PlanetData

Network of Excellence

FP7 – 257641

D8.5 2nd PlanetData Program Assessment Report

Coordinator: Anna Fensel (UIBK)
With contributions from: Zaenal Akbar (UIBK), Alice Carpentier (UIBK), Nelia Lasierra (UIBK), Elias Kärle (UIBK), Ioan Toma (UIBK), Francisco J Lopez Pellicer (UNIZAR), Jesus Barrera (GEOSLAB), Loris Bozzato (FBK), Luciano Serafini (FBK), Marta Sabou (MODUL)

1st Quality reviewer: Max Schmachtenberg (UMA)
2nd Quality reviewer: Elena Simperl (SOTON)

Deliverable nature:	Report (R)
Dissemination level:	Public (PU)
Contractual delivery date:	M48
Actual delivery date:	M48
Version:	1.0
Total number of pages:	25
Keywords:	Data Dynamicity, Quality and Provenance, Access Control, Privacy and Trust, PlanetData Programs Call 2

Abstract

The deliverable is focused on presenting the PlanetData Program assessment for the three proposals (LinkedMap, MetaReasons, ETHIQ) accepted as part of the PlanetData Program 2 (Call 2): Data Dynamicity, Quality and Provenance, including Access Control, Privacy and Trust. The deliverable offers an assessment framework split in three main categories: Performance of the Project (overall accordance to requirements and overall performance), Impact of the Project and Evolution of Requirements (assessment of the ways the applications' requirements and features changed during the implementation stage, such as changes that were made due to technical motivations, glitches that have hinged the progress of a project, as well as unexpected or surprising results). The framework has been applied on the three proposals.

Executive summary

The deliverable presents the PlanetData Program assessment for the proposals accepted as part of PlanetData Program 2, Second Call: Data Dynamicity, Quality and Provenance, including Access Control, Privacy and Trust. It presents an overview of PlanetData (a short description of the project, its objectives and reasons for creating calls), the defining features of Call 2 (description of the call's objectives, main topics of interests, formats of the projects, application procedures and facts, as well as the evaluation criteria utilized to select the proposals), descriptions of the proposals accepted, as well as an evaluation framework (an assessment approach, assessment criteria, comparisons and an evaluation matrix) to assess the progress of each project implicated in the Call.

The analysis of the three proposals shows that they have fulfilled the research goals they have set out to fulfil. Moreover, the end-products of the proposals are three applications which have received both positive end-user assessments, as well as interest from commercial enterprises.

Document Information

IST Project Number	FP7 - 257641	**Acronym**	PlanetData
Full Title	PlanetData		
Project URL	http://www.planet-data.eu/		
Document URL			
EU Project Officer	Leonhard Maqua		

Deliverable	**Number**	D8.5	**Title**	2nd PlanetData Program Assessment Report
Work Package	**Number**	WP08	**Title**	Administrative and financial management

Date of Delivery	**Contractual**	M48	**Actual**	M48
Status	Version 1.0		final X	
Nature	prototype □ report X dissemination □			
Dissemination level	public X consortium □			

Authors (Partner)	Alice Carpentier (UIBK)			
Responsible Author	**Name**	Alice Carpentier	**E-mail**	alice.carpentier@sti2.at
	Partner	UIBK	**Phone**	+43 512 507 53708

Abstract (for dissemination)	The deliverable is focused on presenting the PlanetData Program assessment for the three proposals (LinkedMap, MetaReasons, ETHIQ) accepted as part of the PlanetData Program 2 (Call 2): Data Dynamicity, Quality and Provenance, including Access Control, Privacy and Trust. The deliverable offers an assessment framework split in three main categories: Performance of the Project, Impact of the Project and Evolution of Requirements. The framework has been applied on the three proposals.
Keywords	Data Dynamicity, Quality and Provenance, Access Control, Privacy and Trust, PlanetData Programs Call 2

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
30-June-2014	0.1	Alice Carpentier	Initial version drafted
		Alice Carpentier	Updates towards its quality improvement and addition of the missing details.
19-Sep-2014	0.2	Alice Carpentier	Deliverable modified using Call 2 reviewer indications.
20-Sep-2014	0.3	Alice Carpentier	Deliverable modified using Call 2 reviewer indications.
29-Sep-2014	0.4	Alice Carpentier	Deliverable modified using Call 2 reviewer indications.
30-Sep-2014	0.5	Anna Fensel	Internal release check and update
06-Oct-2014	0.6	Max Schmachtenberg	External review
09-Oct-2014	0.7	Alice Carpentier, Anna Fensel	Update according to external review

Table of Contents

List of tables

1 Introduction

In this deliverable we present the PlanetData Program assessment for the proposals accepted as part of PlanetData Program 2 (Call 2): Data Dynamicity, Quality and Provenance, including Access Control, Privacy and Trust.

The aim of the Network of Excellence PlanetData is to establish an interdisciplinary community of researchers helping organizations to get their data exposed online in a useful form. This is achieved through an integrated program comprising research, data and technology provisioning, as well as training and dissemination, pushing forward the state-of-the-art in large-scale data management and its application to the creation of useful, open data sets; supporting data owners in mobilizing their data following Linked Data principles; and enabling consumers to leverage such data sources in a variety of scenarios and domains. The network brings together thirteen core partners spread over ten European countries, with a renowned profile in the database, data and Web mining, and semantic technologies communities, and an ecosystem of collaborators worldwide.

