ~~(~~e.g., volume vs. variety~~)~~;
- the definition of the test suite, with an indication of the data size of each ~~composing~~ to be used for the scalability tests; these datasets will be uploaded to the HDFS-Hadoop Distributed File System;
- a reference to the corresponding showcase workflows (for prioritization);
- references to data sets, available at the time of writing, to be used as test data;
- references to the scientific literature and/or state-of-the-art methods (if available), in addition to what is reported in the toolkit reports already).

The information mentioned above, discussed and integrated during face-to-face meetings (Paris, 2-4 February, 2015) and several TELCOs (for details, see the document history), have beed used to produce this document.

The document is organized as follows. In Section 2, we re-state briefly the specific goals and roles of processing services with respect to the overall ~~project~~ IQmulus project, taking into account the feedback we ~~got~~ received during the review. In Section 3, we describe the rationale we adopted to select services that are expected to demonstrate the capability to process *big data*, in at least one of its aspects. In Section 4, we discuss~~,~~ and propose~~,~~ ~~a~~ concrete ~~effort~~ actions towards a to address *tiling&stitching*, as the proposed overall IQmulus ~~proposed~~ solution to address big the main conclusion of the discussions we had on this topic after the review meeting. Finally, in Section 5, we present the timeline that was adopted to achieve the expected results, as an integrated pipeline involving not only service development~~m~~ but also scalability tests, validation, and integration. In the Annex, we include the tables summarizing the discussion and analysis of each service~~s~~ developed~~,~~ or planned so far.

THE ROLE OF THE PROCESSING SERVICES IN IQMULUS

Therefore, IQmulus is targeting to enable optimized use of large, heterogeneous geo-spatial data sets for better decision making through a high-volume fusion and analysis information management platform. This platform will transpose approaches and IT standards from distributed computing to enable distributed, service-oriented geospatial processing. We will determine optimal execution and distribution parameters for different geospatial processing tasks and to ensure that the IQmulus system can transparently execute processing on different architectures like GPGPU clusters or clouds. Methods will be developed to connect processing and visualization into a tight loop, ensuring high interactivity in the process to enable users to better understand correlations between heterogeneous data sets.

FIGURE 1: FROM THE DOW - IQMULUS ABSTRACT

The text above summarizes the planned goals of IQmulus. The aim is therefore to demonstrate how ***service-oriented*** geospatial processing can be exposed to **distributed computing**, as an

approch to cope with **big data** in the geospatial context, where data volume and heterogeneity are concrete problems.

The role of WP4 processing services is therefore to demonstrate how processing services can be turned into service-oriented processing to be integrated in a distributed computing infrastructure. We ~~remark~~ note the importance of a well-designed testing phase for services as a "*determine* ***optimal execution and distribution parameters*** *for different geospatial processing tasks and to ensure that the IQmulus system can* ***transparently execute processing on different architectures.***"

The capability to process big data is, therefore, not only the responsibility of the single processing services, rather~~,~~ it is the result of the integration of several components, the most important being the computing infrastructure, with its JobManager engine which injects *intelligence* in the workflow execution adopting the best strategy to distribute the computation and/or the data among the available computing nodes.

As suggested during the review, however, a more structured approach to scalability testing, and performance evaluation in general, is needed in this important project phase. This effort will strongly reinforce the original goals of WP4.

Concerning scalability, we would like to remark that the results of the scalability tests should be analyzed with respect to *data sizes that match the sizes that are considered critical in the use cases* ~~-~~ – ~~no need to~~without ~~-~~articificially inflat~~e~~inge the data volume if there is no need for that. regard has been duly documented in the D1.2.3 deliverable report, where representative datasets for each showcase have been listed. Note that all the datasets mentioned there will be used as reference datasets for the analysis of the scala~~n=~~bility results, at service and at workflow level.

Also, scalability tests will provide an overview over the services~~,~~ that correlates execution time versus input data size, given a specific setting of the underlying computing infrastructure. Thi~~e~~is~~e~~ analysis will be useful to identify services that might be candidates for being handled in a kind of "batch execution mode" (typically, pre-processing services, requiring a long execution time, but not critical for the context of usage) ~~with respect~~as opposed to services that need a quicker instance, analysis of streams of environmental data). Finally, as pointed out by the Reviewers already, it is not expected that every single service in WP4 should handle big data: there are ~~cases cases~~ settings that do not ~~point~~ relate to scalability issues at all, rather, they focus on smarter heterogeneous data or smarter classification methods.

Beside scalability, ~~Task~~ the task leaders and developers are committed to thoroughly checking the quality of the service metadata tables and asking for revisions ~~for~~ of any metadata that do not properly characterize important performance aspects of services, as deemed necessary for a serious validation of the results. Revisions of the metadata tables, with the inclusion~~s~~ of metrics evaluate performance criteria will be made available in the next toolkit releases. As described in D4.1.2 and D4.1.3, the performance metrics are expected to go beyond textual descriptions and will be released as executables, or shell scripts, that implement a quantitative evaluation of the property declared in the metadata tables (e.g., executables that compare the results obtained with respect to a ground truth, or ~~-~~shell scripts that provide a comparison of the results obtained with small variations/perturbations of the input data).

PROCESSING SERVICES AND "BIG DATA": THE THIRD YEAR PERSPECTIVE

This section contains the Consortium's reply ~~to~~ concerning the selection of the set of services that will challenge IQmulus to demonstrate concrete advancements towards its goals, that is, a set of services that, by running in the IQmulus distributed computing environment, will make it possible to handle big geospatial data. In the following, we provide a first set of selected services that exhibit particularly innovative aspects with respect to their approach to big data processing and that will be tested on big data in the third project year.

1.1 ~~WHAT~~ Services

(Q1) Which specific services will address at least one big data aspect, and the specific aspects to be addressed for such service?

To address this question, we have analysed WP4 services, either already released or to be released soon, and we have identified a few ones that exhibit particularly innovative aspects with respect to their approach to big data processing, or services that address crucial processing tasks with the potential to enable a quick uptake of large point clouds, images and collections thereof in applications. The selected services are listed according to the t~~T~~ask/toolkit in which they are developed.

