

Evaluating a collaborative IT based research and development project

Zaheer Khan¹, David Ludlow¹ and Santiago Caceres²

¹Faculty of Environment and Technology,

University of the West of England,

Coldharbour Lane, BS16 1QY, Bristol, UK

Email: {Zaheer2.Khan, David.Ludlow}@uwe.ac.uk

Tel: +44 117 328 2636

²New Technologies,

ETRA I+D

Tres Forques 147 - 46014 Valencia - SPAIN

Email: scaceres.etra-id@grupoetra.com

Tel: +34 96 313 40 82

Corresponding Author: Zaheer Khan – Zaheer2.Khan@uwe.ac.uk

Evaluating a collaborative IT based research and development project

Abstract

In common with all projects, evaluating an Information Technology (IT) based research and development project is necessary in order to discover whether or not the outcomes of the project are successful. However, evaluating large-scale collaborative projects is especially difficult as: i) stakeholders from different countries are involved who, almost inevitably, have diverse technological and/or application domain backgrounds and objectives; ii) multiple and sometimes conflicting application specific and user-defined requirements exist; and iii) multiple and often conflicting technological research and development objectives are apparent. In this paper, we share our experiences based on the large-scale integrated research project - *The HUMBOLDT project* – with project duration of 54 months, involving contributions from 27 partner organisations, plus 4 sub-contractors from 14 different European countries. In the HUMBOLDT project, a specific evaluation methodology was defined and utilised for the user evaluation of the project outcomes. The user evaluation performed on the HUMBOLDT Framework and its associated nine application scenarios from various application domains, resulted in not only an evaluation of the integrated project, but also revealed the benefits and disadvantages of the evaluation methodology. This paper presents the evaluation methodology, discusses in detail the process of applying it to the HUMBOLDT project and provides an in-depth analysis of the results, which can be usefully applied to other collaborative research projects in a variety of domains.

Highlights:

- Planning evaluation of collaborative and integrated research and development projects
- Generic evaluation methodology applicable to different components of the project
- Participatory evaluation performed in the HUMBOLDT project
- Evaluation results, lessons learned and recommendations

Key words: *Integrated Project Research and Development, Collaborative Research Project Evaluation, Participatory Evaluation Methodology, Questionnaire and User-based Evaluation, User and Technical Evaluation.*

Evaluating a collaborative IT based research and development project

1. Introduction

i) Problem Context

One critical aspect of an Information Technology (IT) based research and development project is the evaluation of its outcomes against its stated objectives and requirements by using a specific evaluation methodology. Undertaking the evaluation of an integrated system is not a simple task for a number of reasons, for example, on time delivery of the different components of the system when: i) interconnections exist between the various components; ii) stakeholders are involved from different countries that have diverse backgrounds in respect of technology or application domain; iii) multiple and sometimes conflicting user-defined and applications specific requirements must be accommodated; iv) multiple and often conflicting technological research and development objectives must be addressed; and v) and the evaluation is highly dependent on the availability of resources. All these considerations demand a coherent evaluation methodology, planning to adapt to the varied requirements of collaborative integrated research projects, and use of the available resources in an efficient manner.

In the literature a large number of research and system evaluation approaches, methods and techniques are described (Dobrica L and Niemela E 2002; Babar M A et al 2004; Sommerville I 2006; Juristo N 1997; Juristo N and Morant J 1998). However, assessment of the outcomes of large-scale integrated and collaborative projects using ad-hoc evaluation approaches has its shortcomings. It may not fully reveal, for example, the limitations and disadvantages of the system. Furthermore any single evaluation methodology or technique may not offer sufficient capacity to fully assess the various aspects of the system that need to be evaluated. As a consequence, it is critical to define the most appropriate evaluation methodologies and techniques for such systems. This paper presents the Criteria Indicators Metrics (CIM) evaluation methodology used to evaluate the EU FP6 HUMBOLDT project (HUMBOLDT Project 2006-2011; Khan Z et al 2012), a large-scale integrated and collaborative research project, with 27 partner organisations plus four sub-contractors from public and private sector representing 14 member states of Europe. The main objective of this paper is to present in-depth analysis of the process of applying CIM to the HUMBOLDT project, and reveal its strengths and weaknesses. Furthermore, the paper assesses the extent to which CIM methodology has successfully evaluated the project outcomes against the project's stated and implied objectives. Specification of the overall research outcomes of the project in terms of software tools, components and services are not within the scope of this paper, and are presented elsewhere (Villa, P et al 2012; Khan Z et al 2012; Fichtinger, A et al 2011).

ii) Background and Introduction to the HUMBOLDT Project

In Europe, the INSPIRE Directive mandates all member states to ensure compliance between their geo-data and INSPIRE specifications (INSPIRE 2007), so that cross-border unified information can be produced to support better decision-making and policy development. INSPIRE supports a bottom-up perspective for information gathering and monitoring that is complementary to the GMES initiative (Global Monitoring for Environment and Security) (GMES 2012) that follows a top-down approach for environmental monitoring, for specific application domains such as land, marine, risk management and security.

In relation to the above, the HUMBOLDT project aims to address geospatial data harmonisation by developing state-of-the-art tools, components and services compliant to the INSPIRE and OGC (OGC 2012) standards. HUMBOLDT Framework services and tools (HUMBOLDT Framework Tools 2010) include HUMBOLDT GeoModel Editor, HUMBOLDT Alignment Editor (HALE), Workflow Design and Construction Service, Conceptual Schema Translation Service, Edge Matching Service etc. The process for the design and development of these HUMBOLDT tools and services includes: i) performance of state of the art review of the technology, and ii) acquisition of stakeholder requirements in respect of various environmental application fields related to GMES applications, known in the HUMBOLDT project as scenarios. Considered only from the perspective of technology, a state-of-the-art literature review could have derived the major requirements for the development of HUMBOLDT Framework. However, such a system would not meet the requirements for full implementation of some solutions, for example those concerning spatial data harmonisation issues faced by various environmental application domain communities. As a consequence, a user perspective was seen as essential to derive the spatial data harmonisation requirements in relation to the various environmental application domains. As a result in addition to the accommodation of user defined requirements from the selected uses cases (e.g. 9 HUMBOLDT scenarios) for various GMES related environmental application domains, it was also possible for HUMBOLDT project partners to derive an evaluation framework for the HUMBOLDT framework and its application to scenarios.

The principal objective of the HUMBOLDT scenarios is deployment of the framework components in the real world. Fundamentally, the HUMBOLDT scenarios provide a community driven research environment that offers feedback on the functionality, usability, benefits and relevance of the HUMBOLDT tools, which thereby contribute towards the full sustainable development of the HUMBOLDT Framework. The HUMBOLDT scenarios include: Environmental Risk management, Forest management, Protected Areas management, Trans-boundary Catchment, Ocean, Air quality (also referred as Galileo), Urban planning and Urban Atlas, and Border Security (HUMBOLDT Scenarios 2010). These various scenarios deploy real-life considerations to varied European urban and regional contexts, in particular addressing cross-border and cross-thematic data harmonisation challenges that limit the planning and decision making capabilities of policy makers. Essentially, a scenario is

a “requirement generator” and therefore central to the user-driven development process. At the same time the scenario is a demonstration activity that articulates the results from the framework and software development process via the applications, which can thereby be used and evaluated by domain expert users and intermediate level users. As a result the HUMBOLDT framework and scenarios can be assessed and proven by demonstrating that they facilitate application developers in the creation of useful software that permit the integrated usage of heterogeneous services and data sources.