PlanetData mission revolves around three main objectives: (1) research – create approaches to large-scale data management from different disciplines and develop holistic solutions to the challenges of dealing with planetary-scale data; (2) data provisioning and management – offer software to support large-scale data provisioning (via the PlanetData Lab), create definitive vocabularies for the description of data sets and their context; build a catalogue of data sets, publicize guidelines and best practices for provisioning (to enable the consumption of available data sets by end-users); and (3) impact – to provide a medium through which the research results and empirical findings of the PlanetData network can be used to improve the education level related to large scale management in both academia and industry; to create a community of researchers from different disciplines; to encourage (industrial) uptake through standardization, strategic dissemination and network events [1].

To strengthen the interaction with data owners, technology and application providers, and researchers working on large scale data management topics, PlanetData has set-up an open partnership scheme. This includes a thematic PlanetData Program, through which new partners can join the consortium to undertake tasks that expand and further develop the scope of the project. In this respect, PlanetData Second Call was opened on the 15 August 2012. The current deliverable will present the elements which compose Call 2 (its description, topics of interest, and acceptance requirements), present the projects which have been accepted by the consortium, as well as offer an assessment of the success of such proposals.

The remainder of the deliverable is structured as follows: Section 2 focuses on Call 2 on the subject and requirements of the call (presented in Section 2.1), as well as on the evaluation criteria for the submitted proposals (presented in Section 2.2). Section 3 presents in detail the proposals funded, following the call: (1) “Evolving web maps” – LinkedMap, (2) “A unified framework for representation and reasoning over provenance, access control, privacy and trust metadata” – MetaReasons, and (3) “Exposing Tourism Indicators as High Quality Linked Data” - ETHIQ. Section 4 describes the evaluation framework and evaluation matrix used in the assessment of the accepted proposals. The section is split in three subsections: Section 4.1, which deals with the assessment framework which will be applied on all three proposals, Section 4.2, which mentions how the assessment has been conducted (i.e. which deliverables have been used, where the data for the assessment originated, etc.) and Section 4.3, which offers the results of the application of the evaluation framework on the three proposals. Section 5 presents the conclusions of the deliverable.

2 Call 2: Data Dynamicity, Quality and Provenance, including Access Control, Privacy and Trust

The call solicited proposals for research projects that deal with any of the dimensions explored by PlanetData, with a special focus on the combination of aspects related to data dynamicity, quality and provenance, including access control, privacy and trust [3]. Proposals should propose methods and techniques, as well as develop services and applications that consider any of these dimensions or combinations of them. Proposals could use existing data sets available as Linked Data, or open up datasets that fill a gap in existing publishing initiatives (in the context of the Linked Open Data Cloud or beyond) with respect to the types of data they cover (e.g., sensor networks, social media streams), or in terms of the quality of self-descriptive metadata available. Three main topics of interest were identified, which the proposers could consider in their submissions; this list was, however, by no means exhaustive and aimed at providing a starting point for the type of research and development work solicited through the call.

The call was opened for submissions from the 15 August 2012 until the 19 September 2012. The expected duration of the projects participating in the call ranged between 6 and 12 months, starting earliest in October 2013. Requirements for participation in the call included that the proposals were to be submitted in English and could include one or more organisations eligible for EU funding. The indicative funding for the proposals was selected as between EURO 15,000 and EURO 50,000 [2].

2.1 Description

Examples of topics of interest relevant for Call2 proposals included (but were not limited to):

Data dynamicity

- Generation and exploitation of data streams across different data modalities (e.g., text, numerical values, sensor data, multimedia).
- Characterization of dynamic data sources.
- Integration of heterogeneous dynamic data sources across different modalities, with a special attention to dealing with large and very large data volumes.
- Data versioning and curation approaches for dynamic data sources.
- Novel applications and services combining static and dynamic data sets, exposed according to Linked Data principles or similar.

Data quality and curation

- Assessment and improvement of the quality of large data sets.
- Applications proposing new, or evaluating existing, quality assessment metrics for large data sets.
- Applications proposing combinations of automatic and crowdsourced techniques to improve the quality of data sets (for instance, by adding labels, correcting facts etc.)
- Empirical studies on the quality of Linked Data and on the trade-offs associated with the additional effort required to curate data vs its expected increased usability.

Data provenance, access control, privacy and trust

- Evaluation and exploitation of provenance information for large datasets.
- Evaluation and exploitation of access control information for large datasets.
- Evaluation and exploitation of privacy information for large datasets.
- Evaluation and exploitation of trust information for large datasets.
- Integration of provenance propagation in SPARQL query and update languages.

Proposals may also aim at providing integrated repositories/portals for specific fields (e.g., geospatial information, public administration, etc.) addressing the topics discussed above.

2.2 Evaluation/Impact Criteria

The evaluation framework chosen to select the submitted proposals included three main criteria:

- Scientific and/or technological excellence (relevant to the topics addressed by the call);
- Quality and efficiency of the implementation plan
- Potential impact.

3 Accepted Proposals

As a response to the call, 13 proposals were received, all of which were valid, and evaluated according to scientific and/or technological excellence, quality, and impact. Out of the 13 proposals, 6 proposals were considered eligible for funding. Taking into consideration the budget available for funding sub-projects in the framework of Call 2, the four best ranked relevant proposals were selected for funding: (1) "Evolving web maps" – LinkedMap, (2) "A unified framework for representation and reasoning over provenance, access control, privacy and trust metadata" – MetaReasons, (3) "Exposing Tourism Indicators as High Quality Linked Data" – ETHIQ, and (4) "Scalable components for the Web of Data that maintain a materialized, dynamically updated and integrated view over a multitude of heterogeneous sources – QUADAPT". The proposers of the selected proposals where to be officially added to the PlanetData project consortium via an EC amendment by the end of September 2013, so that their work could commence on 01 October 2014. The proposers of QUADAPT informed the project coordinator on 13 December 2013 that, due to the long delay of the implementation of the EC amendment, they were no longer in a position to become partners of PlanetData. Consequently the proposal QUADAPT was not funded. The proposers three remaining proposals where officially added to the PlanetData project consortium via an EC amendment on 16 December 2013. More details of the three funded proposals are presented in the remainder of this section.

