



This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 603824.

Evaluation Plan

Deliverable D6.1 :: Public

Keywords: Evaluation, Assessment, Metrics, Demonstration Pilots

**Linked Open Data for environment
protection in Smart Regions**

Document Metadata

Contractual Date of Delivery to the EC:	31/07/2014		
Actual Date of Delivery to the EC:	26/08/2014		
Editor(s):	Marco Amadori, FBK		
Contributor(s):	MAC, SAZP, IMCS, TRAGSA, ARPA		
DOCUMENT HISTORY			
Version	Version date	Responsible	Description
0.1	25/07/2014	FBK	Draft version
0.2	08/08/14	FBK	After partners contribution
0.3	13/08/14	FBK	Tragsa contribution added
0.4	13/08/14	FBK	Acknowledges
0.5	13/08/14	FBK	Italian pilots information added
0.6	18/08/14	FBK	FMI contribution added
1.0	26/08/14	TRAGSA	Minor editorial changes

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Table of Contents

Table of Contents	3
List of Figures.....	4
List of Tables.....	4
1 Executive Summary	5
2 Introduction.....	5
3 Background. ISO/IEC SQuaRE	6
4 Evaluation Methodology	8
4.1 <i>Scope - Evaluation requirements</i>	9
4.2 <i>Purpose of the evaluation</i>	9
4.3 <i>Stakeholders identification</i>	10
4.3.1 <i>General Public</i>	10
4.3.2 <i>Data Resource (content) Providers</i>	10
4.3.3 <i>Data Service Developers</i>	10
4.3.3.1 <i>Consulting Services Operators</i>	11
4.3.4 <i>Administrators</i>	11
4.4 <i>Elements to be evaluated</i>	11
5 Evaluation Specification and Design.....	12
5.1 <i>Background</i>	12
5.2 <i>SmartOpenData Evaluation Specification and Design</i>	14
5.2.1 <i>Open Data Resources Evaluation Design. Quality Measurement</i>	15
5.2.1.1 <i>Satisfaction measures</i>	15
5.2.1.2 <i>Functional suitability measures</i>	16
5.2.1.3 <i>Usability measures</i>	17
5.2.1.4 <i>Reliability measures</i>	20
5.2.1.5 <i>Security measures</i>	20
5.2.1.6 <i>Maintainability measures</i>	21
5.2.2 <i>Open Dataset Evaluation Specification and Design</i>	22
5.2.2.1 <i>Open Dataset Evaluation Specification. Quality Model.</i>	22
5.2.2.2 <i>Open Dataset Evaluation Design. Quality Measurement.</i>	22
5.3 <i>Pilot specific Evaluation</i>	24
5.3.1 <i>Ireland Burren National Park Pilot</i>	25
5.3.2 <i>Slovakian Pilot</i>	28

5.3.3	<i>Italian Pilot</i>	30
5.3.4	<i>Spanish Pilot</i>	31
5.3.5	<i>Czech Republic Pilot</i>	32
6	Evaluation execution	34
6.1	<i>Preparation of evaluation</i>	34
6.2	<i>Execution of Evaluation</i>	35
6.2.1	<i>Questionnaires</i>	35
6.2.2	<i>Interviews with users</i>	36
6.3	<i>Feedback</i>	36
6.4	<i>Use feedback for further development</i>	37
7	Control groups	37
8	Plan of activities	38
9	Conclusions	38
10	References	39

List of Figures

Figure 1:	Organisation of SQuaRE series standards	7
Figure 2:	Software product quality evaluation reference model [ISO/IEC-25040]	8
Figure 3:	Software product quality evaluation reference model	8
Figure 4:	Relationship between quality measures types [ISO/IEC-25023]	12
Figure 5:	Quality in use model [ISO/IEC-25010]	13
Figure 6:	Product Quality Model of ISO/IEC 25010 [ISO/IEC-25010]	13

List of Tables

Table 1:	Elements to be evaluated by target users	11
Table 2:	Open Dataset evaluation metadata	22

1 Executive Summary

This document describes the evaluation plan and assessing of the Demonstration Pilots developed in WP5. Thanks to the Eurosentiment Project¹ some guidelines from its Evaluation Plan have been used in this deliverable.

Since the pilots are essentially software services based on high quality open data, the evaluation plan must provide an evaluation metric for both data and software quality. The plan follows an iterative verification & validation process, based on ISO/IEC SQuaRE Software Quality Evaluation standards.

The verification phase is done through an internal evaluation process of the artefacts developed in the project, both software and data resources. Software is evaluated through adoption of automated source code checking and automated unit testing tools. Data resources are evaluated through several metrics, such as evaluation of its quality in reference to other sources of similar data (think of 'official' issued data, when applicable), coverage of a domain and other quality indicators provided by the authors or the gatherers.

The validation phase will be carried out through an external evaluation process taking into account the potential users of the Pilots. The main activities of this phase are surveys to these kinds of potential users.

The evaluation plan finishes with an initial scheduling of the identified evaluation activities.

2 Introduction

The aim of the evaluation plan is to create a methodology that allows assessment of the project evolution and results. It determines how to measure the success of the project results, what kind of data will be collected and how they will be analysed.

The evaluation plan is part of WP6 Evaluation, Assessment and User Group. It consists of a set of reports and contributions to the other activities. In this sense, the plan defines how the other activities will perform the control and evaluation of the project objectives.

This plan also supports the quality of the project, providing valuable data to determine whether the direction of the project is correct and the results fulfil the users' needs. Therefore it makes possible the identification of risks in different steps of the project. In this plan, scientific and technical objectives are taken into consideration, both from the user and developer point of view.

Evaluating a project is an iterative process, of which the first step is the definition of the methodology. This methodology will be applied in an iterative way covering three different phases:

1

<http://eurosentiment.eu/>, Grant Agreement no: 296277

1. Collecting data: the partners & users provide data.
2. Evaluating the results: the data collected are compared with the metrics values.
3. Providing assessment: a report describing the results is submitted.
 - When the value is below a predefined threshold, the report will also provide recommendations to elaborate a recovery plan that allows enhancing the results.
 - When it is over that threshold, the recommendations of the report will be to follow the activities.

The information gathered as a result of the evaluation will be duly reported in the annual evaluation plans.

The remainder of this document is structured in order to describe what, who, how and when the results of SmartOpenData will be evaluated.

3 Background. ISO/IEC SQuaRE

The Pilots provided in SmartOpenData WP5 will consist mainly of software products as well as open dataset resources and formats. Our aim in this document is to propose an evaluation plan in order to monitor the quality of the SmartOpenData results.

The topic of software quality in the software life cycle has been widely discussed. As a result, the discipline of software product quality engineering is emerging with the aim of achieving the required quality of products through the definition of quality requirements and their implementation, measurement of appropriate quality attributes and evaluation of the resulting quality. The most popular standard for software product quality engineering is ISO/IEC 9126 Software engineering - Product quality. This standard has evolved to respond to feedback from practitioners and the academic community based on a survey conducted in 2001/2002. As a result, a second generation of quality standards, referred to as SQuaRE has been defined. Even though SmartOpenData results are not only software standards, the standard provides a well accepted conceptual evaluation framework based on the definition of quality models and measurements, that can be applied to the evaluation of Open Data resources. Thus, we have adopted this specification in order to develop the SmartOpenData evaluation plan.

The SmartOpenData Evaluation Plan is based on the IOC/IEC Software product Quality Requirements and Evaluation (SQuaRE)² set of standards. The general goal of the SQuaRE set of standards is to cover three complementary processes: requirements specification, measurement and evaluation.

The SQuaRE series of standards consists of the following divisions (Figure 1):

- **ISO/IEC 2500n Quality Management Division.** The standards that form this division define all common models, terms and definitions referred further by all other standards from SQuaRE series.

2

<https://www.iso.org/obp/ui/#iso:std:iso-iec:25000:ed-2:v1:en>

- **ISO/IEC 2501n Quality Model Division.** The standards that form this division present a detailed quality model including characteristics for internal, external and quality in use.
- **ISO/IEC 2502n Quality Measurement Division.** The standards that form this division include a software quality measurement reference model, mathematical definitions of quality measures, and practical guidance for their application.
- **ISO/IEC 2503n Quality Requirements Division.** The standards that form this division help specify quality requirements.
- **ISO/IEC 2504n Quality Evaluation Division.** The standards that form this division provide requirements, recommendations and guidelines for software product evaluation.