The HUMBOLDT evaluation addresses three elements: i) automated technical validation (HUMBOLDT Technical Evaluation 2010) using Continuous Integration Server - HUDSON server (HUDSON 2010), and Code Analysis Server – SONAR (Sonar 2010); ii) cost-benefit analysis (Ostreika A, 2010); and iii) user or participatory evaluation. The central interest of this paper is on the participatory or user evaluation element of the overall evaluation framework identified above. This user evaluation was performed on both the HUMBOLDT Framework components, as well as the scenarios. However, the scenario evaluation was complicated e.g. due to licensing issues concerning access to application specific data and proprietary software used by specific scenarios. As a consequence, this evaluation required rigorous collaborative planning, and a structured evaluation methodology to assess the project outputs. This is elaborated in the following sections.

iii) Related work

The process of defining the evaluation methodology for the HUMBOLDT project builds upon an existing foundation of evaluation methods. The methodology has similarities to GQM (Basili R V et al 1994), and the concerned-based approach (Kotonya G and Sommerville I 2002) utilised by (Khan Z et al 2011). In addition, other evaluation methodologies and experience was collected and analysed, especially that applied in Information Technology (IT), Geographical Information Systems (GIS) and Spatial Data Infrastructure (SDI) related research projects. Other considerations included the evaluation concepts developed by the EU funded projects CASCADOSS (CASCADOSS Consortium (ed.) 2007), RISE (RISE (ed.) 2007), and WebPark (WebPark consortium (ed.) 2002), relevant international standards (ISO/IEC 9126-1 2001) and additional literature (e.g RESPECT Consortium 1998; Tiits 2003).

In addition to the above, there are a number of research and system evaluation frameworks, methodologies and techniques in the literature (Dobrica L and Niemela E 2002; Babar M A et al 2004; Sommerville I 2006; Juristo N 1997; Juristo N and Morant J 1998) which can be applied to the assessment of specific areas of IT orientated collaborative development projects such as HUMBOLDT. For example, Falessi D et al (2009) emphasized the view that each software engineering area (i.e. system architecture, reverse engineering, etc.) has specific assessment issues, suggesting that assessment solutions specific for each research area should be developed. Overall we support the views of Falessi, D. et al, but also argue that this approach may not deliver a holistic evaluation necessary for integrated development projects, particularly as such

projects frequently involve multiple and cross-disciplinary research areas. All of this indicates the need for an evaluation methodology that unifies the assessment aspects and objectives of different research areas within integrated projects. In addition, IT project evaluation often does not fully accommodate the evaluation objectives of different stakeholders. In this respect, Bryson J, et. al. (2011) emphasized the need to identify the potential evaluation stakeholder's interests, needs, priorities, concerns, etc in planning, designing and applying evaluation based on different techniques. These techniques facilitate step-by-step participatory evaluation for collaborative decision-making and impact assessment and ensure that stakeholders' perspectives are fully considered in the evaluation of IT projects.

Some researchers (Gustedt J et al 2009) argue the importance of assessing a system through experiments in computer science and especially in distributed computing. In this context, they compare and contrast various experimental validation methodologies such as real-scale experiments, emulation, simulation and benchmarking. These methodologies can be useful to answer specific evaluation criteria but require additional efforts in planning, structuring and performing a comprehensive evaluation for answering different research questions and validating the research objectives.

Juristo N et al (Juristo N 1997; Juristo N and Morant J 1998) have defined an evaluation framework that can be used for the evaluation of both knowledge-based and conventional software domains. This framework determines 'what to evaluate, who is to evaluate it, how it is to be evaluated and the implications of the outputs of the evaluation'. One of the most positive aspects of Juristo et al's evaluation framework is that it can be tailored to enable the evaluation of a particular system (i.e. conventional software systems, etc.) using specific criteria (e.g. correctness, consistency, completeness, etc.) supported by different evaluation techniques (e.g. walkthroughs, inspections, dynamic testing, etc). However, it is not clear how to operationalise this approach for an integrated project developed in a collaborative environment.

Other researchers (Lindvall M et al 2007) consider that in most sciences, an established mechanism for performing experimentation exists, with a well-adapted dedicated set of tools (e.g. experimental validation methodologies in the distributed computing domain as reviewed by Gustedt J et al 2009). However, software engineering has no such established protocol for experimental validation. Lindvall M et al 2007 used a Unified Model of Dependability (UMD) method (Basili R V et al 2004), which aims to capture stakeholder's concerns i.e. dependability requirements and scope, external events and measures, and reactions to mitigate faults. This approach is useful for technical evaluation, for example experimental validation of system architectural violations, and model checking for design verification. However, it does not address user evaluation necessary to assess the extent to which the outcomes and objectives of the development project have been attained.

Among others, the Goal Question Metric (GQM) paradigm (Basili R V et al 1994) provides a general approach to

specify a measurement system focusing on goals, viewpoints and specific questions for the quantitative measurement of a software product. Similarly, the GQM/MEtric DEfinition Approach (GQM/MEDEA) (Briand L et al 2002) is driven by the GQM based explicit measurement goals to be addressed, and the set of explicit empirical hypotheses that need to be validated. Furthermore, explicit attributes of interest are identified and theoretically valid measures are defined for these goals, which are subsequently used to validate the experimental hypotheses. This approach requires careful development of measurement goals, viewpoints and empirical hypotheses but it is not clear whether or not qualitative assessment can be performed using this approach.

While these other evaluation processes have proven useful in other contexts, it was found necessary in the HUMBOLDT project to incorporate user requirements, and adapt the framework based on specific stakeholder use cases. These approaches (Juristo N 1997; Juristo N and Morant J 1998; Basili R V et al 1994; Briand L et al 2002) appear very close to the CIM evaluation methodology presented in this paper, excepting that the CIM approach criteria and questions are driven by user requirements, which are investigated by specific stakeholders in order to address relevant questions using both qualitative and quantitative metrics. The prime characteristics of the CIM approach are its generic application in various software engineering areas of collaborative projects, and its characteristic ability to adapt to different evaluation methods/techniques, including questionnaire, code inspection, unit testing, model checking, walkthrough, white and black box testing (Sommerville I 2006).

iv) Paper structure

The structure of the paper is as follows: in section 2, the user evaluation methodology is briefly elaborated, followed by a description of user categories and evaluation planning in section 3. In section 4, an example of the application of the CIM based evaluation of a selected HUMBOLDT Framework tool is presented. Subsequently, in section 5 selected examples of scenario evaluation using the CIM approach are detailed. A critical analysis of the evaluation methodology and evaluation outcomes of the HUMBOLDT project are presented in section 6, followed by discussion in section 7. Finally, section 8 provides the conclusion.

2. Participatory Evaluation Methodology

As the development of the HUMBOLDT Framework and Scenarios is both user-driven (based on actual user requirements) and technology-driven (based on a state of the art analysis), both these different perspectives are addressed in the evaluation of the HUMBOLDT outcomes. As a consequence the evaluation results are based on two perspectives: a

technology-oriented “perspective A: software functionality and quality” and a user and context-oriented “perspective B: relevance and benefits”.

Perspective A addresses software functionality and quality including usefulness and usability. This perspective aims to answer the central question: “Does the software developed meet the requirements defined in the specification?”. Perspective A is based on the main evaluation criteria as defined in the ISO 9126 quality model including functionality, maintainability, portability, reliability, usability, efficiency etc.

Perspective B addresses the HUMBOLDT results from the perspective of potential users and the HUMBOLDT context. This perspective is especially important for the evaluation of the HUMBOLDT scenarios, which are developed according to user-defined criteria. This perspective should answer the central question: “To what extent have the stated and implied objectives of the project been achieved?” Relevance and benefits are the main evaluation criteria under Perspective B.

In general, the evaluation of both the perspectives A and B targets the HUMBOLDT Framework components, and scenario applications, as well as other relevant resources (including training material, test cases, component or scenario software documentation, etc) which can facilitate evaluation, as depicted in Figure 1. In order to develop understanding of the HUMBOLDT project research and development activities, Figure 1 also depicts the project lifecycle by indicating the sequence of development steps and evaluation activities performed. At first, scenarios were identified from the GMES domain which provided a user perspective facilitated by the state of the art technological perspective, as a driving force for the research and development of the HUMBOLDT Framework components, tools and services. Subsequently the HUMBOLDT Framework was applied to the scenarios and detailed documentation and training material were prepared. These activities can be considered as evaluation enablers, which facilitated evaluating of the extent to which HUMBOLDT Framework tools, components and services could meet the stated and implied technology and user objectives.