3.1 Proposal 1: LinkedMap

The first proposal selected was "Evolving web maps", led by Universidad Zaragoza and Geospatiumlab S.L. [3]

The proposers described their objectives and goals as follows: A Linked Data map is an interactive Web application that mixes maps and Linked Data. Two illustrative examples are the projects Opendatamap and LinkedGeoData. The first makes use mainly of authoritative Geographic Information (GI) datasets that describe the University of Southampton and surroundings. These datasets contain data from UK government agencies such as Ordnance Survey and Royal Mail. The second publishes as Linked Data the content of OpenStreetMap, a well-known Volunteered Geographic Information (VGI) initiative that aims to map the Earth. The term VGI identifies the creation, assembly and dissemination of GI by a growing number of amateurs with the help of Web-based crowdsourcing communities. The quality and provenance of VGI data has been a recent topic of research. In some scenarios VGI data is better than GI data provided by an authoritative provider. Large GI producers, such as USGS and TomTom, have recently started to evaluate the quality of VGI data, the effort required to combine VGI data with GI datasets, and how to deal with challenges such as VGI provenance.

The mix of GI and VGI with Linked Data has enormous potential. However, this combination has a slow pace of adoption in large GI producers (NASA, ESA, national mapping agencies, cadastres, etc.) even with success stories such as Ordnance Survey. Linked data is perceived as a disruptive technology not interoperable with existing infrastructure investments based on standard geographic Web services, such as OGC Web map services. Additionally, Linked Data approaches usually offer read-only access to the data. These technical issues are hampering the adoption of Linked data.

The project envisions a platform based on standard geographic Web services, in particular. Web map services, enhanced with read-write Linked Data support on such services, and able to improve through crowdsourcing methods the quality of an automatic combination of GI and VGI data is a killer app for the adoption of both VGI and Linked Data in large GI producers.

The following four goals were identified to fulfil the vision:

1. The development of methods involving Linked Data for integrating authoritative GI with VGI, and its application to large datasets.

2. The development of guidelines for extending standard geospatial Web map services with read-write Linked Data support, and the implementation of a reference service. We name this service Linked Map.

3. The development of a Web platform for testing crowdsourcing techniques based on standard geospatial Web services and Linked Data with provenance and access control support.

4. The design of metrics and experiments for the evaluation of crowdsourcing techniques on the platform, and its application to evaluate crowdsourcing for the curation of the integration of an authoritative dataset with several VGI datasets.

The integration of a large National Map with several VGI datasets was identified as an application scenario. The National Map of choice is the Spanish Numerical Cartographic Database (BCN25/BTN25), which can be downloaded free for non-commercial use. Its use in the production of Linked Data is well documented (e.g. GeoLinkedData.es). The VGI datasets to be integrated with the BNC25/BTN25 shall include at least a "pure" VGI dataset (e.g. OpenStreetMap), a cross-domain dataset with a geographic subset (e.g. Wikipedia), and the continuous user edits of both datasets. The data provision are to be based on Linked Data counterparts (e.g. LinkedGeoData, DBpedia) if they are not outdated.

3.2 Proposal 2: MetaReasons

The second proposal selected was "A unified framework for representation and reasoning over provenance, access control, privacy and trust metadata", led by Fondazione Bruno Kessler [4].

The proposers described their objectives and goals as follows: As the use of Semantic Web/LOD data expands in dimensions and the number of related tools and applications increases, it is expected that the management of different aspects of such data become more and more urgent. Among these aspects, we can find notions of provenance, access control, privacy and trust relative to the use of combinations of possibly heterogeneous datasets and sources of information.

In the case of the above aspects, however, we note a basic feature that is common to all of them: such aspects basically consist of meta information that qualify or relate to each other different datasets. As such, sources of data may be seen as atomic elements in the representation (and reasoning) of such aspects, in which moreover other subjects (e.g. the users accessing such sources or the evolution of data over its processing) may be represented. In other words, when such aspects are represented (and possibly, when reasoning takes place over such representations), this modelling may be considered to assert about and not inside each single information source.

Reasoning takes clearly a major role in the flexible management of such aspects. After a suitable model representation has been formulated and formalized for each aspect, reasoning over such model is useful in order to infer implicit information from the available metadata. Reasoning over meta information is useful to extract implicit relations among datasets (for example, deriving all of the datasets directly or indirectly accessible from a set of users) or for classifying newly added elements to the previously known model of the system (e.g., upon discovery of a new dataset, infer if this can be trusted on the base of its relations with other components of the domain). The problem of reasoning about provenance, access control, privacy and trust of data sources has clearly already been considered in the area of SW/LOD data. Different formalisms and (OWL) ontologies for reasoning over such different aspects have been largely proposed. Many of these representations are in fact formalizations of known models for such aspects, adapted to the SW/LOD scenario or expressed in an ontology language.