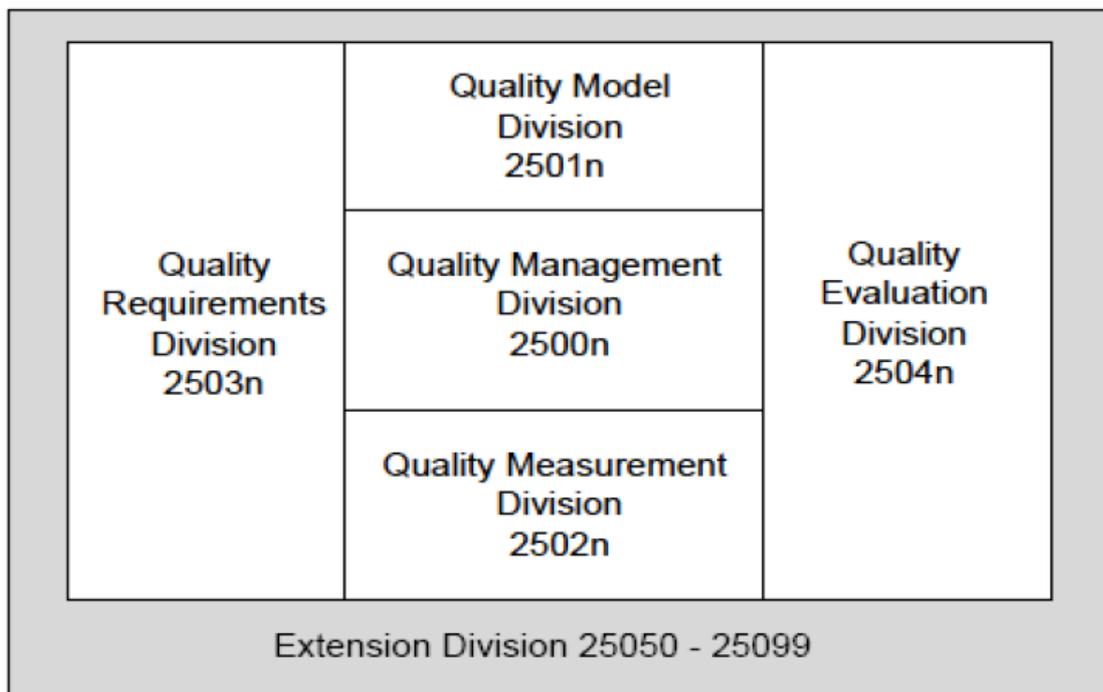


Figure 1: Organisation of SQuaRE series standards 39

4 Evaluation Methodology

The ISO/IEC 25040 “Systems and software Quality Requirements and Evaluation (SQuaRE) – Evaluation process” standard [ISO/IEC-25040] provides a reference model for software product quality evaluation that is depicted in Figure 2.

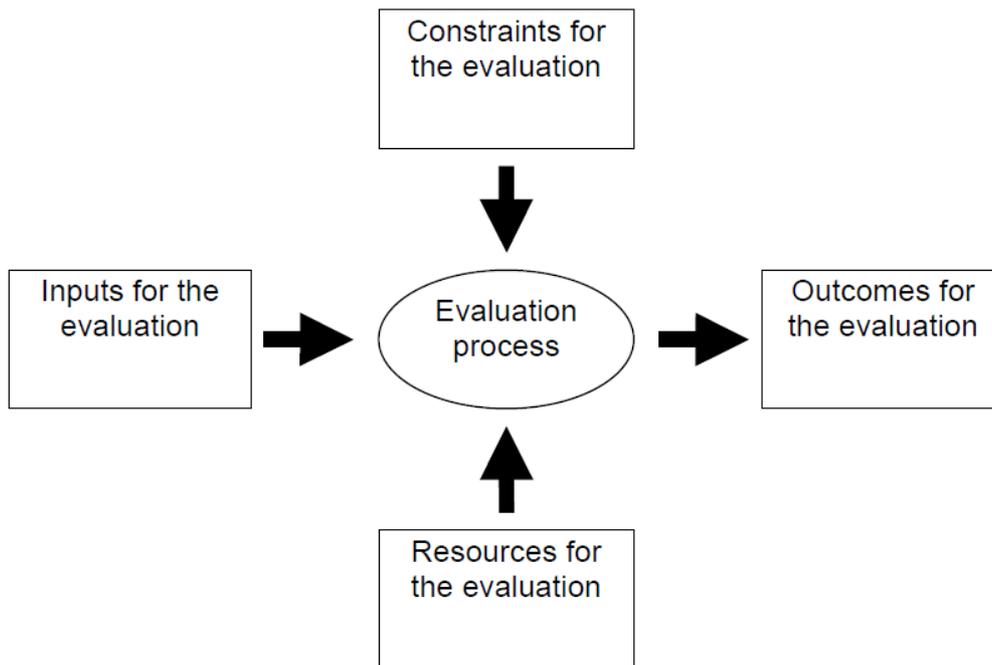


Figure 2: Software product quality evaluation reference model [ISO/IEC-25040]

The software product quality evaluation reference model intends that the evaluation should be based on a product software quality requirement specification by using ISO/IEC 25030 [ISO/IEC-25030] before the evaluation and making clear the objectives and criteria of evaluations. ISO/IEC-25030 provides requirements and recommendations for software product quality requirements specification and applies other SQuaRE specifications, such as ISO/IEC 25010 [ISSO/IEC-25010] and ISO/IEC 2502n.

The evaluation methodology reference model is shown in Figure 3.

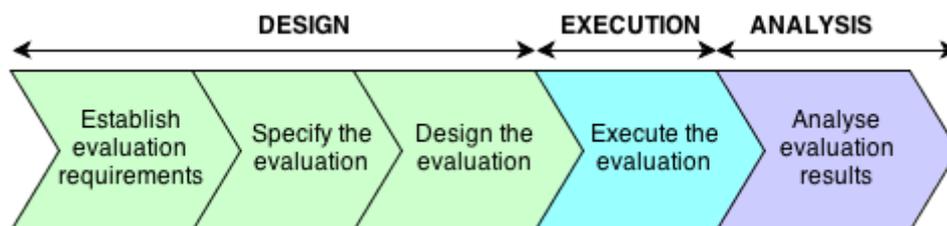


Figure 3: Software product quality evaluation reference model

This methodology is developed in the following sections, where some parts of the methodology are generic and thus applicable to all of the pilots, while the latest part regards some ideas or suggestions related to the specific pilot and are not applicable directly to other ones.

4.1 Scope - Evaluation requirements

The first step of the evaluation model is the definition of evaluation requirements. This can be achieved by addressing the following steps [ISO/IEC-25040]³:

- Establish the purpose of the evaluation. The first step is identifying what is the purpose of the evaluation, such as decide on the acceptance of the project, and decide when to release the product, compare the product with competitive products or select a product from among alternative products.
- Identify stakeholders of the software product.
- Obtain the software product quality requirements, using a quality model.
- Identify product parts to be included in the evaluation. Identify the intermediate software products to be included in the evaluation, such as software modules, design diagrams or test specifications.
- Define the stringency of the evaluation. The extent to which the quality evaluation covers the specified software quality requirements, taking into account evaluation budget, target date for the evaluation, purpose of the evaluation and use criticality of the software product.

4.2 Purpose of the evaluation

The SmartOpenData evaluation purpose is to provide assessment of the project progress so that the products developed in the project meet the required quality to reach the expected impact, according to the SmartOpenData objectives.

The ultimate purpose of this evaluation plan is to define how the assessment of the project results will be carried out. In order to ensure that the project results fulfil the objectives proposed, the users of each intermediate tool and demonstrator within the SmartOpenData project will be requested to evaluate the status of the work. In this regard, two main elements will be assessed: resources quality and overall system usefulness.

The methodology will provide the tools (guidelines to create the questionnaires and interviews to be used for the evaluation and assessment). Moreover the methodology will identify and select the means of getting the appropriate information in order to measure the acceptance of all the targeted groups once the pilot studies are completed. The following sections try to answer the questions “who should and will evaluate the system?”, “what do we want of the final system?”, and “what are the indicators and elements for evaluation?”

3

Steps 2 and 3 are proposed together in ISO/IEC-25040, they have been divided into two steps, so that the quality model is further developed in a different section given its extension.

4.3 Stakeholders identification

The different nature of the users of the platform highly influences the way the evaluation is performed, and how the information will be retrieved from each of the different groups. In the following sections we will give some details on how each type of user differs from the rest, and how to approach them.

4.3.1 General Public

This could be the easiest target for evaluation (depending on how the evaluation would be communicated to attract this kind of shareholder), as there isn't any requisite or condition, and their evaluation can be self-contained. These users can be asked to submit response to a survey and rate content and ease of use while using the platform. They can provide useful information about the results of the services: accuracy, relevance, etc. It can help measure the adequacy of the data models used, the state of the technology and the relevance of the resources and services available through the platform.