Task 4.2 (PM 36)

formázott: Címsor 3, Balra zárt, Nincs felsorolás vagy számozás

- ***#70 Radiometric Enhancement** (FOMI) – Aspect: Size*

For the Year 3 toolkit this service ~~is~~ will be implemented in~~ing~~ a parallel Hadoop version will improve the currently available single node Year 2 version. This service will be applied on large collections of satellite imagery that require processing at FOMI. Within IQmulus this service is part of the Land Showcase 3.

- ~~***#8 Point cloud to point cloud distance** (TUDelft) – Aspect: Size*~~
- ***#98 Point cloud intersection** (TUDelft) – Aspect: Size*

formázott: Behúzás: Bal: 1,27 cm, Függő: 0,63 cm, Térköz Utána: 6 pt

The reason for focusing on services #8 and #98 is the generality of the operations they implement, which can have a large impact on a number of other subsequent operations or applications. In addition, for ~~Services~~ services #8 and #98 an additional tiling and stitching is required to enable the services to properly work on data sets consisting of many files.

Task 4.3 (PM 36)

- ***#94** **Confidential -** ~~**Vector layer partitioning of a point cloud** (IMATI) – Aspect: Size~~*
- ***#59 Multi-object classification of 3D point clouds** (IGN) – Aspect: Size*

This service faces an even larger dataset than the raw input point cloud, as each point record has been expanded with a number of local features to be used by the classification process. This service is central to the ~~US2~~ urban showcase US2 and is the basis of multiple recent publications:

- M. Weinmann, S. Urban, S. Hinz, B. Jutzi, C. Mallet. *Distinctive 2D and 3D for a~~A~~utomated l~~L~~arge-s~~S~~cale s~~S~~cene a~~A~~nalysis in u~~U~~rban a~~A~~reas*. Computer & 2015.
- M. Weinmann, B. Jutzi, S. Hinz, C. Mallet. *Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers*. ISPRS Journal of Photogrammetry and Remote Sensing, 2015.
- M. Weinmann, B. ~~JUtzi~~Jutzi, C. Mallet. *Semantic 3D scene interpretation: a combining optimal neighborhood size selection with relevant features*. ISPRS

Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3, Zürich, Switzerland, September 2014. (Best Paper Award)

This service is implemented on Apache Spark. It uses an in-memory distributed Random Forest classifier in a Hadoop cluster. This enables us to handle arbitrary sizes of input data by scaling the cluster size. A prototype implementation of this approach was succesfully demonstrated at the 2nd year review meeting.

Task 4.4 (PM 30)

- ***#9 LR B-spline approximation from point cloud*** *(SINTEF) – Aspect: Size*

An LR B-spline approximation of a point cloud is an adaptive and shape sensitive representation that is able to produce a surface that has an overall lean representation with more information in areas with high variation in the input data. Used on the types of terrains where the format is well suited we achieve a very good data reduction effect while maintaining the important information carried by the data. The adaptivity differs from the raster format which is the state-of-the art representation for this purpose, but which is a completely uniform format.

The current implementation is single node, but to take advaentage of the representation properties the approximation service and related services must handle large data sets. There is ongoing work with tailored tiling of point clouds and stitching of the surfaces approximating the tiles to get a regular surface model that is well suited for maneuvering. The computations will utilize distributed computing in the cloud.

- ***#48 Confidential – see Appendix***
  Multi-resolution for land monitoring *(IMATI) – Aspect: Size*
- ***#40 Approximation of rainfall data*** *(IMATI)- Aspect: Heterogeneity*

This service is demonstrating the use of heterogeneous sources of rain measures and the delivery to the users of enhanced models of the rain field, showing the source sites and of the different sources. Heterogeneity and speed of computation are the challenges here, having in mind that rain data may come in an almost continous stream of data.

This service implements a pre-processing step that selects the optimal local scale of analysis of the point cloud based on the significance of the ACP analysis of each point's nearest neighbors. The challenge here is to optimize its computation in terms of speed, memory consumption and possible boundary artefacts at the ~~limit~~ borders of tiles.

- ***#110 3D Delaunay triangulation*** *(IGN) – Aspect: Size*

This new service is a tetrahedral decomposition of space based on the input point cloud. The challenge is to scale out this processing so as to enable its computation on datasets of an arbitrary size in reason~~n~~able time.

- ***#95 Watertight surface reconstruction*** *(IGN) – Aspect: Size*

MMS, terrestrial and aerial volumetric point clouds. Handling the heterogeneity of the sampling rates and the measurement accuracy is a key challenge here. Being based on #110, the goal is to be able to scale to volumetric point clouds of arbitrary sizes.

Task 4.5 (PM 30)

- ***#45 Topological change detection*** *(IMATI) – Aspect: Size/Heterogeneity/Processing Speed*

This service is demonstrating the analysis of an almost continous stream of data, with improved analytic capabilities for rain field data.

- ***#60 2D Displacement measurement using DSM/DTM/DEM and optical images*** *(UBO) – Aspect: Size/Processing Speed*

The service is already exploiting multi-threading and has been demonstrated to excel over alternative solutions frequently used in the geoscience community in terms of robustness and scalability (see D4.5.1). It enables the analysis of multi-temporal sea-bottom surface models (Marine showcase MS4 as presented at the 2nd review meeting ~~in November 2014~~) images (Land showcase LS4) and is therefore addressing important user requirements. The targeted distribution of the service among multiple nodes with the J~~j~~ob M~~m~~anager show the enhanced scalability of ~~a~~ multi-threaded algorithms when integrated in the IQmulus infrastructure.

For more details on these services, we refer the reader to the t~~T~~oolkits of the WP4 tasks 4.2-4.5 in which some of them were already described. In the Annex, the complete list of comments on all the other services is provided. Obviously, the majority of the services released so far will be chained in workflows, and their usage demonstrated in showcases. Since big data is an issue of every showcase, in principle, every service will be therefore exposed to big data and should find its solution to address this issue.