Another issue of the representation of such meta information is that there exist different models and interpretations for each of the single considered aspects. This is also due to the fact that different models may be useful in different situations and can be formalized in different formalisms for capability of reasoning (e.g. trust may be represented as general property for a given entity or as trust in executing a specific task/accessing a specific resource).

As we have described them so far, each aspect is to be considered separately and may be even formalized in a different language than the others. On the other hand, many of the meta level aspects are clearly dependent one with the other (e.g. one may require to access a set of data with a specific trust and privacy criteria). Thus, they should not be considered as separate aspects in reasoning, even if we may consider different formalisms for inference in each of these aspects. This is reflected also by the fact that some existing ontologies and models consider and couple more than one aspect. Relations between each of the different aspects may be defined by the modeller: these relations define policies with respect these aspects or in general can be seen as rules that bridge and constrain different aspects. This means that for the verification of such policies, reasoning under

different theories is necessary. Moreover, we also have to consider the relations between reasoning (and querying) about datasets in the meta level and reasoning inside each dataset.

Indeed, reasoning and query answering in single datasets can now be influenced by all of the reasoning at the meta level: one has to consider all of the inferable meta information about the considered datasets when querying or accessing knowledge from the scope of a dataset to another. The latter issues clearly call for a combination of reasoning and querying between the meta level (i.e. about datasets) and the knowledge level (i.e. inside datasets). Moreover, in the meta level, it is necessary to provide the means for a combination of the meta theories for all of the considered aspects in a coherent meta knowledge.

Considering these motivations, we want to develop an architecture for representation and reasoning over aspects of provenance, data access level, privacy information and trust meta information under an integrated meta level structure. The goal is to define an architecture and its implementation for combined reasoning and querying between such dataset meta information and the contents of the data sources.

Consequently, the main of objectives are:

1) Integrated representation of aspects for provenance, access control, privacy and trust

The first objective is the theoretical definition of the proposed ideal system architecture. In particular, it will be necessary to study how to define a general formal framework for the integration of the meta theories and a suitably expressive definition of policies for their interconnection. The experience of the participant in the development of contextual frameworks can be used for this definition: as previously noted, the basic architecture and notions of the proposed system are common to the developed contextual frameworks and this structure may be seen as a generalization of them.

2) Meta theories modelling and encoding

After the basic architecture of the integration framework has been defined, the second objective regards the recognition (and possibly definition) of models to be encoded in the meta theories representing relevant models from provenance, access control, privacy and trust. As previously described, such meta theories may be already available as, e.g., OWL ontologies that can be readily "plugged in" the proposed framework. On the other hand, encoding of different models may require a formalization and encoding of further meta theories. It is our objective thus to define a reasonable set of meta theories that can be combined to represent common modelling scenarios in the set of the considered aspects.

3) Prototype implementation for the integration framework

After such theoretical definitions for the framework structure and meta theory components, our objective is also to provide a prototype implementation for it. Such implementation will cover the aspects of reasoning and querying over the combination of the different meta theories and the union of meta and object knowledge. Again, as in the case of the theory development, the starting point for the definition of such implementation will be the current implementation for the contextual framework CKR developed by the proposing participant. On the other hand, issues of scalability and applicability to a large number of distributed datasets have to be considered in order to meet the applicability to the objectives of the main ICT project.

4) Evaluation of the framework

Finally, another objective of our work regards the evaluation of the developed theory and prototype with respect to real representation needs. To reach this goal, an assessment with respect to known test beds and use cases for each of the considered aspects (and combination of them) will provide a proof of the suitability of the framework for modelling real-life situations. On the other hand, it is our objective also to provide an evaluation of the implementation with respect to the scalability of reasoning and querying, for assessing the applicability of the approach to several and large datasets.

3.3 Proposal 3: ETHIQ

The third proposal accepted was “Exposing Tourism Indicators as High Quality Linked Data”, led by the MODUL University Vienna [5].

The proposers described their objectives and goals as follows: Tourism is a key sector of the European economy accounting for five percent of the EU gross domestic product. The tourism domain is a highly complex and dynamic domain where decision-makers often rely on forecasting models to predict future demand or on decision support systems to analyze and compare the relevant stakeholders (e.g., competing regions). Tourism statistics (also known as indicators) such as the number of tourists that arrive to and sleep at a destination are important for various decision making related tasks such as (i) understanding the contribution of tourism to the destination’s economy or (ii) promoting and marketing a destination by forecasting tourism demand, setting marketing goals and exploring potential source markets. In addition, tourism planners and public agencies can use tourism statistics to decide on planning tourism related facilities and infrastructure such as airports, highways, bridges and water treatment facilities.

The decision making processes described above often require combining data from multiple data sources. However, while tourism indicator data sets are provided by various organizations, they primarily exist in isolation and they are often difficult to combine and compare automatically thus hampering effective decision making. The reasons for these difficulties are twofold. Firstly, at the data encoding level (syntactic level), while most data sets are published as open data, they use syntactic encoding formats that lead to substantial manual effort when integrating them (e.g., data dumps, custom APIs). Secondly, at the data semantics level, they contain data of different geographic granularity, time frequency or they employ different ways of measuring the same indicator, but all these differences are not made explicit in a machine readable format and it remains the task of an analyst to understand them.