4.3.2 Data Resource (content) Providers

For these stakeholders, it will be vital to be able to provide resources to the platform and other users in the easiest way, while keeping control of their resources. Apart from the actions already specified for the General Public, Data Resource Providers should be interviewed and looked after more closely, in order to assess the final usability and features from their standpoint.

There are several types of resources, and different types of owners, so the test cases must be adapted to a wide range of them in order to provide the most accurate and complete evaluation. Usage statistics and questionnaires have to be made in accordance to the type of resource, and always include such information to be able to generate proper results.

Due to the nature of these users, their use of the SmartOpenData Pilots will be more sporadic than constant. Thus, evaluation will be mainly done in specific evaluation sessions (the demonstration pilots). Hence the test cases can be better elaborated and quickly gone through.

4.3.3 Data Service Developers

The key difference of a Data Service Developer from the General Public is that they access the information through APIs and other advanced methods instead of the usual Web User Interface. In other words, they represent the role of resources transformers generating added value on top of the resources published by Data Resource Providers. In this case, we will have to also cover the ease of access to the information via their different frameworks, programming languages and other tools.

4.3.3.1 Consulting Services Operators

The role of consultants in the demonstration pilots could be orthogonal to the other kind of stakeholders and their contribution to evaluation could be in providing more insight on quality measures of ease of use. In addition this type of stakeholders can help evaluation of fitness of purpose as well as the wider impact of the project outcomes.

4.3.4 Administrators

To ensure consistent communication framework with the internal project consortium as well as external stakeholders close collaboration will be established with the activities of tasks T6.2, user groups set up and analysis and T6.3 User group maintenance.

4.4 Elements to be evaluated

There are three main elements to be assessed:

- The quality of the open data resources and the suitability of the Linked Open Data model and Geospatial Information;
- the facilities provided by the SmartOpenData consortium to access existing services and resources in the SmartOpenData project;
- The SmartOpenData Demonstration Pilot. Its usefulness and features exposed to all the target groups. This obviously includes several roles and sections of the pilots.

Each of the different groups that we will be targeting with the SmartOpenData pilot platforms will be focused on different aspects, which will translate into different requirements and means of evaluation.

To give a better insight into the different points each of the groups will most probably be interested in, and what kind of elements within the platform they are linked to, we provide the following table. However, it is not intended to be an in-depth description, but a set of highlights.

Group	Elements to evaluate
General Public	<ul style="list-style-type: none"> • Ease of use and accessibility • Accuracy and quality of the results, added value • Protection of content, security of communications • Consults, ease of use and accessibility for the general public • Different services provided
Data Resource Providers	<ul style="list-style-type: none"> • open data resources management and visibility • Treatment of assets • Visibility
Data Service Developers	<ul style="list-style-type: none"> • Ease and means of integration • Quality of service (QoS)
Consulting Services Operators	<ul style="list-style-type: none"> • Accuracy and quality of the results • General Quality
Administrators	<ul style="list-style-type: none"> • Control, management and traceability

Table 1: Elements to be evaluated by target users

5 Evaluation Specification and Design

5.1 Background

The standard ISO/IEC 25010, Systems and Software Quality Requirements and Evaluation (SQuaRE) [ISO/IEC-25010] defines a quality in use model and a product quality model with the aim of providing a consistent terminology for specifying, measuring and evaluating system and software product quality.

The quality models of a system define the degree to which the software system satisfies the stated and implied needs for its various stakeholders, and thus provides value. The standard SQuaRE represents these stated and implied needs by quality models that categorize product quality into characteristics, which in some cases are further subdivided into sub-characteristics. The measurable quality-related properties of a system are called quality properties, with associated quality measures. To arrive at measures of quality characteristic or sub-characteristic, it might be possible to measure one or more quality properties corresponding to that quality characteristic or sub-characteristic.

The ISO/IEC 2504n series of standards provide methods for measurement, assessment and evaluation of system or software product quality during the software life cycle. They are intended for use by developers, acquirers and independent evaluators, particularly those responsible for system or software product evaluation.

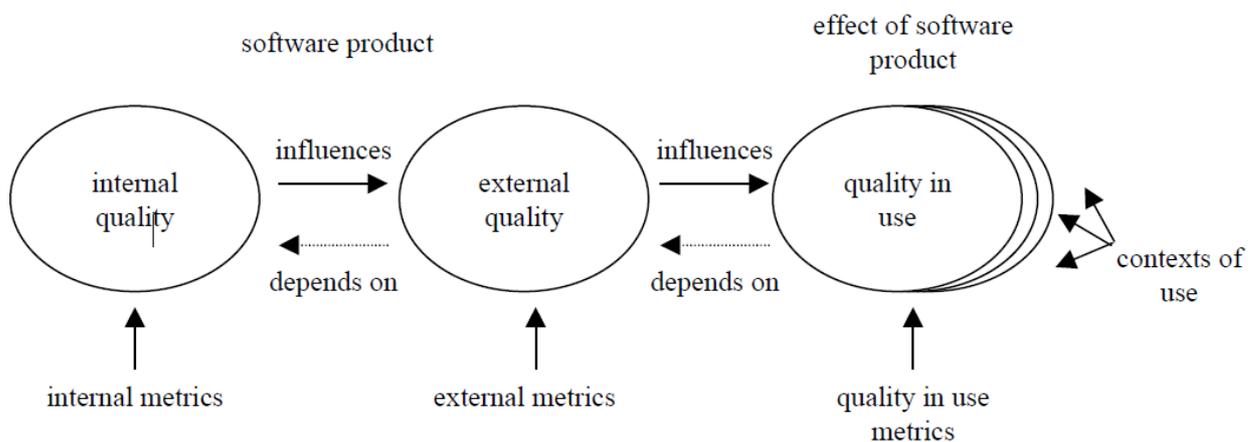


Figure 4: Relationship between quality measures types [ISO/IEC-25023]

The internal quality measures may be applied to a non-executable system / software product during its development stages (such as request for proposal, requirements definition, design specification or source code). Internal quality measures provide the users with the ability to measure the quality of the intermediate deliverables and thereby predict the quality of the final product.

The **external quality measures** may be used to measure the quality of the system / software product by measuring the behaviour of the system of which it is a part. The external quality

measures can only be used during the testing stages of the life cycle process and during the operational stages.

The **quality in use quality measures** measure whether a product meets the needs of specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in a specified context of use. This can only be achieved in a realistic system environment.

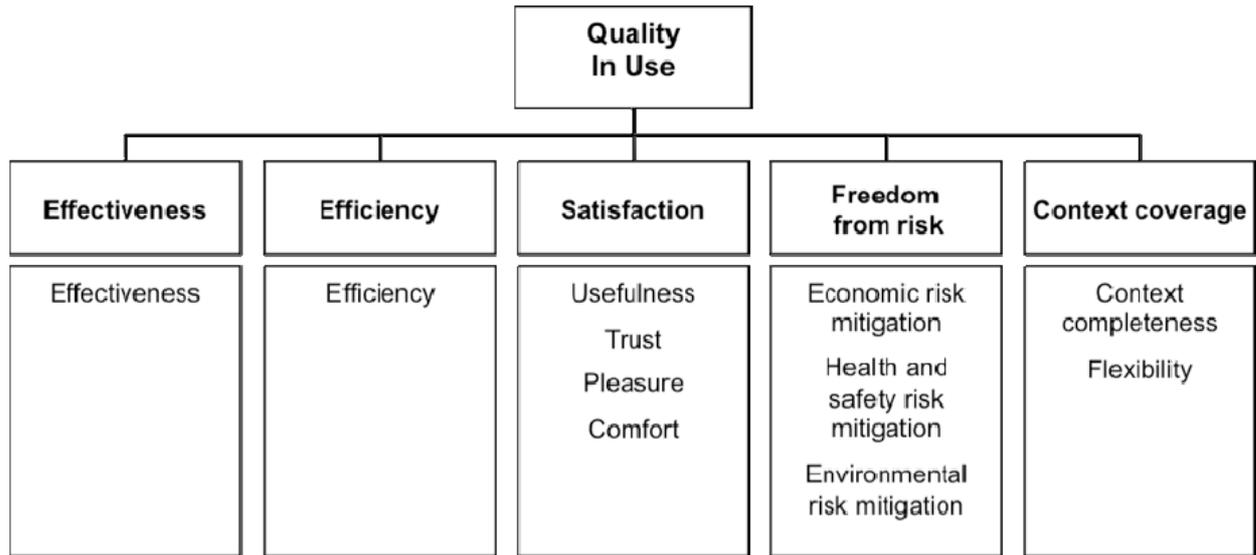


Figure 5: Quality in use model [ISO/IEC-25010]

The quality in use model (Figure 5) defines five characteristics (effectiveness, efficiency, and satisfaction, freedom from risk and context coverage) which are subdivided into sub-characteristics. The product quality model of SQuARE ISO/IEC 25010 [ISO/IEC-25010:2011] defines eight characteristics (functional suitability, performance suitability, compatibility, usability, reliability, security, maintainability and portability) which are further subdivided into sub-characteristics, as shown in Figure 6.