Evaluation and Impact Assessment

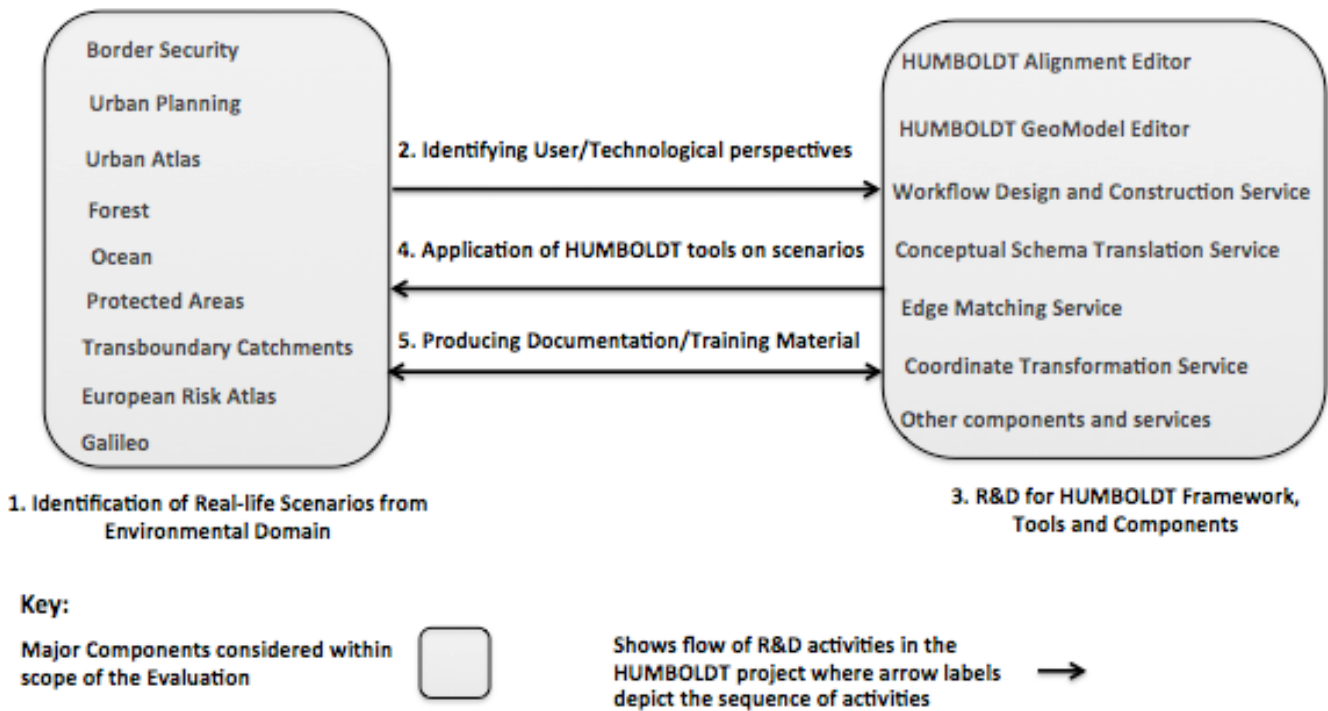


Figure 1: Scope of Evaluation in the HUMBOLDT Project

In order to perform the above evaluation in the HUMBOLDT project, the CIM methodology has been applied (Khan Z, 2012). This methodology primarily defines the means of securing evaluation results. Figure 2 depicts the overall conceptual schematic of the CIM methodology. This approach defines a set of criteria based on stakeholders' requirements (e.g. user functional and non-functional requirements), which need to be considered, for example considerations related to usability, functionality, performance, relevance etc. Each criterion may have additional associated sub-criteria to address other aspects. Each sub-criterion is made operational by defining one or more indicators which consider the evaluation criteria in question. To understand more fully the context for the evaluation for any particular aspect, each indicator is represented by one or more questions. Furthermore for each indicator (and associated questions) quantitative and qualitative metrics are defined to judge whether or not the result is regarded as satisfactory, good or bad.

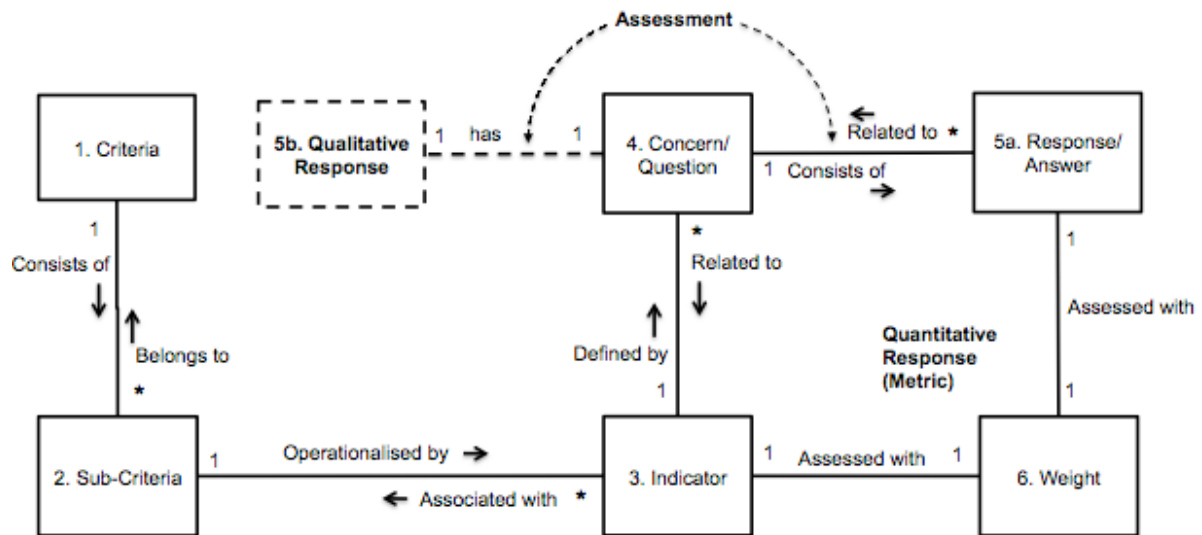


Figure 2: Evaluation Methodology: Criteria-Indicator-Metrics

In Figure 2, each block represents a major component of the CIM methodology. In order to show the start point for the evaluation design, each block is labelled with a specific component name and numbered to depict the sequence of activities in the evaluation processes. For example, the first step in the process is to identify a set of criteria followed by derivation of related sub-criteria and so on. These components are associated with each other using labelled lines which indicate the relationship between them. These associations are bidirectional, shown by labelled arrows that enable the linking of individual components in both backwards and forwards manner, and so help in relating evaluation responses to individual criterion. Furthermore, each association has specific multiplicities which indicate possible number of instances that an individual component is associated with a specific number of instances of another component. These multiplicities are indicated by labelling the lines with ‘1’ for ‘one’ and ‘*’ for ‘many’. For example, one instance of a specific ‘criterion’ can consist of many instances of ‘sub-criterion’. This example can also be interpreted in a backwards direction, for example, many instances of specific ‘sub-criterion’ can belong to one instance of a certain ‘criterion’. Likewise, one instance of a ‘sub-criterion’ can be operationalised by many instances of ‘indicators’, or many instances of specific ‘indicators’ can be associated with one specific instance of a ‘sub-criterion’ and so on. The dotted block in Figure 2 indicates that responses to specific question may also be of a qualitative nature, so enabling the CIM methodology to perform quantitative as well as qualitative assessments.

In Figure 3 shows an example where ‘usability’ is one of the main elements, as derived from ISO 9126 (ISO/IEC 9126-1 2001), and selected as a main criterion with two further sub-criteria defined as: ‘learnability’ and ‘understandability’. As an example, each sub-criterion identifies indicators, also considered as concerns to be evaluated. In order to define metrics for indicators, specific questions are defined which address the aspects to be evaluated for each indicator. These questions can have specific response options each defined with both qualitative values and quantitative (and measurable) weights.