Supporting data integration at both syntactic and semantic level is the goal of the Linked Data (LD) paradigm, which will be used in this project in the tourism setting. Concretely, the aim is to support tourism data owners to publish, connect and, subsequently exploit their data sources for supporting decisions by following Linked Data principles. This aim translates into the following objectives:

> Objective 1 Publish tourism indicators as high quality Linked Data: While many tourism indicator datasets exist, these are either not available as LD or have been exposed by third parties as low-quality datasets. The objective is to create a high quality Linked Data repository of tourism (and related) indicators based on TourMIS2, the largest and most detailed European dataset of tourism indicators. To ensure a high quality publication, the focus will be on (i) a high data interoperability through detailed semantic specifications and (ii) improved traceability by specifying data provenance.
>
> Objective 2 Extend and connect tourism data with external indicators: Tourism decision makers often analyze tourism indicators in tandem with indicators that are external to the domain, for example, those that shed light on sustainability or economic developments within a target region. The objective is to extend the ETIHQ repository with indicators extracted from external sources, convert these to LD, record their provenance and link them to the ETIHQ data.
>
> Objective 3 Build a tourism decision support application: Decision support applications that allow simultaneous analysis of indicators from diverse domains are scarce in tourism and generally as LD applications. The project will leverage the ETIHQ data for offering decision support to tourism experts by combining and visualising data from different sources. By creating a usable prototype, the ETIHQ dataset is validated and provide an insight into the trade-offs of LD for the tourism domain and beyond.

The project builds an integrated repository of Linked Data for tourism-specific and tourism related indicators. The publication process will aim at creating high-quality, semantically rich data sources as well as to record provenance information. An integrated architecture and a prototype decision support system evaluates the quality of the published data and create value for tourism stakeholders.

4 Evaluation Framework

The current section presents the evaluation framework selected to create the assessment of the sub-projects based on the three funded proposals. The first step in creating an evaluation framework is to determine what needs to be verified.

The remainder of this section is structured as follows: Section 4.1 presents the criteria selected for the evaluation, as well as short descriptions for each (why they were selected, what the criteria intends to measure, etc.), Section 4.2 focuses on the methods employed to assess the proposals (based on the criteria selected), while Section 4.3 presents the results of the assessment process.

4.1 Assessment Framework

The use of the same criteria across the evaluation serves the purpose of reducing the variations in approaches and reporting format between the three proposals assessed. Moreover, the consistent application of the evaluation framework aims to ensure that the three projects' impacts are addressed in a systematic way, the results and performance are comparable (across the three proposals) and that the performance and results are more feasibly provided.

In order to achieve these goals, the evaluation framework is split in three main categories: (a) Performance of the Project, (b) Impact of the Project, and (c) Evolution of Requirements. Each category is divided into a number of elements (or sub-criteria) (see Table 1 for an overview of the framework and grading system). The assessment of each criterion will be approached separately (for each of the proposals).

Table 1: Summary of evaluation framework and grading system

Category	Criterion	Sub-criterion	Grading System
Performance of the Project	Relevance	-	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
	Effectiveness	-	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
	Design and Delivery	Functionality	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
		User friendliness	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
		User interface	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
		Best practices proposal	Yes (which?)/No
		Performance	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
	Reliability of results	-	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
	Usage of PlanetData Facilities	-	Yes (which?)/No
	Usage of pre-existing facilities	-	Yes (which?)/No
	Familiarity with related work	-	Yes (which?)/No

	(re)use of technologies from other EU projects	-	Yes (which?)/No
Impact of the Project	Usefulness of application	-	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
	Potential fields of application	-	Open ended question: either answer or "None"
	Advance in technologies for using Linked Data	-	Open ended question: either answer or "None"
	Closeness to the market	-	Open ended question: either answer or "None"
	Impact of proposal results	-	1 – 5 scale, where 1 is "Not at all fulfilled" and 5 is "Completely fulfilled"
	Contributions	-	Open ended question: either answer or "None"
Evolution of requirements	Open ended category	-	Provide detailed descriptions of the elements relevant to this category.

The Performance of the Project is evaluated through the following criteria:

- **Relevance**: degree to which the project is consistent with individual project requirements;
- **Effectiveness**: degree to which the project has been effective in reaching its objectives;
- **Design and Delivery**:
    - Is it functional? Does the application do what it set out to achieve (overall assessment of requirements – verification of whether the applications have fulfilled the goals they set up to achieve).
    - Is it user friendly? This criterion focuses mostly on the concept of usability, particularly its positive aspect. The mark for this criterion will be given based on whether the application is in accordance with the following five elements: (1) learnability ("Does the application have a learning curve? How easy it is for users to accomplish basic tasks the first time they encounter the application?"), (2) efficiency (determine how quickly users can perform tasks after they have familiarized themselves with the design), (3) memorability ("Can the user perform tasks with the same efficiency after not using it for a period of time?"), (4) errors ("How many errors the user make when using it?") and (5) satisfaction (how pleasant the design is).
    - User Interface – the marks for this criterion are given based on whether the application follows the following rules: clarity (the content is conveyed quickly and accurately), discriminability (the displayed information can be easily distinguished), conciseness (users are not overloaded with unnecessary details), consistency (a design that is in conformity with what the user would expect for a particular application), detectability (the user can identify the information that is important), legibility (easy to read), and comprehensibility (the meaning is clear, concise, unambiguous).
    - Best practices proposal – "Does the project propose solutions and ideas for best practices?". From the best practices proposal point of view, the three projects have been evaluated based on the principles proposed in D2.1 ("Conceptual model and best practices for high-quality metadata publishing").
    - Performance – refers to the degree of accuracy, completeness, etc.
- **Reliability of results** - The current criterion refers to the degree to which the results presented by the developing teams are trusted and reliable. In order to measure this feature, two aspects have been taken into consideration: correctness and trustworthiness of software validations and those of the end-user validation.