1.2 WHAT Approaches

(Q2) For each service and big data aspect to be addressed, the solutions to be pursued to improve with respect to the as-is status; indicate also the targeted improvement and how the improvement is to be assessed and validated.

To address this question, for each selected service, a discussion took place on the suitability of the current service implementation to handle big data, and on the most suitable appropriate approach improve the existing implementation in the service revision plan. The approach to service implementation is obviously bound to the current architectural choices of IQmulus). The viable options therefore are:

- ***Ddistributed execution.:*** *Ggiven the characteristics of the computing hardware, the Job Manager splits the input data and to distributes the chunks to the various processing Also, the Job Manager is responsible of spawning the single-core processes and to distribute them to the different nodes in a way that best utilises the possibilities of the available hardware.*

  KEY ISSUES: the Job_Manager has to know enough of the service behaviour to be able to handle the distribution properly; service metadata are important channels of information. ACTION REQUIRED: interaction with WP3 for service metadata revision.

- ***Multi-core/GPU algorithms****. These processes should be treated like single-core algorithms, with the difference that the Job Manager has to keep track of used CPU/GPU cores in order to decide how many processes can run on one node at the same time.*

- ***Distributed algorithms****. Such processes can be implemented with frameworks for distributed computing such as agent-based frameworks (Akka, etc.) or MPI. The Job Manager has to keep track on which nodes these processes run.*

  KEY ISSUES: the service itself manages the distribution of the computation..

- ***Hadoop jobs****. The Job Manager forwards jobs (MapReduce, Spark,...)Spark ...) to Hadoop. responsible for executing the jobs and for distributing work to the different nodes in the cloud. The Job Manager configures the infrastructure (i.e., spawns new nodes if necessary) and oversees the job execution.*

Concerning the services selected so far, they will implement the following strategy:

Task 4.2

- ***#70 Radiometric Enhancement*** *(FOMI) – Solution: Hadoop/MapReduce*
- ***#8 Point cloud to point cloud distance*** *(TUDelft) – Solution: Tiling&Stitching*
- ***#70 Radiometric Enhancement*** *(FOMI) – Solution: Hadoop/MapReduce*
- ***#98 Point cloud intersection*** *(TUDelft) – Solution: Tiling&Stitching*

Task 4.3

- ***#94 Confidential - Vector layer partitioning of a point cloud*** *(IMATI) – Solution:*
- ***#59 Multi-object classification of 3D point clouds*** *(IGN) – Solution: Parallelization through decision tree*

[A1] megjegyzést írt: Michel K.
please check and comment

formázott: Behúzás: Bal: 1,27 cm, Függő: 0,63 cm, Térköz Utána: 6 pt, Az azonos stílusú bekezdések között van térköz

formázott: Betűtípus: Félkövér

formázott: Betűtípus: Félkövér

- ***#27 Tree crown recognition from mobile mapping point clouds*** *(UCL) – Solution: Amazon EC2 cloud computing and a parallelized version of Random Forrest classification.*

- *Task 4.4*

  - ***#9 LR B-spline approximation from point cloud*** *(SINTEF) – Solution: Tiling&Stiching*
  - ***#48 confidential – see Appendix~~Multi-resolution triangulation for land~~***
  - ***#40 Approximation of rainfall data*** *(IMATI) - Solution: Tiling&Stitching*
  - ***#110 3D Delaunay triangulation*** *(IGN) – Solution: MapR~~r~~educe, Tiling&Stitching*
  - ***#95 Watertight surface reconstruction*** *(IGN) – Solution: Sub modular function optimization with cloud implementation*
  - ***#66 Point cloud dimensionality*** *(IGN) – Solution: Tiling&Stitching*

- *Task 4.5 ~~(Leader: UBO)~~*

  - ***#45 Topological change detection*** *(IMATI) – Solution: D~~d~~istributed computation*
  - ***#60 2D Displacement measurement using DSM/DTM/DEM and optical images*** *(UBO) – Solution: Distribution by the J~~j~~ob M~~m~~anager*

Considering the whole list of services released ~~so far (~~by PM24~~)~~, most of the executables are with only few exceptions~~.,~~ I~~i~~t is expected that most of the effort in the third project year will be devoted to re-implement the services to make them fit to one of the schemas above.

The results of the scalability tests will provide ~~a~~ crucial insights ~~in~~ for the best approaches towards the re-implementation of the services, when needed. Preliminary scalability test results will be made available early enough to guarantee corrective actions and to indicate the most appropriate directions for development.

As far as geo~~-~~spatial data processing is concerned, it is quite difficult to compare the planned approaches to handle big data to state-of-the-art techniques. Unless the parallelization is straightforward (e.g., embarassingly parallel problems) there are just a few algorithms that address big data in the cloud computing environment pushed ~~forward~~ by IQmulus.

At the time of writing, it is not possible to provide a quantitative analysis of the advancement foreseen over the state-of-the-art methods. Some of the references in the literature are cited in the tables in the Annex. Thanks to the infrastructural support to scalability testing, however, it will be possible to run specific comparions to available code. Moreover, the definition and adoption of test suites for scalability tests will constitute ~~a~~ benchmarks for the comparison of the services developed with respect to the state-~~-~~of-~~-~~the-~~-~~art.

The comparison will be fully documented with respect to the as-is analysis already discussed in the D~~.~~1.2.3 ~~document~~deliverable, and with respect to user validation results (see ~~see~~ deliverable

For each of the above-~~previous~~ mentioned services, we have identified possible solutions to previous work and how the improvement will be assessed. This information will be improved during the third year in WP4 toolkits~~,~~ according to the outcomes of the scalability testing and the updates to the IQmulus scenarios. The validation of results ~~of the results will be~~is discussed in the

1.3 ~~WHAT~~ Evaluation/performance measures

(Q4) A description of how the IQmulus system will excel over competing systems in terms of its support for big data.