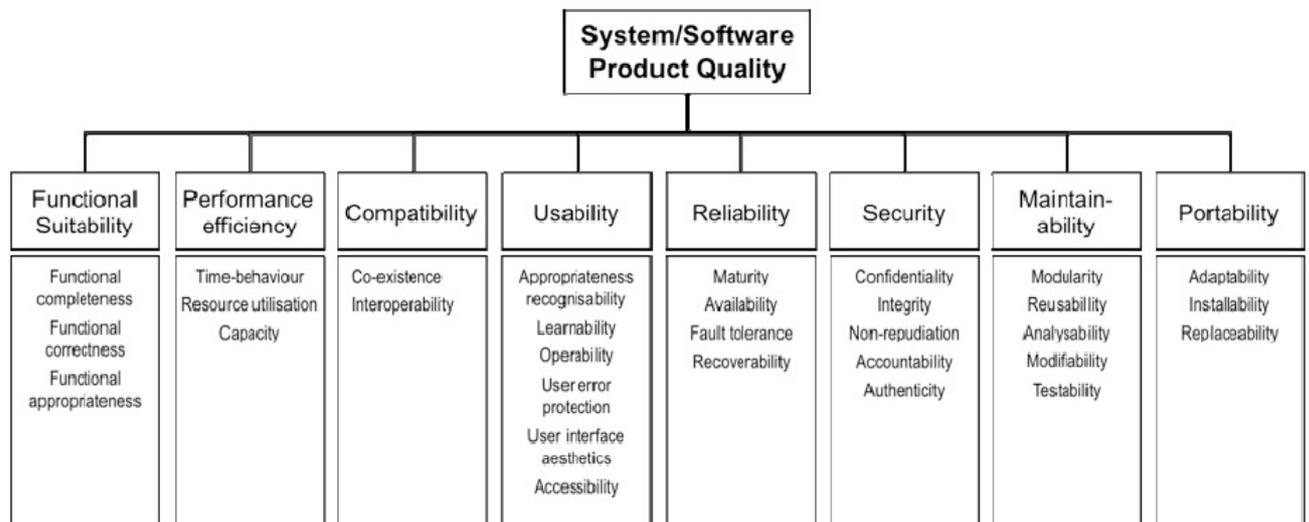


Figure 6: Product Quality Model of ISO/IEC 25010 [ISO/IEC-25010]

The set of standards ISO/IEC 2502n define the Quality Measurement Division. ISO/IEC 25020 provides a reference model and guide for measuring the quality characteristics defined in ISO/IEC 2501n Quality Division Model. The associated standards provide suggested measures of quality throughout the product life cycle. ISO/IEC 25021 offers Quality Measure Elements (QME) that can be used to construct software quality measures. ISO/IEC 25022 provides measures for the characteristics in the Quality in Use Model defined in ISO/IEC 25010. ISO/IEC 25023 provides measures for the characteristics in the product quality model defined in ISO/IEC 25010. Finally, ISO/IEC 25024 provides measures for the characteristics in the data quality model defined in ISO/IEC 25012.

In the next section, a Quality Model and Quality Measurement is proposed, following the ISO/IEC 25000 guidelines.

5.2 SmartOpenData Evaluation Specification and Design

Users need and demand that some basic requirements are met in order to be able to use the system, and to be satisfied with the results. Those requirements enforce the adoption of the principles in the developed system:

1. Accessibility, visibility and search-ability of the resources
2. User interaction with the system by means of feedback
3. Localisation of the tools (multilingual user interfaces)
4. Multilingual access to quality resources in their own language (when applicable)

On the other hand, content providers and service developers will use the platform as well, which leads to the following criteria:

5. Secure communication channels
6. Well defined and visible authorship and privileges

In particular, effectiveness, efficiency and satisfaction will be evaluated. With regards to freedom from risk and context coverage have been excluded from the evaluation plan since most of them are not applicable, with the exception of economic risk mitigation that even though it is interesting from evaluating the exploitability of the project, it should be evaluated once the project is finished.

A key issue to evaluate in the pilot trials is the confirmation of the assumptions made on the sustainable continuation and exploitation of the pilots' services after the project ends.

Regarding the Product Quality Model, taking into account that the aim of SmartOpenData is to provide a research prototype that will be exploited after the project's conclusion, the evaluation focus will be put on the quality attributes that lead to stakeholders acceptance of the Open Dataset Pool, as well as the software quality of the developments within the project.

5.2.1 Open Data Resources Evaluation Design. Quality Measurement

The standard ISO/IEC 25022 [ISO/IEC-25022] defines quality measures categorised by the characteristics and sub-characteristics of the Quality in Use model defined in ISO/IEC 25010 [ISO/IEC-25010].

5.2.1.1 Satisfaction measures

Satisfaction measures assess the degree to which user needs are satisfied when a product or system is used in a specified context of use. For SmartOpenData, there are three identified measures for satisfaction evaluation measurement: usefulness, trust and pleasure.

Usefulness measures assess the degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Satisfaction scale	How satisfied is the user?	$X = A / B$ A = questionnaire producing psychometric scales B = population average	Calculated	External	Questionnaire
Satisfaction questionnaire	How satisfied is the user with specific system features?	$\chi = \sum A_i / n$ A = response to a question n = number of responses	Calculated	External	Questionnaire

Trust measures assess the degree to which a user or other stakeholder has confidence that a product or system will behave as intended.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Trust scale	Does the user trust the system?	$X = A / B$ A = questionnaire producing psychometric scales B = population average	Calculated	External	Questionnaire
As above	Does the user trust the LOD data sources?	As above	As above	As above	As above
As above	Does the user	As above	As above	As above	As above

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
	trust the Geographic sources?				

Pleasure measures assess the degree to which a user obtains pleasure from fulfilling their personal needs.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Pleasure scale	Does the user obtain pleasure from using the system?	$X = A / B$ A = questionnaire producing psychometric scales B = population average	Calculated	External	Questionnaire

Moreover the standard ISO/IEC 25023 [ISO/IEC-25023] defines quality measures categorised by the characteristics and sub-characteristics of the Product Quality model defined in ISO/IEC 25010 [ISO/IEC-25010].

5.2.1.2 Functional suitability measures

Functional suitability measures should be able to measure the degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions.

Functional completeness measures should be able to measure the degree to which the set of functions cover all the specified tasks and user objectives.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Functional implementation coverage	How complete is the implementation according to requirements specifications	A = number of functions stated in requirement specification B = number of missing or incorrect functions	Count / Count	External / Internal	Trace implemented use cases .

5.2.1.3 Usability measures

Usability measures should be able to measure the degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

SmartOpenData Open Datasets will be evaluated using a sample of users which are representative of the previously identified user groups, as well as by surveys in evaluation sessions once the demonstrators pilots are available.

Appropriateness recognisability measures assess whether new users can understand whether the software is suitable and how it can be used for particular tasks. The following measures are defined.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Description completeness	What proportion of functions are described as understandable in the product description?	A = number of functions described as understandable B = total number of functions	Count / Count	External	Questionnaire
Demonstration capability	What proportion of functions requiring demonstration have such capability?	A = Number of functions implemented with demonstration capability B = total number of functions requiring demonstration capability	Count	External	Questionnaire

Learnability measures should be able to measure the degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use.

Learnability measures should be able to assess how long users take to learn how to use particular functions, and the effectiveness of help systems and documentation.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Completeness of user documentation and/or help facility	What proportion of functions are correctly described in the user documentation and/or help	A = number of functions described correctly B = total number of functions	Count	External	Questionnaire

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Auto-learning capability	facility What proportion of functions are correctly executed by users?	A = number of correctly finished tasks B = total number of tasks	Count	External	Questionnaire

User Error Protection Measures should be able to measure the degree to which the system provides users against making errors.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Input validity checking	What proportion of input items provide checking for valid data?	A = number of input items checked for valid data B = number of input items which need checking for valid data	Count / Count	External / Internal	Interface inspection
Avoidance of incorrect operation	How many functions have incorrect operation avoidance capability?	A = number of functions implemented to avoid critical or serious malfunctions being caused by incorrect operation B = total number of incorrect operation patterns	Count / Count	External / Internal	Interface inspection

Accessibility measures should be able to measure the degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use. The range of capabilities includes disabilities associated with age.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Physical accessibility	What proportion of functions can a user with a physical	A = number of functions accessible by the disabled person	Count / Count	External / Internal	Accessibility checking tool. Level WAI-AA ⁴ will be required.