- **Familiarity with related work** – the creators of an application must be familiar with similar implementations and research, with their strengths and weaknesses, in order to make a plausible contribution.
- **Usage of PlanetData facilities** – mention which PlanetData tools have been used by an application;
- **(re)use of technologies from other EU projects** – mention which EU project facilities and technologies have been used.
- **Usage of pre-existing facilities** – mention which data-related facilities have been used.

Taken together, the criteria presented above capture how well the implemented proposals met the requirements of the Call, and intends to answer the following question: "Where the things done right?". As such, it focuses on the quality of the project objectives (they have done the "right things") and the extent that the objectives were achieved at the required time (they have been "done right").

The Impact of the Project category is split in the following sub-criteria:

- **Usefulness of the application** - a quantitative mark that gives the reader an idea on the commercial potential of the application. The current criterion is couples with the following one ("Potential fields of application") to create a complete view of what impact the application might have on the market.
- **Potential fields of application**. The current criterion is concerned with the domains in which the application can be utilised. In order to measure just how versatile the proposals are, the criterion will be applied for the applications offered by each proposal.
- **Advance in technologies for using Linked Data** – one of the requirements of the Call.
- **Closeness to the market** – verify whether the final product is a proof of concept or ground research. Additionally, if possible, mention the companies that would be interested in exploiting the results.
- **Impact of proposal results** – a quantitative mark to determine the potential success of the application on the market.
- **Contributions** – enumerate the contributions brought forth by the project.

The Evolution of Requirements category refers to the assessment of the ways the applications' requirements and features changed during the implementation stage (changes that were made due to technical motivations, glitches that have hinged the progress of a project, as well as unexpected or surprising results). The purpose of this category is to address how well the individuals responsible for the implemented proposals have identified, prepared and supervised their respective projects. Additionally, this category is also intended to clarify certain aspects in the development of the project: "why the project has proceeded in a certain way (if the approach was different than what was expected)?", "what shortcomings (if they were) can be identified and how can they be avoided in the future?" etc.

The evaluation criteria have been summarized in Table 1. For the criteria that do not have a binary form of assessment (i.e. "Yes/No"), the deliverable introduces a system of rating intended to quantify or qualify the assessment. The proposed scale is symmetrical (there are two positive and two negative ratings), and the distance between ratings is conceptually equal. The scales chosen are: High (5), Substantial (4), Modest (3), Negligible (2) and "None" (1). Additionally the deliverable will employ the use of open ended questions, designed to offer a full, meaningful answer, as opposed to closed-ended questions, which encourage short or single word answers.

4.2 Evaluation Methodology

The evaluation of the three proposals selected is performed in two stages. First, the requirements and objectives of the proposals are identified. Second, the proposals are evaluated on the criteria according to the information provided by the following deliverables:

LinkedMap

D15.1 Call2: LinkedMap Requirements definition and conceptual architecture

D16.1 Call2: LinkedMap VGI Provenance schema

D16.2 Call2: LinkedMap Provisioning service

D16.3 Call2: LinkedMap Authoritative dataset

D16.4 Call2: LinkedMap Data access/update service

D17.1 Call2: LinkedMap Read-write Linked Data enabled OGC Web map server

D17.2 Call2: LinkedMap Read-write Linked Data enabled OGC Web map client

D18.1 Call2: LinkedMap Platform Alpha version

D18.2 Call2: LinkedMap Platform Beta version

D18.3 Call2: LinkedMap Platform Monitoring report,

D19.1 Call2: LinkedMap Report on VGI data quality factors

D19.2 Call2: LinkedMap Report on crowdsourcing trade-offs for geospatial data curation

D20.1 Call2: LinkedMap Web site, social media channels, fact sheet

D20.2 Call2: LinkedMap Community awareness plan

D20.3 Call2: LinkedMap Exploitation plan

D20.4 Call2: LinkedMap Standardization report

D20.5 Call2: LinkedMap Web enabled public showcase

D21.1 Call2: LinkedMap Project Handbook

D21.2 Call2: LinkedMap 6-montly report

D21.3 Call2: LinkedMap Final report and project showcase

MetaReasons

D31.1 Call2: MetaReasons Theoretical architecture description

D31.2 Call2: MetaReasons Metat heories list and description

D32.1a Call2: MetaReasons System prototype (version 1)

D32.1b Call2: MetaReasons System prototype (version 2, final)

D33.1 Call2: MetaReasons Prototype evaluation: modelling and scalability,

ETHIQ

D26.1 Call2: ETIHQ Semantic Modelling of Tourism Indicators

D27.1 Call2: ETIHQ The Repository

D28.1 Call2: ETIHQ Tourism Decision Support using Linked Data

D29.1 Call2: ETIHQ Project Website

D29.2 Call2: ETIHQ Dissemination and Exploitation Report.

D30.1 Call2: ETIHQ Project Report

The evaluations have been performed by the experts, who have not been involved in the project teams, and had no other actual or potential conflicts of interest. When applicable, also the produced prototypes have been examined. The reader should note that the assessment of each criterion has been mainly done in a subjective manner by the deliverable authors, as the deliverable intends to quantify mostly qualitative data. Finally, however, the assessments have been checked by the project leaders, for ensuring the factual completeness and correctness. This deliverable and its assessment part also in particular had been available to the PlanetData consortium over an adequate time period before the submission, so that any possible concerns about inaccuracies could be raised.

4.3 Results

The current section presents the results of the evaluations for the deliverables and tools implemented for each proposal.