~~Results~~ The results of the scalability tests will provide a proper quantitative evaluation of the of the services with respect to ~~the~~ computational aspects, and these results will be matched **realistic data sizes** as described in deliverable D~~.~~1.~~.~~2.3.~~,~~

The qualitative improvement, its assessment and validation will be achieved through:

- a comparison of the proposed approach with respect to the state ~~-~~of ~~-~~the ~~-~~art and the as-is analysis;
- a definition of metrics that will consider specific aspects of the processing services, as identified in deliverables D4.1.2 and ~~,~~D4.1.3, ~~and~~ as summarized below.

Task 4.2

- quality include registration error after alignment (i.e., point distances between registered point clouds); shape fitting (e.g., plane, sphere, etc.) for known geometries (e.g., facade, street). Finally, results of the fusion should be delivered together with a quality indicator;
- scalability include processing time; numbers of parameters; computational and storage complexity. Furthermore, temporal updates should probably only require the new information and the last state; in this way, not all old data needs to be reprocessed;
- degree of human intervention include the reduction of manual intervention during processing, but expect manual checking (10% of overall effort).

Task 4.3

For **feature extraction, classification and correlation ~~(Task 4.3)~~,** performance indicators for

- quality include the percentage of false positive/negatives, ROC curves, and comparison with ground ~~-~~truth (if available);
- robustness in time and space include the percentage of misclassification updates;
- scalability include the processing time per unit of volume data; the classification time per unit of volume data; hierarchical approaches for handling features and classes;
- degree of human intervention include the reduction of manual intervention during processing, but expect manual checking (10% of overall effort); manual corrections' feedback to the training of classifiers, which keeps the classification accuracy better than 90%. However, some human intervention is needed, as features and classes may need to be defined on the spot, depending on the requirements.

Task 4.4

For **multivariate surface generation ~~(Task 4.4)~~,** performance indicators for

- quality include the accuracy in the adaptation of the surfaces to the point clouds, smoothness and curvature of the reconstructed surfaces;
- scalability include processing time and the possibility to parallelize the computation through data partitioning;
- degree of human intervention include the percentage of manually selected parameters, which is expected to be low. However, if large changes in the model are detected and

automatic decisions cannot be taken, then we foresee that decisions are to be taken by the user.

Task 4.5

- quality include change detection accuracy vs. reference (e.g., ancillary punctual measurements);
- scalability include processing time;
- degree of human intervention include the human/manual intervention time.

TILING&STITCHING AS A CONCRETE EFFORT TOWARDS BIG DATA HANDLING

Processing large geo~~-~~spatial data is definitely a complex task, which cannot be solved by a *one-fits-all* solution; parallel or distributed computation may work nicely for some problems but may cause a considerable overhead for some other problems. Scalability tests will contribute to the analysis of what approach is best ~~-~~suited to which problem, identifying what are the limits (in terms of execution time or available in-core memory) of services on the IQmulus hardware.

In the analysis of services released so far, in relation to the input data types and sizes, it is clear that:

- most of the services need to handle efficiently large collections of moderately-sized data files;
- in order to exploit the potential of the HDFS, Hadoop I/O access should be implemented for the services;
- the partition used to store large collections does not necessarily coincide with the optimal partitioning for distributing the computation on the computing cloud;
- the services and the Job Manager need to know how the partition of the input data is defined.

Having in mind these considerations, we have clearly identified as the key for the project's next development period~~,~~ the development of **smart methods for tiling, stitching and indexing data sets**, which exploit the architectural choices made so far. Few services already address similar issues: for instance, services implementing MapReduce jobs or ~~the Vector Layer~~

We believe it is crucial for the project to coordinate the efforts towards tiling&stitching so that WP4 can work jointly ~~to~~ on the development of libraries of services that provide *general purpose* methods to manage large collections of data files, supporting a more computing-efficient partitioning, or a more efficient access to data sets.

The Consortium, the WP4 Leader, and the WP4 Task Leaders have brought to the attention of the Managing Board the proposal to "continue" Task 4.1 for the next two years with a specific focus on the development of strategies for tiling and stitching. FOMI is proposed as the new leader for this task, given the expertise of the human resources that could be assigned to the t~~T~~ask, providing perfect support for the implementation of HDFS-oriented partitioning strategies.

The task should address the following aspects:

- in-depth analysis of the data access patterns (global/local/focal/zonal) of the services ~~,~~ could indicate groups of services requiring a similar tiling&stitching strategy;
- ~~-~~based on the scalability tests, identification of the bottle~~-~~necks to be addressed with priority;
- review of the indexing methods for big data collections and a proposal of an IQmulus partitioning and indexing schema for collections, in collaboration with WP3;
- proposal of methods for partitioning and indexing big data suitable to the HDFS/MapReduce framework; in this case, the partitioning should possibly support HDFS I/O operations;
- implementation of services/libraries to perform partitioning and to re-assemble global results out of partial results (stitching).

The proposal of the continuation of Task 4.1 is under discussion and awaiting formal approval, due to the slight change in the DoW and resource allocation that it implies.

TIMELINE

(Q3) A time plan for these activities

The development of WP4 services is one part of the overall implementation plan, and scalability tests, at service and workflow level, have a crucial role in guiding the future development of the services.

The timeline for testing discussed so far, and agreed with the IQmulus Consortium for testing theis

- 6 March, 6, 2015: infrastructure ready for testing the functional prototype (user feasible)
- 15 March , 15, 2015: definition of the test suites by WP4 task leaders for stress analysis (scalability tests)
- 1 April 1, 2015: release of the test suites for stress analysis in scalability testing
- 1 April , 1, 2015: infrastructure ready for scalability testing (i.e., automation of service logging enabled at service level, test suites available on HDFS)
- End of April, 2015: first results of the scalability tests available (priority to the services selected in section 2)
- End of June, 2015: scalability tests finished (target: every service released and testable with priority to services integrated in the showcases - see WP6)
- End of September, 2015: analysis of the results of scalability tests available
- October, 2015: presentation of the main results at the ISPRS GeoBigData workshop and its IQmulus Processing Ccontest special session.