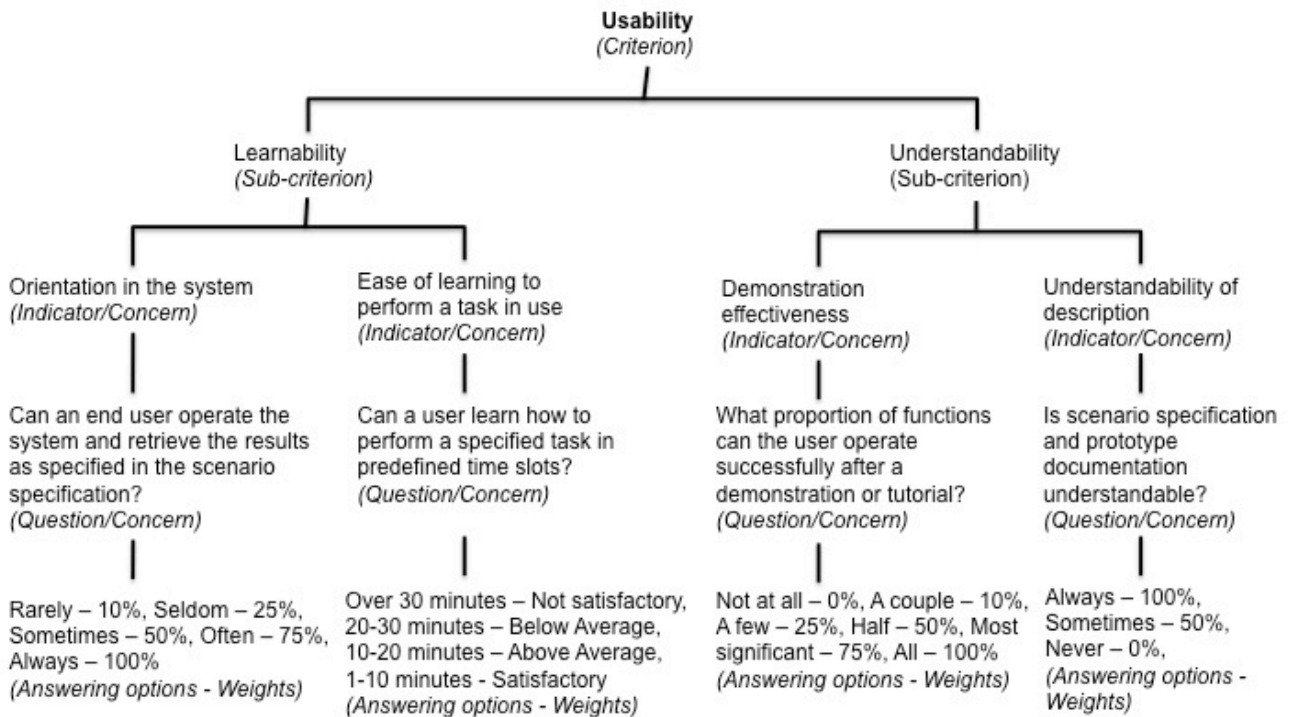


Figure 3: CIM methodology: An Example of Usability Criterion

Overall, the evaluation methodology identified above adopts a black-box testing approach, and requires stakeholder participation in defining criteria, indicators and weights for the overall impact assessment. In evaluating the HUMBOLDT project the prime evaluation considerations are based on the following criteria for both A and B evaluation perspectives:

Relevance – Relevance of the HUMBOLDT scenarios for the attainment of the HUMBOLDT project goals, for the HUMBOLDT users, and for depicting the goals and potentials of INSPIRE and GMES;

Benefits – Potential benefits of the HUMBOLDT scenarios as compared with the pre-HUMBOLDT situation (if applicable);

The software quality perspective was also considered based on the following criteria for both the HUMBOLDT Framework and scenarios:

Usability - Capability of the software to be understood, learned, used and appreciated by the user, when used under specified conditions;

Functionality - Capability of the software to provide functions which meet stated and implied needs, or requirements when the software is used under specified conditions.

Data gathered during the application of the evaluation methodology from different stakeholders was analysed by taking into consideration the different weights allocated to the different indicators as discussed in the following sections 3.2.1 and 3.2.2.

3. HUMBOLDT User Categories and Evaluation Planning

Bryson J, et. al (2011) stressed the need to thoroughly identify different categories of evaluation stakeholders to be used for evaluation planning, design, implementation and decision making process. This section describes aspects of the user evaluation in relation to both the HUMBOLDT user categories, as well as evaluation planning and preparation for the Framework tools and scenario applications.

3.1 HUMBOLDT User Categories

In order to secure proper engagement with different stakeholders, HUMBOLDT project stakeholders were categorized as: i) Developers (Dev): developing the HUMBOLDT Framework component (primary developers) and scenario applications (secondary developers); ii) Data Custodians (DC): in institutions providing data which has been adapted to given standards due to legal or market requirements (harmonised); iii) Data Integrators (DI): individuals or institutions using heterogeneous geodata for the requirements of their work; iv) End-user of Geo-Data (EGD): who are directly working with geodata (e.g. already in harmonised form); and v) End-user of Spatial Information (ESI): who do not use geodata directly, but are more interested in the information arising from it. These user categories permitted the specification of more targeted questions for relevant user groups. In addition, the HUMBOLDT user community (HUMBOLDT Project Community 2006-2011) has been using HUMBOLDT tools throughout the project, and so contributing to the identification of bugs (i.e. software errors) and additional features required in respect of the tools and services. However, this user category is more related to technical validation and is not addressed in this paper.

3.2 Evaluation Planning and Preparation

3.2.1 Common Approach for the HUMBOLDT Framework and Scenarios

The conceptual underpinnings for defining criteria, indicators and metrics were based on standards including ISO 9126, as well as an analysis of stakeholder requirements. For example, the project stakeholders were involved in the process of defining indicators and questions directly associated with the HUMBOLDT Framework and scenarios requirements, and determining the weights of the indicators based on relevance and importance to scenarios. As a result, a detailed set of questionnaires were derived from the requirement specifications for each criterion, and included in the template depicted in Figure 4, below.

Criterion ID	Main Criterion	Sub-Criterion	Users					Indicator ID	Indicator	Description/Questionnaire	Answering Options	Importance High (H), Medium (M), Low (L), Not Relevant (NR)
			Dev	DC	DI	EGD	EGI					
A3	Functionality	Interoperability	Yes	NO	Yes	NO	NO	A3-16	Interoperability of components	Can scenario prototype interact with components of the HUMBOLDT framework?	GeoModel Editor: <i>(yes/no/partially)</i> , HALE: <i>(yes/no/partially)</i> , WDCS: <i>(yes/no/partially)</i> , CST: <i>(yes/no/partially)</i> Edge Matching: <i>(yes/no/partially)</i> , Information Grounding: <i>(yes/no/partially)</i> , Coordinate Transformation: <i>(yes/no/partially)</i> , Context service: <i>(yes/no/partially)</i> Language Transformation: <i>(yes/no/partially)</i> , Mediator service: <i>(yes/no/partially)</i> ,	8H
A5	Usability	Understandability	Yes	Yes	Yes	Yes	Yes	A5-13	Completeness of description	Does scenario prototype documentation provide detailed, clear and understandable functional description of the scenario application?	<i>always,</i> <i>sometimes,</i> <i>never</i>	7H, 1M
B1	Relevance	Relevance for illustrating goals and potentials of INSPIRE	No	Yes	Yes	NO	NO	B1-11	Conformance of services	Is data set used in the scenario prototype accessible through INSPIRE compliant view services (WMS) and/or download services (WFS)?	<i>yes,</i> <i>no,</i> <i>sometimes,</i> <i>not relevant</i>	4H, 1M, 2L, 1NR
B2	Benefits	Qualitative benefits compared to pre-HUMBOLDT state	No	Yes	Yes	Yes	Yes	B2-14	Increased usability	Is scenario prototype easy to use?	<i>yes,</i> <i>no,</i> <i>sometimes,</i> <i>never,</i> <i>not relevant</i>	6H, 1M, 1NR

Figure 4: Extract from the User Evaluation Questionnaire using CIM methodology

Figure 4 shows how evaluation data is prepared in compliance with the CIM methodology. For example, ‘Criterion ID’ uniquely identifies each criterion in ‘Main Criterion’ column, which has further sub-criterion in the ‘Sub-Criterion’ column and so on. The ‘Users’ column is included in Figure 4 to indicate who should respond to specific assessments. An allocation of weights to each indicator and its associated metrics is facilitated by the ‘Importance’ column, which indicates how important it is to make an assessment based on a specific indicator and its respective question for a particular HUMBOLDT scenario. For example, ‘8H’ means that ‘Interoperability of components’ indicator and its respective question has ‘High’ importance for 8 HUMBOLDT scenarios. Similarly, ‘7H, 1M’ means that ‘Completeness of description’, and its respective question, is ‘Highly’ relevant to 7 HUMBOLDT scenarios, whereas it has ‘Medium’ importance for one scenario. Entries for the ‘Importance’ column are collected from individual scenario stakeholders as discussed in detail in section 3.2.2 and depicted in Figure 5.