4

<http://www.w3.org/WAI/>

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
	handicap access	B = total number of functions implemented			Suggested WCAG 2.0 ⁵

Multilingualism measures should be able to measure the degree to which the user interface is consistently localised in several languages.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Multilingual accessibility	What proportion of functions are consistently provided in the user language	A = number of functions correctly localised B = total number of functions implemented	Count / Count	External / Internal	Inspection

Fault tolerance measures should be able to measure the degree to which a system, product or component operates as intended despite the presence of hardware or software faults.

When software components complexity is high, fault proneness increases rapidly [El-Emam02]. So, developers and maintainers use SLOC (Source Lines of Code) in order to determine understandability of code, in conjunction with other source code metrics, such as McCabe cyclomatic complexity or Nested Block Depth (NBD). These metrics are usually measured by using automatic tools. Examples of open source products are Checkstyle⁶, PMD Tool⁷ or CPD⁸. Checkstyle is a tool to help programmers write Java code that adheres to a coding standard. It automates the process of checking Java code to spare humans of this boring (but important) task. This makes it ideal for projects that want to enforce a coding standard. Checkstyle is highly configurable and can be made to support almost any coding standard. An example configuration file is supplied supporting the Sun Code Conventions. Moreover, there are Checkstyle plug-ins available for NetBeans (nbCheckstyle) and for Eclipse (eclipse-cs).

Depending on the language chosen for the pilot codification, other tools are available, e.g. for python pylint⁹, pychecker¹⁰, pyflakes¹¹ are available and they perform different

5

<http://www.w3.org/TR/WCAG20/>

6

<http://checkstyle.sourceforge.net/>

7

<http://pmd.sourceforge.net/>

8

<http://pmd.sourceforge.net/pmd-5.0.2/cpd.html>

9

<http://www.pylint.org/>

10

<http://pychecker.sourceforge.net/>

compliance checks on the code (e.g. pep8¹² style checking among others).

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Software complexity	How complex is the software?	- Cyclomatic complexity - KLOC	Count	Internal	Language dependent tool
Code readability	How is easy to read the codebase according to a standard	Style violations	Count	Internal	Language dependent (e.g. pep8 for python)
Common errors avoiding	How probable that known and trivial errors are present in the code	Common violations in a particular tool	Count	Internal	Language dependent (e.g. pylint violations)

5.2.1.4 Reliability measures

Reliability measures should be able to measure the degree to which a system, product or component performs specified functions under specific conditions for a specific period of time.

Internal reliability measures are used for predicting if the system / software product in question will satisfy prescribed reliability needs, during the development of the system / software product.

5.2.1.5 Security measures

Security measures should be able to measure the degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access appropriate to their types and levels of authorisation.

Confidentiality measures should be able to measure the degree to which a product or system ensures that data are accessible only to those authorised to have access.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Access	How controllable	A = number of	Count / Count	External /	Penetration tests

11

<http://pychecker.sourceforge.net/>

12

<http://legacy.python.org/dev/peps/pep-0008/>

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
controllability	is the accesses to the system?	detected different types of illegal operations B = number of types of illegal operations in the specification		Internal	

Accountability measures should be able to measure the degree to which the actions of an entity can be traced uniquely to the entity.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Access auditability	How complete is the audit trail concerning the user access to the system and data?	A = number of accesses to system and data recorded in the system log B = number of accesses actually occurred	Count / Count	External / Internal	13 JUnit

5.2.1.6 Maintainability measures

Maintainability measures should be able to measure the degree of effectiveness and efficiency with which a product or system can be modified by the intended maintainers.

Testability measures should be able to measure the degree of effectiveness and efficiency with which test criteria can be established for a system, product or components and tests can be performed to determine whether those criteria have been met.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Functional completeness of embedded test functions	How completely are test functions and facilities implemented.	A = number of test functions implemented as specification B = number of required test functions	Count / Count	Internal	JUnit
Autonomous	How	A = number of tests	Count / Count	Internal	Mock Unit Testing,

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
testability	independently can software be tested	that can be simulated by stub out of the tests depending on other systems			with tools such as 14 Mockito or 15 SOAPUI .

5.2.2 Open Dataset Evaluation Specification and Design

5.2.2.1 Open Dataset Evaluation Specification. Quality Model.

The project aims at developing a large shared data pool for open data resources meant to be used by the pilots, in order to bundle together what could be scattered resources. In the table below are listed a set of quality measures that should be evaluated for each data source that is leveraged by the demonstration pilot.

Quality property	Description
Multilinguality	Languages targeted by the resource (if applicable)
Compliance	Compliance with standards, e.g. XML well-formed, INSPIRE data specifications
Specificity	How a dataset succeeds in giving the information specific to the application of the pilot
Quality indicators	- Number of features - Popularity (number of downloads or number of users) - Evaluation method of the resource (performance and uptime in case of services, statistical measure errors in case of sampling ...)
Licence	Licence of the resource

Table 2: Open Dataset evaluation metadata

5.2.2.2 Open Dataset Evaluation Design. Quality Measurement.

The evaluation of a dataset leveraged by the pilot will follow these steps:

1. If the dataset resource license is not compatible with the (at the moment, to be decided) SmartOpenData licensing model, the resource will not be accepted.
2. If the resource is not compliant with the SmartOpenData model, the resource cannot be directly integrated. The first option is that the resource provider would adapt its resource to SmartOpenData model. In case this is not feasible, the SmartOpenData business manager will analyse the business interest in its adaptation and could propose a commercial agreement for this. In case the resource is not adapted, the resource will not be accepted.

14

<http://code.google.com/p/mockito/>

15

<http://www.soapui.org/>

3. Other quality indicators, such as popularity or size, will provide insight to end users for its usage.

Multilinguality measures should be able to measure how several languages are represented. This is not always applicable where metadata are absent or not applicable (pure geographic point with poor or no text data within).

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Multilinguality singularity	Does the Open Dataset contribute to provide open data resources in new languages?	1 if this is the only resource available in one language 0 otherwise	Count	Internal	Inspection

Compliance measures should be able to measure how an open data resource can be easily integrated in the Open Dataset pool.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
SmartOpenData model compliance	Is the Open Dataset available in a compatible format?	1 if the resource is compliant with SmartOpenData accepted formats 0.5 if the resource is based on a open standard easy to convert 0.25 if the resource is based on a closed std 0 otherwise	Number	Internal	Inspection

Specificity measures should be able to measure how many topics are covered in the Open Dataset pool.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Specificity singularity	Does the Open Dataset contribute to provide open data resources in new topics?	1 if this is the only resource that addresses that topic 0 otherwise	Count	Internal	Inspection

Quality measures should be able to measure how the Open Dataset has been created.

Measure Name	Description	QMEs	Measure Type	Measurement focus	Tool
Development method	What is the quality of the method to obtain the resource?	1 if this is the only resource available in one language 0 otherwise	Count	Internal	Inspection
Size	What is the size of the resource?	Number of Features, or POIs (cardinality)	Count	Internal	Inspection
Popularity	Is the resource good because there other users trust on it?	N = number of users using the resource or number of downloads	Count	Internal	Inspection
QA Method	What is the precision of the resource?	N = Precision provided by the Dataset owner	Count	Internal	Inspection

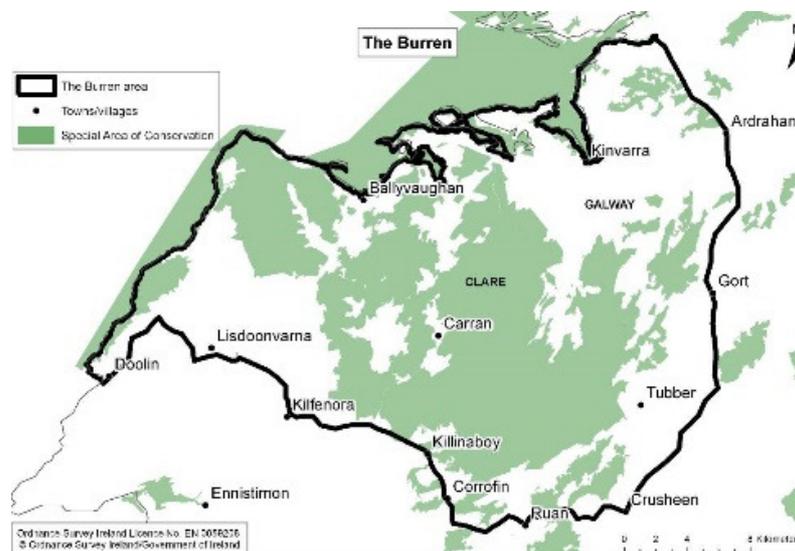
5.3 Pilot specific Evaluation

The 5 pilots have different areas of interest and different environmental topics, and may have different specific quality and evaluation measures that could be interesting to provide. This chapter provides an evaluation methodology based on the objectives defined for each pilot in the D5.1 Rationale of the Pilot, Evaluation set-up and test metrics. In addition, where relevant, this document also addresses the Quality metrics defined in D 1.1 Quality Assurance Plan

Pilot	Scenario / Users	Evaluation Peculiarity
1. Spain & Portugal	Agroforestry Management	Focused on the needs of forest owners, foresters, public administrations, seed-harvesting companies and seed nurseries in order to obtain the best species for a given plot and select the best seeds.
2. Ireland	Environmental Research and Biodiversity	Focused on the needs of the decision makers, researchers and other stakeholders in managing and sustaining the Burren National Park.
3. Italy	Water Monitoring	Emphasis on stakeholder engagement in data publishing and semantic interoperability between datasets from a range of dataowners.
4. Czech Republic	Forest Sustainability	Focused on sharing of the data of NFI using SmOD principles for the public sector (ministries, universities, research institutions), foresters, private companies in forestry sector, general public etc.
5. Slovakia	Environmental Data reuse.	Addressing creation of spatial linked data and their further re-use including the possibilities to improve search for environmental geospatial data.