ANNEX - SERVICES

Task 4.2

Id & Service Name	Lead Show-case	Test Data URI	Size of test data	Aspect of Big Data	Comment on the big data strategy to be adopted (parallelization, compression, etc.) and implementation (Hadoop, Spark, …)
#8 Point cloud to point cloud distance	TU Delft None (at the moment)	Test data set 1 http://146.140.214.126/fs/TuDelft/TuDelftCampusData/LMMSdata/LMMSinIntersection http://146.140.214.126/fs/TuDelft/TuDelftCampusData/AHNdata/ Test data set 2: http://146.140.216/fs/CNR_IMATI/Liguria_LAS	Test data set 1: 852.3 MB + 37 MB Test data set 2: 1 TB	Size	Single node implementation using KD-trees. An improved implementation, making use of additional voxel constraints is possible in future. For the toolbox in PM36 we will design and implement a tiling and stitching strategy.
#70 Radiometric enhancement	FömiFOMI LS3	**Test data set 1** hdfs://146.140.214.127/FOMI/SPOT5_Partitioned/ **Test data set 2** hdfs://146.140.214.127/fs/FOMI/SPOT5_Mosaic/mosaic_20130426_077_254+255_sp5_1_eov_nn.tif **Test data set 3** hdfs://146.140.214.127/fs/FOMI/SPOT5_Mosaic/mosaic_20130425+0426_sp5_eov_nn.tif	Test data set 1: 8 files From 201.9 until 268.1 MB Test data set 2: 450.0 MB Test data set 3: 1.0 GB	Size	Currently only a single node version is available. A parallel Hadoop version is about to be released and could be incorporated in the upcoming scalability testing. Services #70 to #74 are designed to run on typical sattellite data sets. State of the art: The parallel Hadoop version is definitely beyond the state of the art.
#71 Convolution Filtering	FömiFOMI LS3, LS4	See Service #70	See #70	Size	This is a single node version. A parallel Hadoop version is under development but is not available for the upcoming round of scalability testing. The upcoming Hadoop version is beyond the state of the art.
#72 2D on 2D Image Registration	FömiFOMI LS3, LS4, MS4	See Service #70	See #70	Size	This is a single node version. A parallel Hadoop version is under development but is not available for the upcoming round of scalability testing. The

formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Internet Link, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Mintázat: Üres
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Internet Link, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus, Mintázat: Üres (Fehér)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Internet Link, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Mintázat: Üres (Fehér)
formázott: Normál, Sorkizárt, Térköz Utána: 0 pt, Sorköz: szimpla
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Internet Link, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Mintázat: Üres (Fehér)
formázott: Sorkizárt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Internet Link, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Mintázat: Üres (Fehér)
formázott: Sorkizárt, Nincs
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)

					upcoming Hadoop version is beyond the state of the art.
#73 Preprocessing of raster data	~~Fömi~~FOMI LS3	See Service #70	See #70	Size	This is a single node version. A parallel Hadoop version is under development but is not available for the upcoming round of scalability testing. The upcoming Hadoop version is beyond the state of the art.
#74 Topological Analysis of Land 2D data	~~Fömi~~FOMI LS3	See Service #70	See #70	Size	This is a single node version. A parallel Hadoop version is under development but is not available for the upcoming round of scalability testing. The upcoming Hadoop version is beyond the state of the art.
#96 Extraction of rain data	IMATI LS2	http://146.140.214.126/fs/CNR_IMATI/Rainfall-datasets/iQumulus_20130929.csv http://146.140.214.126/fs/CNR_IMATI/Rainfall-datasets/iQumulus_20140116.csv	2.1 MB and 10.9 MB	Complexity	This is a conversion service that works on relatively small data sets. In this way it helps in the processing of data that is big w.r.t the complexity of the data.
#98 Point cloud intersection	TU Delft None (at the moment)	http://146.140.214.126/fs/TuDelft/TuDelftCampusData/LMMSdata/ http://146.140.214.126/fs/TuDelft/TuDelftCampusData/AHNdata/ Test data set 2: http://146.140.216/fs/CNR_IMATI/Liguria_LAS	Test data set 1: 1.7 GB and 48.9 MB Test data set 2: 1 TB	Size	Single node implementaton that uses voxels to speed up computation. For the Geospatial Week we plan to report on this approach and compare to other approaches like bounding polygon intersection of brute force Flann distance evaluation. The current methods take single file as input, not yet collection of files. For the toolbox in PM36 we will design and implement a tiling and stitching strategy. It needs additional checking if any literature exists on this approach or if software exists with similar functionality.

Task 4.3

Id & Service Name	Lead Show-case	Test Data URI	Size of test data	Aspect of Big Data	Commentbig data strategy (parallelization, compression, etc.) and implementation (Hadoop, Spark, …)
#94 ~~Vector layer partitioning of a point cloud~~	IMATI ~~LS1~~	~~http://146.140.214.126/fs/CNR_IMATI/Liguria-LAS~~ See Appendix~~600 GB~~ ~~Size~~ ~~Service is exposed to full dataset. A benchmark is not available. Currently the execution of the service on the cloud can be distributed among the processing nodes via a bash script: the main bottleneck is the transfer time of the las files between HDFS and the processing nodes.~~			

formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt, angol (brit)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt, angol (brit)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus, Mintázat: Üres (Fehér)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Nincs aláhúzás, Betűszín: Automatikus, Mintázat: Üres (Fehér)
formázott: Betűtípus: 12 pt
formázott: Középre zárt
formázott: Betűtípus: +Szövegtörzs (Calibri), 12 pt
formázott: Betűtípus: 12 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)