Weight allocations to metrics are kept between ‘0’ to ‘1’, i.e. ‘1’ represents complete satisfaction, that the feature is completely implemented and works perfectly, and ‘0’ indicates that it is not well addressed, and that the user is not satisfied with the results.

3.2.2 Evaluation Planning and Preparation for HUMBOLDT Scenarios

The following steps were undertaken for the evaluation of the HUMBOLDT scenarios. T:

Step 1: Preparation of criteria, sub-criteria, indicators, metrics and questionnaires according to the template depicted in Figure 4. It was not anticipated that responses to all these questions from evaluators could be secured due to: i) the large number of questions (over 65), and ii) no account was taken of the resources required (e.g. access to data) in answering the functionality and usability related questions. Therefore, the following step was undertaken:

Step 2: Revision of criteria, sub-criteria, indicators, metrics and questionnaires. This includes the identification of conflicts, resources to be used for evaluation, relevance or importance of specific criterion, sub-criteria, indicators, metrics and questionnaires by all scenarios as depicted in Figure 5. The main objective of this step was to identify and select critical questions for individual scenarios which are directly related to the functional and non-functional requirements of a scenario. In this regard, the template as shown in Figure 5 aimed to obtain inputs from the primary stakeholders of the scenario in respect of the following considerations:

- i) deriving more suitable and clear questions – enhancing clarity and understanding;
- ii) identifying the appropriateness of the questions and alternative answers – facilitating the elimination of unnecessary questions;
- iii) identifying the relevance and importance of a question for specific scenarios rated ‘High’, ‘Medium’ or ‘Low’ – helping in assigning weights to specific indicators and questions.
- iv) identifying more specific user types (stakeholders) that should answer a specific question – aiming to reduce the number of questions for specific user categories.
- v) identifying potential resources (documents, training material, web links, external software and utilities etc) – facilitating evaluators in answering the questions.
- vi) identifying issues or problems which evaluators may face in evaluating scenario prototypes such as access to a system demonstrator and relevant data, the redeployment of services, access to documentation, and the availability of specific services during certain periods, etc.
- vii) identifying test cases and potential resources required to conduct evaluations for some of the questions, e.g. training material, deployment of system, etc.

Indicator IDs	Description of Questions and Answering Choices (Any suggested changes?)	Importance (Relevance and significance of questions to evaluate scenario i.e. High, Medium, Low)	User categories (Any user category to be excluded and/or included?)	Resources (Indicate resources to perform a test, documents, web-based interface of the scenario, other web links, services, data set, software, training material, etc)	Issues (Do you think that there can be issues in answering a question? E.g. restricted access to data, licensed software, redeployment of services/software, etc.)	Test Case(s) (Any suggested test case(s) with identified resources such as sample data set, web links, software, etc.)	Others (Other considerations)

Figure 5: Extract from the Table used for Collaborative Improvement of the Scenario Evaluation Questionnaire

Feedback was obtained from the stakeholders of each HUMBOLDT scenario for the above exercise, concerning resources available (e.g. demonstrator documentation, software and services, etc), the potential issues arising (e.g. lack of data, access to proprietary software, etc.) and the importance (i.e. high, medium, low, not relevant). This feedback was obtained for all the questions identified for the individual scenarios and inserted in the table (Figure 5). This exercise permitted the refinement and selection of significant questions for different evaluation exercises e.g. questions with high importance for most of the scenarios. This also helped in allocating weights to questions.

Step 3: Preparation of test cases related to questionnaires and dissemination to targeted communities for assessment;

In addition to HUMBOLDT framework components, tools and services evaluation, the following scenario evaluation exercises were planned and undertaken:

- i) The first scenario evaluation exercise selected a set of questions with high importance, mainly to assess the relevance and benefits of the application of the HUMBOLDT Framework to the scenarios.
- ii) The second scenario evaluation exercise (for selected usability, benefits, relevance and functional evaluation questions) was undertaken based on the scenario demonstrations.
- iii) The third scenario evaluation exercise was based on an improved selected set of questions which were formulated based on the outcomes of first two exercises, and targeting specific categories of stakeholder (Dev, DI, DC, EGD, EGI).

Step 4: Analysis of evaluation outcomes, suggesting improvements and deriving conclusions and future directions.

4. CIM-based Evaluation of HUMBOLDT Framework

For the HUMBOLDT Framework user evaluation, the main tool used to undertake the evaluation was the test-case based questionnaire. In this respect, test scenarios were prepared with sample data; user instructions/manual were developed to inform participants of the process of undertaking the testing; testers and stakeholders participating in the evaluation were identified; the timeframe to undertake the evaluation was determined and tele/video-conferencing or face-to-face meetings were organised to perform the evaluation. As a result, sufficient information was collected to plan and manage the evaluation activities of the HUMBOLDT Framework by different stakeholders.

According to the above protocol, all evaluators followed the process outlined below to undertake the evaluation of the HUMBOLDT Framework components:

- i) Download necessary HUMBOLDT software and its documentation, schemas, data, etc;
- ii) Install the tools/services and set up infrastructure;
- iii) Read the questionnaire(s);
- iv) Complete the exercise (test cases) for each one of the tools;
- v) Answer the questions either online or in a document; and
- vi) Confirm to the evaluation manager that the evaluation is complete. The possibility of delivering the above process was high as a result of the ICT nature of the HUMBOLDT framework that could be accessed, installed and evaluated and results submitted remotely via Internet.

The following section, presents the experience of the evaluation of the HUMBOLDT Alignment Editor (HALE), as an example of the HUMBOLDT Framework evaluation. A total of thirteen independent evaluators formally evaluated and assessed HALE, following the methodology described in the previous sections. In addition and informally, a large number of evaluation responses concerning the HUMBOLDT tools e.g. HALE, with feedback including bug (error) reports, and requests for new features etc, were received through the user community website (HUMBOLDT Project Community 2006-2011).

Following the methodology (CIM approach), evaluation test-cases were prepared and associated questionnaires were evaluated from '0' to '1' metric, i.e. '1' represents complete satisfaction, that the feature is completely implemented and works perfectly, and '0' indicates that it is not well addressed and that the user is not satisfied with the results.

For each question the ‘final mark’ was extracted – meaning the average value of results provided by the evaluators on which basis the main conclusions were formulated. Table 1 below provides a few examples for selected questions:

Table 1: Selected Examples from the Evaluation of the HUMBOLDT Alignment Editor

Questions	Final Mark (Metric Value - Average Result)	Analysis
Is HALE GUI translated into your language?	0	Currently the only language used for HALE is English. Users pointed out that they did not find any easy way to translate HALE into their different languages. However, most of them do not see this issue as a high priority.
How easy/difficult is it to install and run HALE?	0.97	Most of the users find the tool easy to install and run. The only concerns identified were: i) the necessity of java programming language run-time environment in order to execute the tool; ii) some problems with the configuration settings for corporate networks; and iii) some problems were identified when the installation of software is performed on a MAC OS X based platform
How understandable is HALE user documentation?	0.65	Mix set of responses were received between average to easy. Evaluators found the quality of the documentation useful but identified some issues for possible improvement, for example: <ul style="list-style-type: none"> - Adding more images in the text, rather than just providing the name or description of specific command buttons - Only pdf format is available, users would prefer a more hypertext (e.g. wiki) format - Some more examples and screenshots would be appreciated - Normal layout (single column) is preferable. Currently this is divided in two columns
How understandable is the user interface in terms of functionality offered by HALE?	0.79	75% of users find the user interface of HUMBOLDT easy to understand, the only concerns included: <ul style="list-style-type: none"> - Some bugs in Ubuntu 9.10 OS that make it impossible to use some command buttons - The Mapping view could be enhanced. When using the default perspective, sometimes there is not enough space on the screen to see the full title of the mapping - Mappings are defined one by one, and it lacks a synthetic view of the source and target relations
Does the application allow you to load source/target schemas?	0.92	Most of users did not find any problem to load the schemas, however this result has to be considered with care, as some of the testers used testing schemas provided by HUMBOLDT in order to carry out the exercise. However those who used their own source and target schemas did not find much difficulty in loading their schemas.
Have you been able to create/add code lists and map reclassifications?	0.82	The results here are quite satisfactory; however some good ideas for improvement were expressed by users e.g.: <ul style="list-style-type: none"> - Having an overview of the values available on source schema, even if they are not really considered as codeLists. - Being able to see the whole reclassification in one window for enhanced usability - Adding the possibility to use operators in the source data attribute (e.g. for value “high” set “temperature>30”, etc)
Have you been able to map attribute merging?	0.71	In most of the cases the mapping has been successful, the only cases where the tool failed were when multiplying source attributes (e.g. integer) by a fixed value (e.g. 100) while merging attributes.
Have you been able to map conversion from ordinates to points? (Meaning a function that creates a geometry point from two numeric values)	0.75	75% of users fully mapped their conversions, but 25% failed to perform mapping from ordinates to points.
Does the mapping from your source and your target schema require additional transformation functions which were not available in HALE?	0.56	Some of the testers found that some transformation functions were not implemented, for example, Calculation of length, Intersections, Constant value, etc. Evaluators recommended that for some transformations, it would be necessary to execute more than one simple transformation. This could only be expressed as a chain of transformation and not by a one step transformation.