LinkedMap

Objective 1:
The development of methods involving Linked Data for integrating authoritative GI with VGI, and its application to large datasets.

Assessment:
Introduced a Linked Map system, comprises of 4 components:

1. Provisioning; responsible of the provisioning persistence of several large VGI data with a large authoritative geographic dataset, including their transformation into RDF data and their integration using RDF links
2. Persistence; responsible to manage the storage of RDF data
3. Data Access API; a Linked Map Service (LMS) as proxy for external Web Map Servers, including a GeoSPARQL endpoint
4. Portal; a web portal as testbed for crowdsourcing experiments

Objective 2:
The development of guidelines for extending standard geospatial Web map services with read-write Linked Data support, and the implementation of a reference service. We name this service Linked Map.

Assessment:
Introduced the first prototype of Linked Map Service (LMS), has functionalities:

- Proxy for Web Map Servers (WMS), which is conforms to the OGC WMS implementation specification 1.3.0
- Linked Data frontend; enables RDF and HTML browsers to navigate the content of the datastore and allows users to make comments or edits the geographic data

Objective 3:
The development of a Web platform for testing crowdsourcing techniques based on standard geospatial Web services and Linked Data with provenance and access control support.

Assessment:
Introduced a Linked Map Platform which combines semantic integration with geographic maps, data querying and visual interface, including a monitoring mechanism for collecting the users' usage statistics of the platform.

Objective 4:
The design of metrics and experiments for the evaluation of crowdsourcing techniques on the platform, and its application to evaluate crowdsourcing for the curation of the integration of an authoritative dataset with several VGI datasets.

Assessment:
Introduced several quality assessments concepts:

- Conceptual model for data quality (based on model for data quality introduced in Deliverable D2.1)
- quantitative dimensions shared with GI (e.g. completeness and resolution, positional accuracy, temporal accuracy and temporal quality)
- non-quantitative dimensions shared with GI (e.g. lineage, purpose, usage and constrains)
- dimensions exclusive for VGI (e.g. believability, compliance, convergence)

Table 2: Summary of LinkedData evaluation

Category	Criterion	Sub-criterion	Grading System
Performance of the Project	Relevance	-	5
	Effectiveness	-	5
	Design and Delivery	Functionality	5
		User friendliness	5
		User interface	4 (*)
		Best practices proposal	5
		Performance	5
	Reliability of results	-	5
	Usage of PlanetData Facilities	-	Yes (data quality model in D2.1)
	Usage of pre-existing facilities	-	Yes (BNC25/BTN25, OpenStreetMap, DBpedia)
	Familiarity with related work	-	Yes (GeoSPARQL, PROV-O)
	(re)use of technologies from other EU projects	-	Yes (Strabon from SemsorGrid4Env, Morph-RDB, formerly ODEMapster, from NeOn project)
Impact of the Project	Usefulness of application	-	5
	Potential fields of application	-	Geospatial data enrichment
	Advance in technologies for using Linked Data	-	Yes, linked the authoritative GI with VGI and cross-domain information
	Closeness to the market	-	Can be adopted by any Geospatial data provider
	Impact of proposal results	-	5
	Contributions	-	- A system to integrate static geographic data with dynamic data - A web platform for read-write Linked Data
Evolution of requirements	Open ended category	-	The duration for experiments was reduced from 5 months to 4 months. This was caused by the delay in the approval of Call 2. Finally, a delay in the development of the Web platform reduced the duration to 3 months. In addition, the project suffered the loss of 3 weeks of experimental data caused by software bugs.

Notes:
(*) The system performance is affected by the slow map loading time.

MetaReasons

Objective 1:
Integrated representation of aspects for provenance, access control, privacy and trust

Assessment:
Introduced an integrated framework, comprises of two layered structures:

1. the upper-level, named metalevel (meta knowledge level) which represents the metadata information of datasets,
2. the lower-level, named object level (object knowledge level) which represents contents of the datasets

Each aspect of datasets is encoded in a distinct theory (called metatheory) and connected to each other through dependency rules (policies). Each information source can be seen as an atomic component inside the metalevel. Unified queries can be performed over both the meta knowledge and object knowledge by spanning the query over the knowledge in the integrated datasets by constraining their properties in the meta knowledge.

Objective 2:
Meta theories modelling and encoding

Assessment:
Proposed a list of meta theories, where some of them were adopted from existing models.

- Provenance models: Flouris et.al, PROV-O
- Access Control models: Papakonstantinou et.al, ROWLBAC
- Trust models: a unifying OWL ontology
- Further aspects:
  - Data quality: model in D2.1
  - Factuality: event factuality

Objective 3:
Prototype implementation for the integration framework

Assessment:
Introduced SPRINGLES (Sparql-based Rule Inference over Named Graphs Layer Extending Sesame) as the base layer for the implementation of MetaReasons framework, where:

- RDF named graphs are used to differentiate each of the datasets and the meta knowledge
- All graphs are expressed in OWL-RL
- Each meta theory model is defined as OWL ontology
- Inference enabled as SPARQL based forward rules

Objective 4:
Evaluation of the framework

Assessment:
The evaluations were performed on different MetaReasons approaches, where an inference could be performed to the metaknowledge, to the dataset (local reasoning) or both:

- mr-rdfs-global, inference is only applied to the metaknowledge, no local reasoning, applies only inference rules for RDFS and metatheory rules
- mr-rdfs-local, inference is applied to the metaknowledge and to datasets, applies only inference rules for RDFS and metatheory rules
- mr-owl-global, inference is only applied to the metaknowledge, no local reasoning, applies all inference rules for SROIQ-RL and metatheory rules
- mr-owl-local, inference is applied to the metaknowledge and to datasets, applies all inference rules for SROIQ-RL and metatheory rules

To assess the impact to system performance, a comparison has been performed using OWLIM Lite and Sesame Native RDF stores, as well as using a real use case data (the OpenData Trentino repository).