#59 Multi-object classification of 3D point clouds	IGN US2	http://146.140.214.126/fs/IGN/Toulouse/LidarDim2	170477.8 MB	Size	Decision tree good for parallelization. The methodology has been published in PCV'14 (Best paper) and accepted in the CGF special issue.
#27 Tree crown recognition from mobile mapping point clouds	UCL US2	http://146.140.214.126/fs/IGN/Toulouse/LidarDim2	170477.8 MB	Size	Uses Amazon EC2 cloud computing. Please refer to "Scalability" document regarding logging. Uses parallelized version of decision tree. It uses 1.1 Mio points to test their classification results with an overall accuracy of 83% (without colour). No runtime is given **References** Matti et al., *Geometry and Colour Based Classification of Urban Point Cloud Scenes Using a Supervised Self-Organizing Map*, (2014)
#42 Detection of Flow Lines and Drainage Basins	IMATI n/a	http://146.140.214.126/fs/CNR_IMATI/Service42-TestData/DTM/Liguria_DTM_25.tif http://146.140.214.126/fs/CNR_IMATI/Service42-TestData/DTM/Liguria_DTM_10if	125 MB (res 25m) 800 MB (res 10m)	Size	The service is parallelized with MPI and run in core. Test dataset derived from http://146.140.214.126/fs/CNR_IMATI/Liguria-DEM; the service can deal with bigger data. Grid size between 10 and 30 meters is considered suitable for hydrological purpose. (Zhang, W., and D. R. Montgomery (1994), Digital elevation model grid size, landscape representation, and hydrologic simulations, Water Resour. Res., 30(4), 1019–1028, doi:10.1029/93WR03553) **References** Tesfa, T. K., D. G. Tarboton, D. W. Watson, K. A. T. Schreuders, M. E. Baker and R. M. Wallace, (2011), "Extraction of hydrological proximity measures from DEMs using parallel processing," Environmental Modelling & Software, 26(12): 1696-1709.
#44 Critical Points	IMATI LS2	http://146.140.214.126/fs/CNR_IMATI/LiguriaPLY/service44-TestData	342.2 MB (tested so far; possible up to memory size)	Size/Processing speed	Test dataset derived from http://146.140.214.126/fs/CNR_IMATI/Liguria-DEM; the service can deal with bigger data. Big data strategy: Embarrassing parallel (Hadoop) **References** T. F. Banchoff, Critical points and curvature for embedded polyhedral surfaces, Am. math. Monthly,Vol. 77, pp. 475-485, 1970; Xinlai Ni, Michael Garland, and John C. Hart. 2004. Fair morse functions for extracting the topological structure of a surface mesh. ACM Trans. Graph. 23, 3, 613-622. Typical data size in state of the art: ~200K vertices for a triangle mesh.
#4 Breakline extraction	IMATI MS	http://146.140.214.126/fs/CNR_IMATI/LiguriaPLY/service4-TestData	342.2 MB (tested so far; possible up to memory size)	Size	Test dataset derived from http://146.140.214.126/fs/CNR_IMATI/Liguria-DEM; the service can deal with bigger data. Big data strategy: Embarrassingly parallel (Hadoop). Typical data size in state of the art: ~100K to ~300K vertices for a triangle mesh.

- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus, Mintázat: Üres (Fehér)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus, Mintázat: Üres (Fehér)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri), 8 pt
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri), 8 pt
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
- **formázott:** Betűtípus: +Szövegtörzs (Calibri), 8 pt
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
- **formázott:** Betűtípus: +Szövegtörzs (Calibri), 8 pt
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)
- **formázott:** Betűtípus: +Szövegtörzs (Calibri)

					References Pedrini, H. & Schwartz, W. Skarbek, W. (Ed.) Topographic Feature Identification Based on Triangular Meshes Computer Analysis of Images and Patterns, Springer Berlin Heidelberg, 2001, 2124, 621-629. Sahner, J., Weber, B., Prohaska, S. and Lamecker, H. (2008), Extraction Of Feature Lines On Surface Meshes Based On Discrete Morse Theory. Computer Graphics Forum, 27: 735–742.
#64 Persistent Critical Points	IMATI LS2	http://146.140.214.126/fs/CNR_IMATI/LiguriaPLY/service64-TestData	342.2 MB (possible up to memory size)	Size/Processing speed	Test dataset derived from http://146.140.214.126/fs/CNR_IMATI/Liguria-DEM; a possible issue in dealing with bigger data can be –the memory available on the computer. Big data strategy: Distributed computation (planned). Typical data size in state of the art-: ~100K vertices for a triangle mesh **References** Yu-Shen Liu, Min Liu, Daisuke Kihara, and Karthik Ramani. 2007. Salient critical points for meshes. In Proceedings of the 2007 ACM symposium on Solid and physical modeling (SPM '07). ACM, New York, NY, USA, 277-282.
#69 Raster thresholding	FOMI LS3	hdfs://146.140.214.127/FOMI/SPOT5_Partitioned	1803.7 MB	Size	The previous Versions 1 of the services were not suited for input data larger than a few GB (depending on memory size), as they are single machine versions. Version 2 has just been released, which is the Hadoop implementation. This is better suited for testing scalability. This service is not directly used in workflows, its functionality is embedded in #80.
#76 Computation of spectral indices	FOMI LS3	hdfs://146.140.214.127/FOMI/SPOT5_Partitioned	1803.7 MB	Size	The previous Versions 1 of the services were not suited for input data larger than a few GB (depending on memory size), as they are single machine versions. Version 2 has just been released, which is the Hadoop implementation. This is better suited for testing scalability.
#80 Flood and waterlogging detection on raster data	FOMI LS3	hdfs://146.140.214.127/FOMI/SPOT5_Partitioned	1803.7 MB	Size	The previous Version 1 of the services were not suited for input data larger than a few GB (depending on memory size), as they are single machine versions. Version 2 has just been released, which is the Hadoop implementation. This is better suited for testing scalability. This service works on only on SPOT5 raster imagery which contains metadata.

formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus

#10 Outlier Classification in Point Clouds	UCL LS1	http://146.140.214.126/fs/CNR_IMATI/Liguria-LAS	600 GB	Size	Currently only single CPU. Needs to be converted to FHC or EC2, planned for April. Sotoodeh, *OUTLIER DETECTION IN LASER SCANNER POINT CLOUDS*, (2006) uses a reference dataset with 115820 points to demonstrate the quality of their approach and a dataset of 12 Mio points to show runtime performance. IGN Paris dataset for the processing contest will be used to asses accuracy IGN Toulouse dataset will be used for scalability test
#77 Raster segmentation	FOMI LS3	hdfs://146.140.214.127/FOMI/SPOT5_Partitioned	1803.7 MB	Size	Version 1 of the services is not suited for input data larger than a few GB (depending on memory size), as they are single machine versions. Version 2 with Hadoop to be released after the first round of testing
#78 Clustering of raster data	FOMI LS3	hdfs://146.140.214.127/FOMI/SPOT5_Partitioned	1803.7 MB	Size	Version 1 of the services is not suited for input data larger than a few GB (depending on memory size), as they are single machine versions. Version 2 with Hadoop to be released after the first round of testing. Not used in current LS3 workflow. Will be used in future version.
#79 Thematic classification of raster data	FOMI LS3	hdfs://146.140.214.127/FOMI/SPOT5_Partitioned	1803.7 MB	Size	Version 1 of the services is not suited for input data larger than a few GB (depending on memory size), as they are single machine versions. Version 2 with Hadoop to be released after the first round of testing. Not used in current LS3 workflow. Will be used in future version.
#17 3D Local Keypoint Extraction from Point Clouds	TUDelft US1	Data Set for Demo: #59: sftp://iqmulus.igd.fraunhofer.deå/data/upload/TUDelft/LMMS-Delft-Campus/lmms-cars/	8.5MB (Small demo data set)	none	We have two different local feature extraction software DoN and SHOT features. SHOT features come with ~~have~~ descriptor vectors which are very suitable for object recognition. Currently it is required to partition the input data into chunks of ~10MB for processing. At the moment, big data processing limitations ~~is~~ are not addressed.
#7 Static Tiling of a LAS File	UCL n/a	http://146.140.214.126/fs/LIGURIA/LIDAR	50 MB	Size	Parallelize with Hadoop over several nodes. Logging in distributed Hadoop untested. This is a pre-processing utility not directly connected to any show case. Typical data sizes can be 100s of GByte
#2 Point Cloud to Regular Grid	UCL LS1	http://146.140.214.126/fs/LIGURIA/LIDAR	50 MB	Size	Parallelize with Hadoop over several nodes. Logging in distributed Hadoop untested. Typical data sizes can be 100s of GByte
#26 Individual tree extraction	TUDELFT US2	#49: http://146.140.214.126/fs/IGN/Toulouse/ #50: sftp://iqmulus.igd.fraunhofer.de/data/upload/TUDelft/AHN2-tudelft-point-cloud/	30000 MB	none	Works on dataset reduced after classification. We have developed a linux version and compiled it on Fraunhofer VM. At the moment we are trying to make a Dockerfile. However, someone can use it on Fraunhofer VM.

formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)

		#59: sftp://iqmulus.igd.fraunhofer.de/data/upload/TUDelft/LMMS-Delft-Campus/			
#16 Quick Visualization	UCL n/a	n/a	n/a	none	Works on samples for developer testing only
#3 Peak Points	UCL n/a	n/a	n/a	none	Uses in-memory kd-tree not suitable for big data
#54 Filter Point Cloud by Attribute & Coordinate	UCL n/a	n/a	n/a	none	Uses a file based filter - single file only.

Task 4.4

Service Name	**Lead & Showcase**	**Test Data URI**	**Size of test data**	**Aspect of Big Data**	**Commentbig data strategy (parallelization, compression, etc.) and implementation (Hadoop, Spark, …)**
#9 LR B-spline approximation from point cloud	SINTEF MS1	473513_utm2_1 from HRW (half of 473513) to be uploaded	1.1G (as las with xyz only)	Size	The current version runs on a single node, but the service should use tiled data, work on tiling and stitching in progress. When this work is completed, several instances of the service will run in parallel and larger data sets will be handled.
#66 Point cloud dimensionality	IGN US2	http://146.140.214.126/fs/IGN/Toulouse/Lidar	83 G	Size	The outcome of this task is not a surface in itself. It is used as input for service #59
#49 Constrained triangulation	IMATI LS1	New dataset based on service 94: set of increasing size point clouds	~ 2/3 GB	Size	Service is single node, single core and limited by the computer memory.
#88 Distance between point cloud and LR B-spline surface	SINTEF MS2	473513_utm2_1 from HRW (half of 473513) to be uploaded surface from #9	1.1G (as las with xyz only)	Size	Same data set as #9. Release in April 2015, will be updated on the cloud. Service should use tiled data as prepared for #9
#90 Trimming of LR B-spline surface	SINTEF MS1	473513_utm2_1 from HRW (half of 473513) to be uploaded surface from #9	1.1G (as las with xyz only)	Size	Same data set as #9. Release in April 2015, will be updated on the cloud. Service should use tiled data as prepared for #9
~~**#48#48 Multi-resolution triangulation**~~	~~IMATI~~				

formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Térköz Utána: 10 pt, Sorköz: Többszörös 1,05 s.
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)