The following Figure 6 indicates that the overall final mark for most of the questions is above 0.5, which suggests the successful implementation of the prime required functionalities in HALE. A major limitation was identified as the lack of availability of HALE in different European languages. This proposed as a new requirement to incorporate in the next version of HALE.

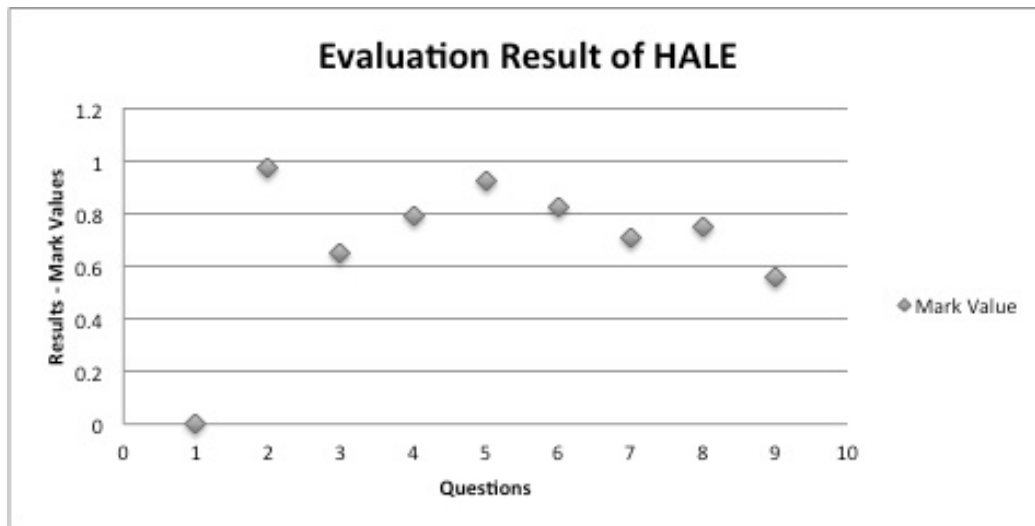


Figure 6: Evaluation Results of HALE

5. CIM-based Evaluation of selected HUMBOLDT Scenarios

Scenario evaluation is performed to assess the extent to which the HUMBOLDT Framework tools and services successfully provide solutions for cross-border spatial data harmonisation in selected scenario applications. A total of 25 evaluators participated in three rounds of HUMBOLDT Scenario evaluation. The following presents the results from the scenario evaluation exercises. Only a selected set of questions are presented due to space limitations:

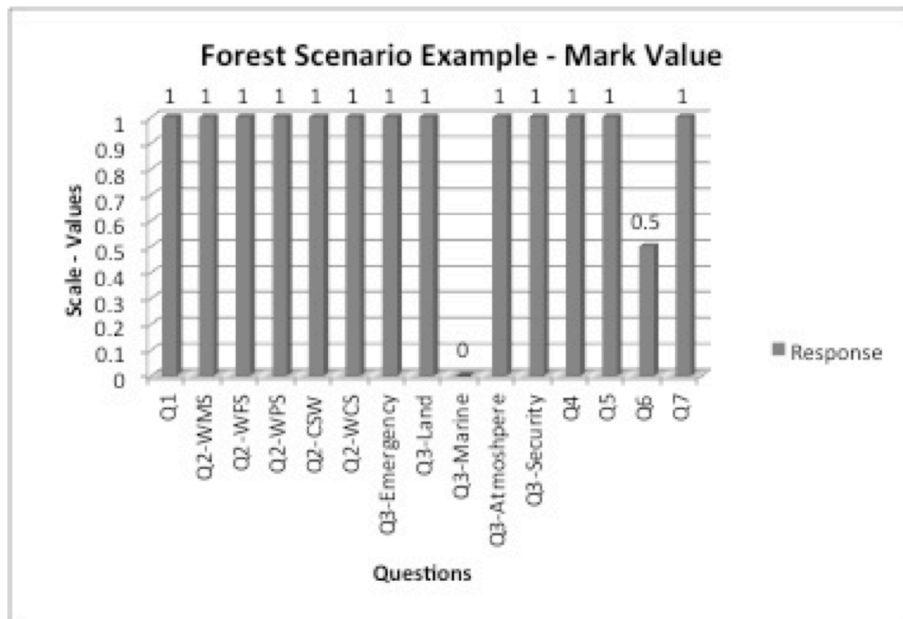
Evaluation Exercise 1: Following the CIM approach, associated questions were evaluated against well-specified response options and metrics e.g. [*Yes – 1.0, Partially - 0.5, No - 0.0*]. These metrics are generalised from ‘1’ to ‘0’, i.e. ‘1’ represents complete satisfaction, that the evaluator fully agrees, and is satisfied with the capability of HUMBOLDT tools and scenarios. ‘0’ indicates that it is not well addressed by the HUMBOLDT tools and scenario implementation. For each question, a ‘final mark’ is extracted – that represents the average value of results provided by the evaluators on which basis the main conclusions were formulated.

An aggregated response for all questions in Evaluation Exercise 1 is summarised and analysed in Table 2 below. Furthermore, Figure 7 (a) depicts an example of the evaluation response for the Forest scenario which shows that the overall response is

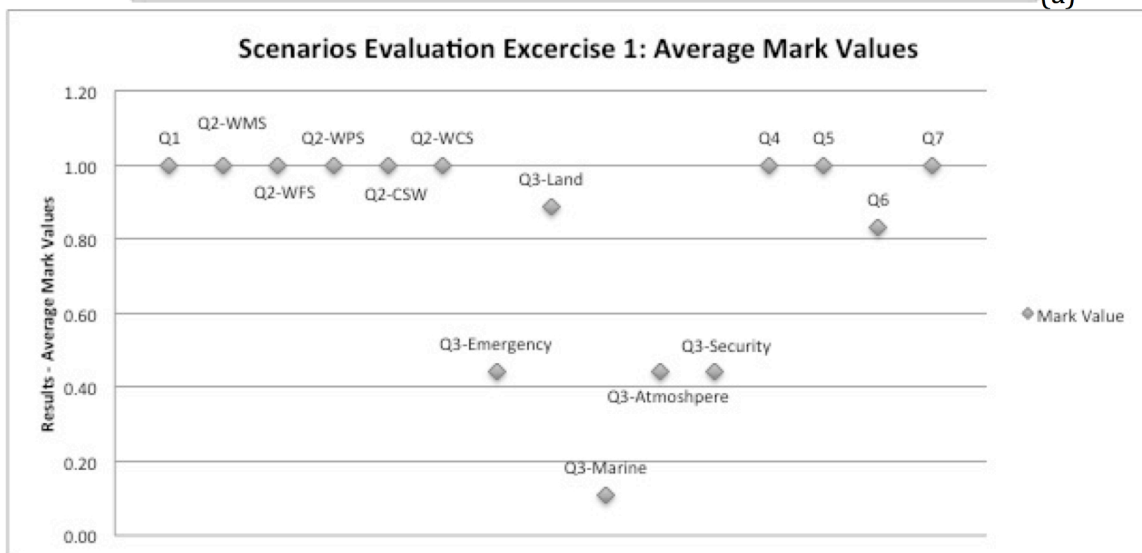
positive. It has mark value '0' for Q3-Marine because the Forest scenario is not relevant to the GMES Marine theme. Similarly it has mark value 0.5 for Q6 as evaluators expect modifications in the scenario solution, before it can be adopted for other projects. Similar responses to the Forest scenario evaluation responses are recorded for all scenarios, and a combined result of all scenarios is shown in Figure 7 (b). Figure 7(b) indicates that most of the evaluation results lie above the satisfactory level. The only concern is indicated in respect of Q3 because not all HUMBOLDT scenarios are related to all GMES application themes. For example, the Ocean scenario is only useful for the GMES Marine theme.