5.3.1 Ireland Burren National Park Pilot

The Irish Pilot will focus on the use of the SmartOpenData infrastructure to provide open data and open INSPIRE-compliant geospatial sources for environmental researchers particularly focused on biodiversity and habitats, building on participative social validation and pilots. Focusing on European protected areas and its National Parks, starting with the Burren National Park in Ireland.



The pilot aims is to demonstrate the value of SmartOpenData in helping Researchers and Decision Makers to better manage, preserve, sustain and use this unique ecosystem in 4 user scenarios:

1. SmartOpenData enabled European Tourism Indicator System (ETIS) Webservice for the Burren & European GeoParks Network.
2. SmartOpenData enabled Farming for Conservation Webservice
3. SmartOpenData enabled App to Ground-Truth potential Protected Monument sites
4. SmartOpenData Platform input to the Irish Open Government Partnership (OGP) process

These four User Scenarios were identified as being most beneficial to the various stakeholder groups in the Burren from various meetings and discussions with them, as explained in D5.1. These User Scenarios will be complemented by various social networking and crowd sourcing mobile apps to engage stakeholders at the local level in particular.

The stakeholders will know that the Irish Pilot's intended outcomes have been achieved when the 4 intended user scenarios are operational and being used by their users, with the functionality as described in D5.1, involving

1. User engagement – as a first step in validating the value of the Irish Pilot's services to its intended users.
2. Direct user interaction with the open data access process – as the next step in user involvement with the GI/LOD sources.
3. Co-design of innovative “demand pull” services – the ultimate engagement of the stakeholders to evolve the Irish Pilot's service beyond the project, and use the SmartOpenData platform to create new opportunities, and in turn sustain the platform.

The Irish Pilot will demonstrate achievement of its intended outcomes mainly through User Social Validation, which will use criteria and indicators of success according to the different standpoints of the actors represented in each usage scenario, as a framework for evaluating the added value of the services that conform to the standards proposed by SmartOpenData.

The indicator sets defined here will be integrated into the composite list of evaluative questions to be used for the pragmatic assessment of impact generated by the Apps and services enabled by the SmartOpenData platform for each of the scenarios – and more broadly, on the environmental related activities users are involved in.

The mapping of each validation approach to each scenario in the Irish Pilot will be broadly as follows:

Pilot & Validation approaches.	Validation driven by the prospect of user engagement	Validation through direct user interaction with the open data access process	Validation driven by the co-design of innovative "demand pull" services
1. SmartOpenData enabled ETIS Webservice for the Burren & European GeoParks Network.	X	X	X
2. SmartOpenData enabled Farming for Conservation webservice	X	X	
3. SmartOpenData enabled App to Ground-Truth potential Protected Monument sites	X		X
4. SmartOpenData Platform input to the Irish Open Government Partnership (OGP) process	X	X	

The Criteria of Success of the Irish Pilot's Scenarios will be as follows:

1. Usage level and User Validation of the Irish Pilot's Services that use SmartOpenData
2. Increased access to harmonised and interoperable GI, L/OD and VGI data
3. Integrate data from users', OD, crowd-sourced and social media.
4. Integration of VGI into existing SDIs and LOD
5. Easy collection of information using smart phones and LOD
6. Reuse and share tourist information resources, channels and tools
7. New tourism activities, visitors and jobs, and new SME developed Apps and Services.

The following table summarises an initial list of the main criteria to be achieved in the Irish pilot by the end of the project in October 2015:

Irish Pilot Scenarios.	Apps/ Services in Operation	No of App/ Svc Users	No GI/LOD datasets in use	No VGI datasets created	Monthly access-es	New Apps/ Svcs
1. SmartOpenData enabled ETIS Webservice for the Burren & European GeoParks Network.	2	15	10	2	100	2
2. SmartOpenData enabled Farming for Conservation webservice	1	10	5	0	25	0
3. SmartOpenData enabled App to Ground-Truth potential Protected Monument sites	1	10	5	1	50	2
4. SmartOpenData Platform input to the Irish Open Government Partnership (OGP) process	1	15	20	2	25	0
Total for Irish Pilot	5	50	40	5	200	4

The final targets for each will be refined and agreed in line with the project's overall targets, as they evolve through the formative and summative evaluation phases, as defined in this Evaluation Plan, and access to various sources by collaborating with, and building on various open data and geo-spatial sources and initiatives that will have a particular value for biodiversity researchers, including:

- The European Biodiversity Observation Network, EUBON project¹⁶ ;
- European Environmental Agency (EEA), Biodiversity data centre (BDC)¹⁷ ;
- PESI¹⁸ ;
- FP7 EUBrazilOpenBio¹⁹ ;
- LifeWatch European research infrastructure²⁰ ;
- The Joinup Portal²¹ ;
- EU Open Data Portal²² ,
- The UK environmental Agency's Datashare²³ ,
- The Global Biodiversity Information Facility (GBIF)²⁴ .

5.3.2 Slovakian Pilot

Regarding the Slovakian pilot the specific evaluation methodology should be extended to provide also indication, how objectives addressed by this pilot are going to be evaluated.

There are two main objectives addressed within the Slovakian Environmental data reuse pilot. The first one is aiming to investigate possibilities to improve search for environmental geospatial data. Second outcome is related to the challenges related with creation of spatial

16 www.earthobservations.org/geobon.shtml

17 www.eea.europa.eu/themes/biodiversity/dc

18 www.eu-nomen.eu/portal

19 www.eubrazilopenbio.eu

20 www.lifewatch.eu

21 <http://joinup.ec.europa.eu/catalogue>

22 <http://open-data.europa.eu/en/data/dataset>

23 www.geostore.com/environment-agency/WebStore

24 www.gbif.org

linked data and their further re-use.

Based on that two subpilots were proposed:

- 1. Biodiversity MashUp Linked Open Data Extension:** Focused on linked geo data creation and reuse in decision making activities.
- 2. Spatial Web Crawler:** Aiming on possibilities to discover spatial information resources through spatial data infrastructure as well as linked data interfaces including the possibility to display and view of discovered spatial data.

In order to fulfill the above indicated objectives and to address the requirements of spatial and non-spatial environment related data producers and consumers, particularly in connection with the SmartOpenData project aims and where possible supporting existing data and tools re-use four use cases were designed aiming to delineate following requirements:

1. How to create spatial linked data? (Support and guidance in creating environmentally related spatially enriched linked data resources).
2. Where to search and discover available spatial data? (Provide the possibility to make existing geospatial resources discoverable for users and vice versa to allow potential searchers to find geospatial content within the SDI, wide web and semantic web infrastructures).
3. Display and view of discovered spatial data (Allow users to see the results of their discovery to help them decide, whether what they have found fits their needs).
4. Help citizens and decision makers to investigate, what kind of biodiversity potential as well as environmental risks can be identified in their area of interest. (Through this use case users will be able to create simple queries on the area of their neighbourhood or the area they will define and receive relevant information about the valuable areas as well areas representing potential risks to the environment (e.g. contaminated sites)).

Use cases 1 and 4 are going to be deployed via sub-pilot 1 Biodiversity Mashup v.2,0, whilst use cases 2 and 3 through the other 2. Spatial Web Crawler subpilot.