Table 3: Summary of MetaReasons evaluation

Category	Criterion	Sub-criterion	Grading System
Performance of the Project	Relevance	-	5
	Effectiveness	-	5
	Design and Delivery	Functionality	5
		User friendliness	4 (*)
		User interface	4 (**)
		Best practices proposal	Not Applicable
		Performance	5
	Reliability of results	-	5
	Usage of PlanetData Facilities	-	Yes (data quality model in D2.1, access control model in D3.1, provenance model in D3.2)
	Usage of pre-existing facilities	-	Yes (Sesame, OWLIM, OpenData Trentino)
	Familiarity with related work	-	Yes (PSL, ROWLBAC, Rei and Kaos, EAC, PPO)
	(re)use of technologies from other EU projects	-	No
Impact of the Project	Usefulness of application	-	5
	Potential fields of application	-	All applications that require to manage the provenance, access control, privacy, trust aspects or other combinations of metalevel aspects
	Advance in technologies for using Linked Data	-	Can be applied to any LOD data
	Closeness to the market	-	Can be adopted by any RDF data-storage provider
	Impact of proposal results	-	5
	Contributions	-	- An architecture for integration of provenance, access control, privacy, trust aspects - A representation for adaptation of different models in a single framework
Evolution of requirements	Open ended category	-	

Notes:
(* / **) No public access to the implementation (REST interface of a MetaReasons server).

ETHIQ

Objective 1:
Publish tourism indicators as high quality Linked Data

Assessment: As understood from deliverable D26.1 some ontologies have been built to match the very specific needs of ETIHQ. In deliverable D27.1 the necessary steps to the publishing of high quality Linked Data are listed and described. For those steps the authors described a lot of different tools available on the market and justified why they used which tools. All the ontologies as well as a dump of the data (1.8 GB) are available on the web (in a zip file, which is not that handy) on: http://data.etihq.eu/.

Objective 2:
Extend and connect tourism data with external indicator

Assessment: In deliverable D27.1 the authors described that they collected and linked data from and with external sources which were World Bank and Eurostat. External indicators they linked were for example: international tourism number of arrivals, international tourism number of departures, nights spent by type of accommodation and more (~120 indicators as mentioned in the introduction). Of particular importance was the linking between tourism indicators and indicators from other domains, namely economy and environment.

Objective 3:
Build a tourism decision support application

Assessment: After testing a demo version of the tourism decision support system, as described in D28.1, the following outcomes can be observed: appealing and rather intuitive user interface design, good performance, steep learning curve in usage of software, and an interesting use case described in D28.1. No bugs or other issues have been observed during testing.

Table 4: Summary of ETHIQ evaluation

Category	Criterion	Sub-criterion	Grading System
Performance of the Project	Relevance	-	5
	Effectiveness	-	5
	Design and Delivery	Functionality	5
		User friendliness	5
		User interface	5
		Best practices proposal	Yes
		Performance	5
	Reliability of results	-	5
	Usage of PlanetData Facilities	-	No
	Usage of pre-existing facilities	-	Yes – Ontologies
	Familiarity with related work	-	Yes
	(re)use of technologies from other EU projects	-	NA
Impact of the Project	Usefulness of application	-	5
	Potential fields of application	-	Tourism prediction, complex tourism decision making
	Advance in technologies for using Linked Data	-	Yes, decision support system for HQ tourism linked data
	Closeness to the market	-	Yes; Since the decision support application uses technology of the webLyzard startup, it is ready for market deployment in tourism and beyond in any domains that need to inspect statistical indicators
	Impact of proposal results	-	5

	Contributions	-	- Published TourMIS data as HQ linked data - Developed tourism decision support system - Performed intensive dissemination of LD technology in the tourism domain (including a survey with 37 tourism practitioners)
Evolution of requirements	Open ended category	-	NA

5 Conclusion

In the current deliverable we presented the PlanetData assessment of the proposals accepted as part of Call 2: Data Dynamicity, Quality and Provenance, including Access Control, Privacy and Trust. The three proposals selected were LinkedMap, MetaReasons and ETHIQ. LinkedMap's purpose is to develop methods and guidelines for Linked Data map - an interactive Web application that mixes maps and Linked Data – including crowdsourcing techniques in this area and their evaluation. The aim of MetaReasons is to design, develop and evaluate a unified framework for representation and reasoning over provenance, access control, privacy and trust metadata . ETHIQ's aim is to support tourism data owners to publish, connect and, subsequently exploit their data sources for supporting decisions by following Linked Data principles, particularly, in the tourism domain.

The analysis of the three proposals shows that they have fulfilled the research goals they have set out to fulfil. Moreover, the end-products of the proposals are three applications which have received both positive end-user assessments, as well as interest from commercial enterprises. Additionally, the applications have fulfilled the goals set out by the Call 2 requirements for acceptance, and partially kept inline with the earlier aspects emphasized in Call 1, namely, best practices of linked data publishing and consumption.

References

[1] http://www.planet-data.eu/

[2] http://www.planet-data.eu/news/call2

[3] proposal document "LinkedMap"

[4] proposal document "MetaReasoner"

[5] proposal document "ETHIQ"