Table 2: Selected Examples from Evaluation Exercise 1

Questions	Final Mark (Metric Value - Average Result)	Summary and analysis of responses
Are selected use cases for the scenario demonstrator relevant to the domain specific real-life tasks?	1.0	All answers were 'Yes' which infers the suitability of use cases selected for the scenarios were relevant to domain specific real-life tasks.
If the dataset (that is used in the scenario prototype) were accessible through web services, would it be useful for you and/or other related initiatives?	1.0	All answers were 'Yes' which indicates that availability of data through OGC services (Web Map Service – WMS, Web Feature Service – WFS, Web Processing Service – WPS, Web Catalogue Service – CSW, Web Coverage Service – WCS) will be extremely useful for a variety of stakeholders.
Is the scenario relevant for one of the Global Monitoring for Environment and Security (GMES) Services application areas?	0.47	Results varied between 'Yes' and 'No' according to the relevance to specific GMES thematic areas (Emergency, Land, Marine, Atmosphere, Security), for example, the Air Quality scenario is less well related, whereas Forest management and the Risk management scenarios can be relevant to most of the GMES themes.
Do you think that HUMBOLDT tools and scenarios are useful for Global Monitoring for Environment and Security (GMES) related initiatives/services?	1.0	The results indicate that the thematic areas of the HUMBOLDT scenarios and HUMBOLDT tools and services are very well related, and useful to the GMES core services.
Do you think that the implementation of scenario prototypes leads to sharing of knowledge among scenario stakeholders and the relevant user community?	1.0	All answers were 'Yes' which indicates that the major outcome from the implementation of the scenario demonstrators is the experience and knowledge that could be utilised for the further development of related applications.
Would you like to reuse the scenario prototype?	0.83	Forest, Risk and Transboundary Catchment management scenarios are most likely to be adapted with modifications, but others can be reused as they are.
Which of the HUMBOLDT Framework components and scenarios training activities would you prefer to see included in the HUMBOLDT dissemination and exploitation period?	1.0	For Scenarios, all answers were 'Yes' which indicates that various stakeholder groups would like to see scenario training activities and further dissemination during the HUMBOLDT extension period. For the HUMBOLDT Framework components, except IGS all other components are useful and have the value 'Yes'.



(a)



(b)

Figure 7: Scenario Evaluation Feedback – Exercise I

Evaluation Exercise II: Due to space limitations, other selected questions and results from Evaluation Exercise II are depicted in Figure 8.

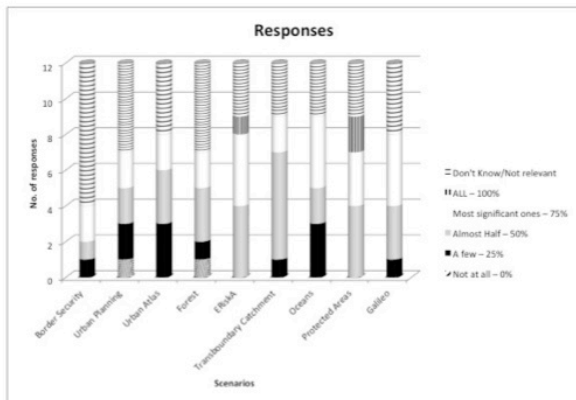
(a) demonstrates the usability aspects, such as learnability, of the scenario. This indicates again a great variation of responses, which may be attributed to various factors including i) knowledge of the specific scenario domain and/or its relevance; ii) the amount of material presented, and the approach adopted to present the training material/demonstration e.g. video, live demo, screen shots, descriptive, etc. In general, the overall response is in the range 50% - 75%, which seems very promising considering the timing constraints and other factors such as availability of demonstrators etc.

(b) demonstrates that the responses for this question were very positive. Based on the participant feedback, it may be concluded that the scenario demonstrators are closely related to domain specific real-life tasks, and their implementation, using the HUMBOLDT Framework, can contribute to further development of harmonised solutions for domain-specific problems. Furthermore, the responses indicate, that for most of the scenarios, relevance to the targeted user communities is sufficiently well identified, and represented in the scenario documentation including the training material.

(c) demonstrates that, based on the material available for the evaluation, most of the evaluators agreed that conformance to the standards for metadata and services in the scenario prototype implementation, can lead to improved interoperability, quality access and sharing of information within the respective scenario thematic domains.

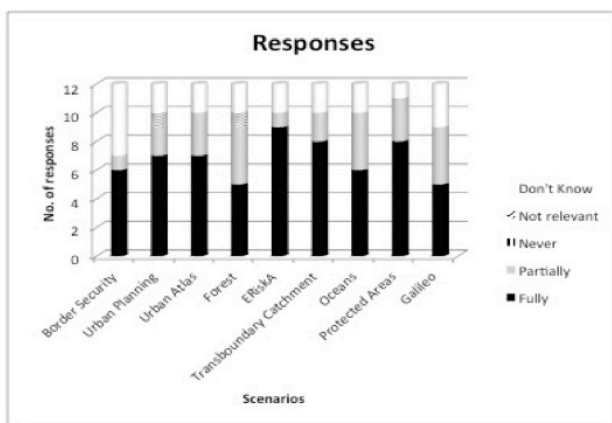
(d) demonstrates that most of the evaluators agree that conformance to INSPIRE standards can lead to a better sharing of geospatial data between cross-thematic application domains. Use cases from different HUMBOLDT scenarios capture various cross-border harmonisation issues, for example spatial data representation, heterogeneous data sources, thematic sectoral boundaries, etc.

Question: What proportion of the functionality of a scenario demonstrator could you successfully perform following a demonstration or tutorial?



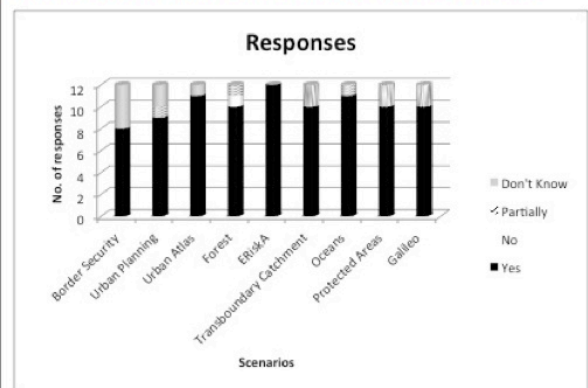
(a) Scenario Demonstrator Learnability

Question: Does conformance to INSPIRE, OGC and ISO standards for metadata and services implementation in the scenario prototypes lead to improved interoperability, quality, access and sharing of information?



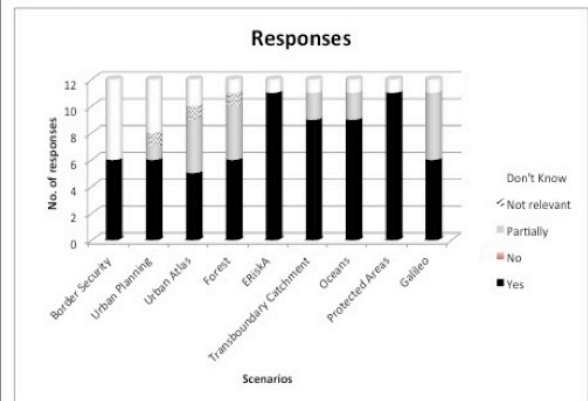
(c) Compliance to INSPIRE, OGC and ISO for Knowledge Sharing

Question: Are the selected use cases for the scenario demonstrators relevant to the domain specific real-life tasks?