~~for land monitoring~~	See Appendix ~~http://146.140.214.126/fs/CNR_IMATI/Liguria-LAS/~~ ~~http://146.140.214.126/fs/CNR_IMATI/Liguria_Basin_map~~ ~~600GB~~ ~~Size~~ ~~The service will deal with management and processing of big data. To be delivered in April 2015.~~				
#95 Watertight surface reconstruction	IGN (US)	http://146.140.214.126/fs/IGN/Toulouse/Lidar	83 G	Size	Using the result of service #110 as input. The surface reconstruction is beyond the state of the art. It extends space carving approaches by taking into account the Dempster Shafer framework and borrows ideas from photogrammetic surface reconstruction. To be delivered in April 2015
#110 3D Delaunay triangulation	IGN (US)	http://146.140.214.126/fs/IGN/Toulouse/Lidar	83 G	Size	The service is currently using a state of the art implementation, but a cloud-enabled version will be provided. First version to be delivered in April 2015
#51 Triangulation of gridded point cloud	IMATI	http://146.140.214.126/fs/CNR_IMATI/Liguria-DEM/ Liguria-DEM	1.5G	Size	Single threaded. Scalable up to memory size of processing node. Tested on 70 million points
#55 LR B-spline approximation from raster	SINTEF	#8 sftp://iqmulus.igd.fraunhofer.de/upload/UBO/PORSMILIN/BATHYMETRY/20110419_Porsmilin_DTM_1m_MBES_L93_IGN69.zip #9 sftp://iqmulus.igd.fraunhofer.de/upload/UBO/PORSMILIN/BATHYMETRY/20100527_Porsmilin_DTM_1m_MBES_L93_IGN69.zip Data sets created from surfaces using service #9 and #91. Initial point clouds: 473513_utm2_1 from HRW http://146.140.214.126/fs/UBO/REUNION/ScalabilityTesting	13 M (ascii) 93 M (ascii) ~100M	(Size)	Service should use tiled data. The specified data sets are too small for scalability testing. Will probably construct data sets using a combination of service #9 and #91 to get similar data sets of increasing size. The service uses that same core as #9. Performing an approximation service on raster data should be used only for dense rasters as the input data will already represent an approximation.
#57 Update LR B-spline surface	SINTEF	473513_utm2_1 from HRW (half of 473513)	1.1G (as las with xyz only)	Size	Service should use tiled data as prepared for #9. It will use same data set as #9 when completed. An updated version of the service will be put on the

formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Térköz Utána: 10 pt, Sorköz: Többszörös 1,05 s.
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Térköz Utána: 10 pt, Sorköz: Többszörös 1,05 s.
formázott: Betűtípus: +Szövegtörzs (Calibri), olasz (olaszországi)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Térköz Utána: 10 pt
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)

with new point cloud		to be uploaded Surface from #9			Fraunhofer cloud.- Currently not prioritized, will be more important before D4.4.3
#84 Parameterize by projection on surface	SINTEF	#9 sftp://iqmulus.igd.fraunhofer.de/upload/UBO/PORSMILIN/BATHYMETRY/20100527_Porsmilin_DTM_1m_MBES_L93_IGN69.zip (subset)	24M (las)	(Size)	The service is related to generation of parameterized 3D surfaces. This option is relevant only in specific areas as it is more time and memory consuming than dealing with surfaces representing elevation. Thus only medium sized data is applicable.
#40 Rainfall using ordinary kriging	IMATI LS2	http://146.140.214.126/fs/CNR_IMATI/Service40-TestData	2MB/day	Variety	Rainfall data by Regione Liguria. The service is not focused on big data, but the ability to run in realtime, i.e. the service must be completed before new rainfall data transmission begins. The raingauges transmit data every 10 minutes.
# 67 Rainfall using radial basis functions	IMATI (LS2)	http://146.140.214.126/fs/CNR_IMATI/Service40-TestData	2MB/day	Variety	Matlab service. Same purpose as #40
#58 Rainfall by LR B-spline approximation	SINTEF (LS2)	http://146.140.214.126/fs/CNR_IMATI/Service40-TestData	2MB/day	Variety	Same data set as service #40, must be checked. Same purpose as #40, not used in work-flow
#56 Parameterize triangulated point set	SINTEF	#9 sftp://iqmulus.igd.fraunhofer.de/upload/UBO/PORSMILIN/BATHYMETRY/20100527_Porsmilin_DTM_1m_MBES_L93_IGN69.zip (triangulated subset)	104M (ply)	None	Works on thinned, triangulated data. The service should be used in cooperation with service #84 and the 3D version of service #9. This reduces the need for big data sizes.
#91 LR B-spline surface to raster	SINTEF	Surfaces from #9 + #90 Initial point clouds: 473513_utm2_1 from HRW http://146.140.214.126/fs/UBO/REUNION/ScalabilityTesting	~1-10M	None	Output from #9 used as input. It will be updated with respect to handle stitched surfaces created from tiled point clouds. At this stage, the data size of the original point cloud is strongly reduced by the surface generation in service #9

Task 4.5

#45 Topological change detection	IMATI LS2	http://146.140.214.126/fs/CNR_IMATI/test_suite_rain_services_IMATI/Test_Data/Liguria_DRM/	2.5MB/day; ~10MB/day; 342.2 MB (underlying surface model tested so far;	Size/Processing speed	Test dataset derived from http://146.140.214.126/fs/CNR_IMATI/Liguria-DEM and http://146.140.214.126//fs/CNR_IMATI/Rainfall-datasets; a possible issue in dealing with bigger data can be the memory available on the computer. Processing speed: the scalability issue here is toi be able to process data in "almost real-time".

formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Hiperhivatkozás, Betűtípus: +Szövegtörzs (Calibri), 11 pt, Betűszín: Automatikus
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri)
formázott: Betűtípus: +Szövegtörzs (Calibri), 8 pt
formázott: Betűtípus: +Szövegtörzs (Calibri)

			possible up to memory size)		**References** Wheaton, J. M., Brasington, J., Darby, S. E. and Sear, D. A. (2010), Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets. Earth Surf. Process. Landforms, 35: 136–156. Hafez A. Afify, Evaluation of change detection techniques for monitoring land-cover changes: A case study in new Burg El-Arab area, Alexandria Engineering Journal, 50(2), Pages 187-195
# 60 2D displacement measurements using DEMs and optical data	LS4/MS4	http://146.140.214.126/fs/UBO/PLEIADES/ScalabilityTesting	2MB/day	Size/Variety	The service is multi-threaded and readily enables fast processing of the largest image pairs currently available (VHR satellite images in < 24h). To enable timely processing of archived data the distribution of multiple instances by the Jjob Mmanager will be enabled.