In order to be able to ensure the above mentioned objectives will be met the following set of pilot specific criteria and their target values are proposed:

Slovak Pilot Use cases.	No. of datasets	No. of md records	No of APIs	No. of applications	No. of users
1. How to create spatial linked data?	5	5	2	0	1
2. Where to search and discover available spatial data?	0	100-1000	4	1	20
3. Display and view of discovered spatial data	25	0	1	1	20
4. Help citizens and decision makers to investigate, what kind of biodiversity potential as well as environmental risks can be identified in the area of their interest.	10	10	1	1	10

5.3.3 Italian Pilot

Regarding the Italian pilot the specific evaluation methodology should be extended to provide information about the Living Lab methodology for engagement of different stakeholders in the process of providing data and exploring the inter-relationships between datasets. The Italian Pilot is based on a unique pilot application domain air and water quality monitoring. The possible scenarios for development start from a “baseline” scenario of engagement of different institutional actors (ARPA, Municipal and Provincial authorities, Health authorities, etc.). All these stakeholders are holding data relevant to the monitoring and management of air and water quality.

From there, the pilot is open to the engagement of other stakeholders (business and sector associations, environmental NGOs, hospitals, businesses and citizens, etc.) who may be holding or producing data relevant to explaining the causes and effects of air and water pollution in relation to the data captured by ARPA. This user-driven approach will require a more nuanced classification of targets of the evaluation exercise, since the “general public” and “data providers” categories overlap significantly. (Indeed, the degree to which they overlap is one of the main evaluation questions of the pilot).

This open-ended approach thus determines a series of specific evaluation questions:

What is the number and variety of stakeholders engaged in the pilot activities and contributing by opening up their data?

Are the stakeholders able to properly format their data and publish it as Open Data?

Do the stakeholders explore relationships between datasets that can enhance and enrich the SmartOpenData semantic framework?

Is ARPA able to publish reports on air and water quality in a more timely and/or accurate fashion?

Does ARPA gain increased ownership of its mission of collecting and disseminating meaningful data?

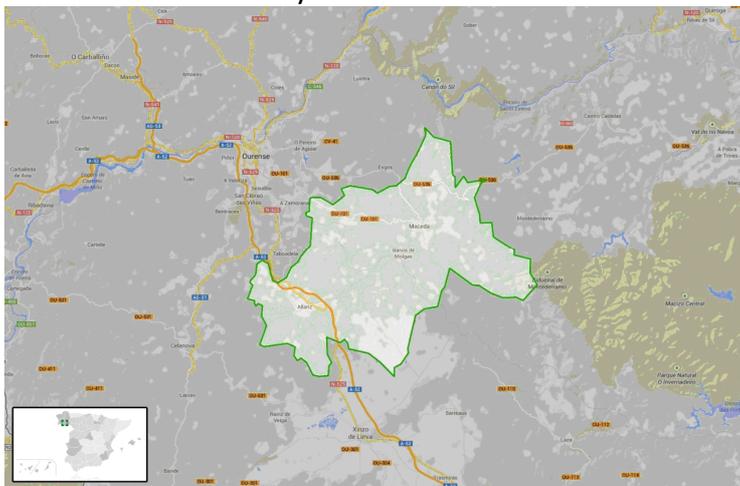
Is there an increase in the usage of ARPA's services by both internal (to the Regional government), institutional and external stakeholders?

Do the functionalities and facilities "returned" by the SmartOpenData infrastructure provide a meaningful and useful support to the issues of interest to the stakeholders.

To answer some of these questions some more informal and qualitative evaluation methods can be used to supplement those already planned, such as reports on user participation in workshops, etc.

5.3.4 Spanish Pilot

This pilot geographically comprises the municipalities of Allariz and Maceda, in the region of Galicia, Northwest Spain. Area management is mainly based on the forest. Tourism and urban planning in this area are secondary.



Location of the Maceda-Allariz area

The existing stakeholders are forest owners, foresters, public administrations, seed-harvesting companies and plant and seed nurseries. Such stakeholders take part in an action chain which starts with the selection of the best species for a given plot and ends with the harvesting of seeds after having determined the location for such harvest.

The main concern for owners and foresters is to know which species is most suitable for a land plot. Then they need to find out where they can get the seeds for that particular species. Obviously, they can resort to seed nurseries, such as the one in Maceda, participating in the pilot, but they can also make use of public seedbeds. These are known as Admission Units, and are controlled by the administration. They cannot be visualized online, so a public officer must be contacted in order to know which are accessible and where they

are located.

Additionally, seed nurseries must be contacted in order to find out which seeds are available, in what amount and in which quality.

The main problem within the process is the lack of public information available for all stakeholders, as the Admission Units and their availability cannot be visualized and the seed stock and quality available at seed nurseries cannot be consulted.

The stakeholders will know that the Spanish Pilot's intended outcomes have been achieved due to the correct results of the following functionalities:

1. Selection of most suitable species for the owner's or forester's land plot.
2. Selection of public seedbed, also known as Admission Unit, based on the species and its availability.
3. Selection of seed nursery, based on species, qualities, stock and proximity.

Species selection today is mostly based on experience or intuition. The paperwork required for seed harvesting in the designated seedbeds and their selection is analog, i.e. not digital.

The desired outcome is that this process becomes easier and cheaper by means of making all relevant information at each stage of the process more accessible.

There are very clear ways to determine whether the pilot has been successful:

1. Cooperation to implement this process and/or adopting it.
2. If an online service has been provided for in order to make queries on the most suitable species, positive queries on admission units, their availability and data update, and queries on seed nurseries, stocks and their updates.
3. Number of visits to these services.
4. Number of visits to these services versus traditional analog queries. Participants will know if the intended outcomes have been achieved using this ratio: visits to each query services versus analog queries. Such comparison can be easily made using figures of previous years.

5.3.5 Czech Republic Pilot

The Czech pilot main focus is to provide guaranteed open data from the NFI through one access point and compliant with the EU standards. In Czech, high-quality, regular update and statistical sound information on forests and landscape based on LOD has not been addressed yet. Czech NFI portal will address these issues in a generic way. It will be based on NFI data and information, which will be transferred effectively to any potential users. Users will be able to identify, search, filter and combine relevant target parameters, the estimates of which would be delivered for specified domains (geographic areas or attribute-based partitions e. g. land tenure, management system, site index, potential and current soil degradation).

We expect an increase of the number of visitors and also the number of applications using our data, therefore the evaluation would be handled by tools allowing measure the access to the data through the web.

In the pilot we have defined target groups e.g. foresters, forest owners, general public, non-governmental organization, universities and so on, these targets should be evaluated separately, because they will have different scope of the access.

We defined some initial steps with deadlines to reach the final presentation of the NFI results, thus we can evaluate if the steps have been reached or not, after the deadline expires. The schedule was approved by our headquarters, therefore is for as determining. Naturally the SmOD add-ons to the presentation have to go along with the presentation itself, because presented data are very complex and prone to misinterpretation and have to be described well. The steps are as follows:

1. Initial version of the portal with LOD (most probably RDF) output is planning to start.
2. Publish an extra data sources e.g. remote sensing products, extended statistics. Define data ontologies within connections to other data providers.
3. Other estimates and filling of source data, which probably generates more relations and ontologies.
4. Other steps are defined out of the time scope of the project.

If the LOD will be published, It could be complicated to evaluate if the goals have been achieved by using an exact statistic, because LOD connections are not so easily traceable. The Evaluation will be also provided by the UHUL colleagues and its partner organisation as well as other stakeholders immeasurable, verbally and otherwise.

We can say that the evaluation will often established immeasurable, but of course there are any indicators.

- Website will be established.
- Number of visits.
- Official feedback from the bodies outside the UHUL.
- Number of connections or applications demanding our data sources if will be possible to track. Some tool can come up during the project and after it ends, as well.
- Number of queries to the deployed services, because an analytical functional within the portal are planned.

6 Evaluation execution

Evaluation has to be done after the project has acquired a level of maturity that allows users to give concrete useful information that will improve the platform, instead of a fistful of bugs and unimplemented features. Hence, before inviting potential users to any evaluation session, the prototypes have to be prepared for evaluation. It is essential to make sure that all technical requirements have been accomplished.

Right after, the SmartOpenData prototypes functionalities will be tested by the rest of the team partners for the sake of enhancement and completeness of the integrated architecture. Then many actions have to be taken before the evaluation is carried out in depth, such as fixing existing bugs.

While this step might seem quite obvious, in practice delivering a set of intermediate prototypes (within a long-running development process that is distributed over multiple collaborating partners) that are both stable individually and consistent if put together is a time-consuming task. In order to mitigate this risk, a continuous integration server will be used, and external evaluation will be focused on the pilot demonstrators.

Verification phases are part of the internal evaluation to be done during development, according to the quality models and measurements defined previously. The following sections will delve into the particulars of executing the evaluation, by collecting feedback and producing a proper and complete evaluation of the results of the project.

6.1 Preparation of evaluation

After making sure that the prototype is finalized, the evaluation sessions need to be prepared. First of all, it is necessary to set specific goals for this evaluation iteration and select the scenarios under test. Depending on which aspects we want to assess, there are many decisions that have to be taken regarding the upcoming evaluation session.

Among the several tasks, there's the need to adjust the target evaluation group, which will also need to be recruited. As part of that recruitment, invitations have to be sent. Depending on the relationship of the participants with the pilots as defined in SmartOpenData deliverable D6.2, the introductory materials need to be adjusted. For the least acquainted with the project, it will be necessary to provide more information, whilst those who have already been contacted as part of the surveys for requirements gathering might already be informed about it and need less information.

Specific metrics, in accordance to the goals of the evaluation at that moment, will need to be selected. And depending on those metrics, the tasks for the participants have to be designed and prepared, along with any extra material that attendees might need to evaluate the system. If necessary, attendees should be allowed to tinker with the system to provide a complete evaluation. That might lead to a more open evaluation. But regardless of that choice, individual questionnaires have to be created for the participants of the evaluation session.

6.2 Execution of Evaluation

In this phase, the evaluation session takes place. These sessions may occur within a workshop or any other dissemination activities, or be undertaken online. Sometimes it is necessary to have an introductory class, a training session or a video introduction. While introducing SmartOpenData, motivations and goals of the whole project have to be stated. The main elements have to be presented in terms of targets and stakeholders.

Then, the user is ready to execute the tasks that they are given under exhaustive observation. While being observed, participants may be subject of eye tracking analysis or “thinking aloud” method. It all will depend on the final form of the system and the evaluation metrics defined in the previous phases. Observations have to be performed by people that have been specially prepared to it. After dealing with these tasks participants to the evaluation have to fill the questionnaire prepared in the previous step.

Because of the wide distribution of the partners some evaluations will be conducted online, using appropriate tools and support by local project members.

6.2.1 Questionnaires

In this section we define the guidelines to elaborate the questionnaires. Those questionnaires will be developed based on the definition of the metadata and technical requirements which will be defined throughout this project.

A simple binary yes/no scale will be employed for some questions that require a non qualitative answer, otherwise the data collection process will be carried out by means of a *Likert* scale 1-to-6 to avoid “mild evaluation”. See the following table for reference:

6	Strongly Agree
5	Agree
4	Moderately Agree
3	Moderately Disagree
2	Disagree
1	Strongly Disagree

The user is invited to check the box on the right side of the scale with a cross or a tick, as shown in the following example:

6	Strongly Agree	X
5	Agree	
4	Moderately Agree	
3	Moderately Disagree	
2	Disagree	
1	Strongly Disagree	

For analytic purpose the respondent must check a single box for each question, according to the one that reflects mostly his/her opinion.

Customised single choice tables will be also employed for specific questions requiring a more technical approach. If multiple choices are available, it will be indicated for each question.

If necessary, keywords and semantic differential will be used for a highly qualitative evaluation of aspects that deal directly with respondents' perceptions.

Furthermore, at the top of each questionnaire, a brief description of the survey goals and precise instructions about the way the participants can fill them out will be provided. Questions must be formulated with perfect clearness and without biasing the user's answer.

Statistical techniques (descriptive statistics such as means, standard deviation) will be used for the quantitative analysis of the results. The quantitative analysis will determine the mean and the standard deviation for each element. If the results do not provide the information necessary to measure the quality of the results and its usefulness, additional activities will be performed. Due to the project characteristics, the main action will consist of the revision of the questionnaires to incorporate new elements to be able to provide recommendations.

6.2.2 Interviews with users

Interviews may be used in conjunction with questionnaires so that the researchers can have the chance for individual discussion with each of the target groups (or some of them) that would allow them to further elaborate on answers given in questionnaires.

The interviews' format will include semi-structured questions, which is a flexible method allowing the interviewer to follow the main axis of the interview and the interviewee to have the chance to argue about his/her positions. Interviews are more suitably addressed to service developers and content providers but they can be used with end-users as well, to get a better insight of how the public part and demonstrators can be improved.

The interviews can be performed by email, telephone or face to face, always following a short presentation of the project environment. The interview format will allow gathering information about the quality, pan-European aspects, technical aspects, usability aspects, business model, etc. It will also allow us to collect information about complementary needs and suggestions to improve the platform.

6.3 Feedback

In this phase, the information resulting from previous steps will be analysed in order to ascertain the current status of the pilots. If the evaluation was well prepared, this phase should provide refined information that can be directly used to improve the pilots, especially on usability and features.

The data has to be adapted and prepared for analysis. The collected information is then subject to close study. During this process it is important to keep in mind that the aim of this analysis is to identify possible usability weaknesses and any other technical issues that need to be solved. Special attention is paid to the participants' feedback and impressions of the SmartOpenData pilots and tools.

6.4 Use feedback for further development

After all of the analysis is done, a set of conclusions will be provided to the technical work packages in the form of an internal report. This report has to include general goals that have to be achieved facing the next iteration of the evaluation process. Based on these goals, technical requirements of SmartOpenData will be extended and adapted to the new needs. Besides, any problems or bugs that are found will be directly reported to the developers via an issue tracking system.

7 Control groups

Before any evaluation procedure, it is important to define the control groups. The quality of the results depends directly on the right choice of sources.

In SmartOpenData, we will use the cooperation of several individuals, companies and other European projects from the early stages of the project. Thus, all of the participants in our requirements gathering surveys and interviews will be invited to participate in the evaluation sessions as defined in D6.2.

To augment that baseline, researchers from the different workshops, congresses and other events included in the SmartOpenData dissemination plan will be invited to join and give feedback right on the spot.

Moreover, professionals and companies will also be invited to join the evaluation phase, as well as to try the demonstrator. This will provide a better understanding of the needs and opinions of the users outside of academia, which is more often than not quite different and biased towards pragmatism.

Lastly, the general public will also play a role in the evaluation of the system, either by participating in the online questionnaires or by using and rating the system, hence providing useful usage information.

8 Plan of activities

In order to give a better overview of the evaluation process, the activities mentioned in the document have been summarised in the following table. It is important to note that the dates given are suggestions and may be subject to changes throughout the project. Since the below mentioned phases are relative to the specific pilot, the evaluation should be the leader of the various phases should be the leader of the specific pilot.

Phase	Activity	Date
Preparation of evaluation	Define test metrics, prepare materials and test cases for the initial version of the demonstrator	M1-M9
Functional tests and bug fixing	Prepare the initial version of the demonstration pilots	M9-M18
Execution of evaluation	Internal evaluation among partners, prior to release	M15-M19
Execution of evaluation	Perform live demonstration with target users	M18-M24
Execution of evaluation	Online polls to measure the satisfaction with the initial version of the demonstrator	M18-M24
Analysis of evaluation results	Gather information from the evaluation	M19-M24
Use feedback	Introduce changes and improvements in the demonstrator to create the final version	M23-M24
Functional tests and bug fixing	Prepare the final version of the demonstrator	M22-M24
Preparation of evaluation	Update the test cases for the final demonstrator	M24
Execution of evaluation	Evaluation with test cases for "Open Dataset Providers" and "Service Developers"	M24
Execution of evaluation	Evaluation with test cases for "End users"	M24
Analysis of evaluation results	Gather information for the final evaluation and assessment	M24

9 Conclusions

This deliverable defines a methodology that will drive the evaluation of the SmartOpenData project.

Feedback collection has been described and defined, and we have described how our test subjects will be chosen, and how to proceed with the evaluation. This includes a detailed description of the goals and tentative needs/goals of each user type, as well as means to evaluate their satisfaction and gather their opinions and contributions.

10 References

- [And99] Anderson P. V. Technical Communication: A Reader-Centred Approach. Harcourt Brace College Publishers, 1999.
- [El-Emam02] Khaled El-Emam, "Object-oriented metrics: A review of theory and practice," in Advances in software engineering. New York, NY, USA: Springer-Verlag New York, Inc, 2002, pp. 23-50.
- [ISO/IEC-25000] ISO/IEC 25000, Software Engineering – Software product quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE, 2004.
- [ISO/IEC-25010] ISO/IEC 25010, Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and software quality models, 2011.
- [ISO/IEC-25022] ISO/IEC 25022, Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – Measurement of Quality in Use, 2012.
- [ISO/IEC-25023] ISO/IEC 25023, Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – Measurement of System and software product quality, 2011.
- [ISO/IEC-25030] ISO/IEC 25030, Software Engineering – Software Product Requirements and Evaluation (SQuaRE) – System and software quality models, 2011.
- [ISO/IEC-25040] ISO/IEC 25040, Software Engineering – Software product quality requirements and Evaluation (SQuaRE) - Quality Requirements, 2007.