(b) Relevance of the Scenario Use Cases to Real-Life Tasks

Question: Does conformance to INSPIRE standards improve the feasibility of cross-border data usage in the different HUMBOLDT scenario applications?



(d) Compliance to INSPIRE for Cross-Border Data Harmonisation

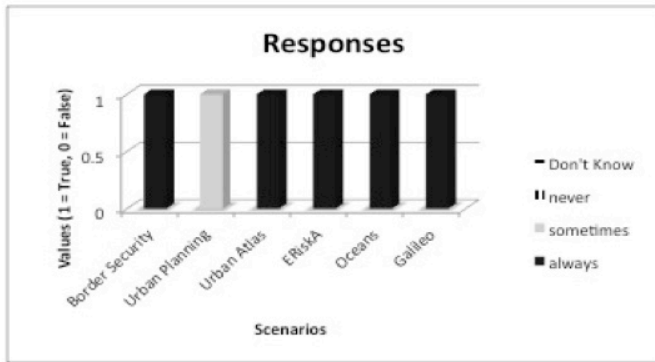
Figure 8: Selected Examples from Evaluation Exercise II

Evaluation Exercise III: The final scenario evaluation exercise primarily focused on the assessment of usability i.e. understandability of the scenario documentation and learnability; compliance to standards i.e. INSPIRE, OGC, ISO, etc; functionality i.e. harmonisation issues handled, use of HUMBOLDT components, common data model creations, etc; relevance to targeted communities and benefits i.e. knowledge sharing, cross-border data usage, etc. Due to space considerations, only selected examples from the Exercise III Evaluation are presented as shown in Figure 9. Figure 9 demonstrates that the evaluation results for most of the scenarios are satisfactory, excepting the urban planning scenario. The

major shortcoming identified for the urban planning scenario was the lack of demonstration, and clear and comprehensive presentation of the implementation of the scenario in the relevant documentation. This led the evaluators to score mostly towards '*partially*', '*sometimes*', '*no*' and '*don't know*' metrics that resulted in a low mark value for the scenario. The principal lesson learned here is that independent evaluation stakeholders rely heavily on the availability of the right tools e.g. software demonstrators, documentation etc at the right time to perform correct assessments. This experience moreover suggests that the development and implementation process of software demonstrators must be completed in time i.e. before the evaluation process commences.

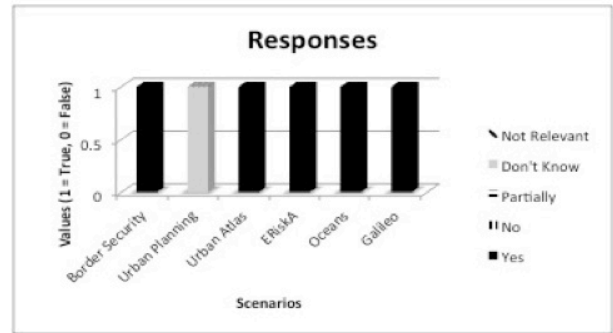
Apart from the urban planning scenario, analysis of the Evaluation Exercise III generally indicates that the scenario demonstrators provide suitable examples of domain-specific geospatial harmonisation problem solving using HUMBOLDT Framework tools. Furthermore, the evaluation exercise also indicates that the scenario training materials, which present in detail the applied solutions for harmonisation problems in real-world environments, are usable and demonstrate the use of various HUMBOLDT components in dealing with different spatial data harmonisation aspects, and which would be useful for the general geospatial community in understanding and adapting the HUMBOLDT solutions to related problems. As the development of the scenario services (prototype) and data modelling is compliant with existing standards, as defined in the HUMBOLDT Handbook of Standards (2010) (e.g. INSPIRE, OGC, ISO, UML, etc.), it would be possible for the external geospatial community to reutilise the existing components and/or extend them for their customised use, where possible e.g. introducing new services, adding new data sources, etc.

Question: Is the scenario prototype documentation e.g. training material understandable?



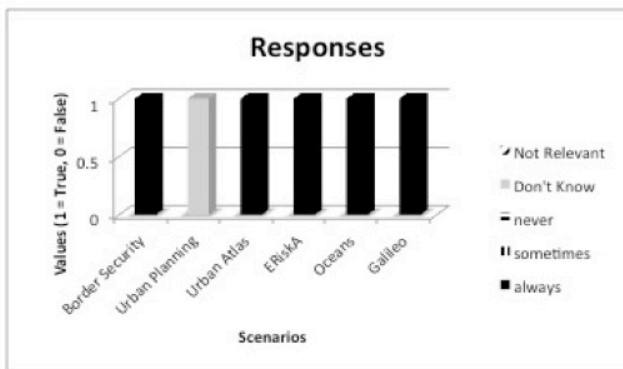
(a) Understandability of Scenario Documentation

Question: Is every service in the scenario prototype compliant with the standards listed as “mandatory” and/or “recommended” in the HUMBOLDT handbook of standards?



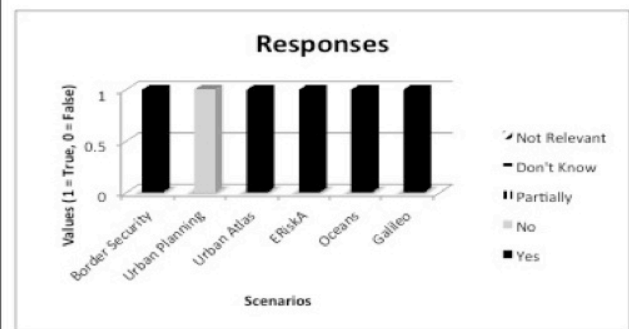
(b) Compliance to Identified Standards

Question: Do you think that it is possible to add new data sources?



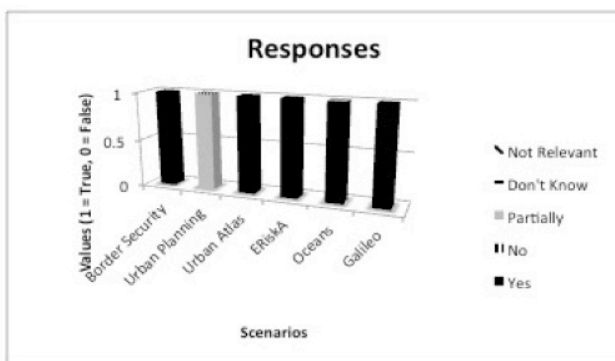
(c) Functional Testing: Adding New Data Sources

Question: Does scenario prototype documentation (training material) provide a detailed, clear and understandable functional description of the scenario application?



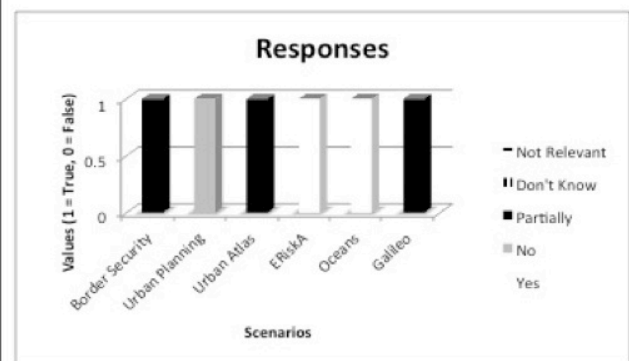
(d) Clarity of Functional Description

Question: Does the documentation of the scenario indicate that specific data harmonisation issues are covered by the scenario prototype using INSPIRE components?



(e) Coverage of Data Harmonisation Issues by Scenarios

Question: Do you think that scenario documentation provides sufficient material to allow the relevant end users to understand and reuse the scenario for their own research?



(f) Understandability and Reusability of Scenarios

Figure 9: Selected Examples from Evaluation Exercise III

6. Critical analysis of the Evaluation Methodology and Lessons Learned

This section provides a critical reflection on the CIM methodology and its use in the HUMBOLDT project, by summarizing the results of the evaluation exercises. In general, the evaluation methodology proved suitable for the HUMBOLDT project user evaluation, and also on this evidence appears to be of substantial benefit for other large scale integrated projects.

Based on the application of the CIM methodology, the following advantages are identified:

Prime advantage that it is **simple and easy to develop and apply** in a collaborative environment where access to resources (i.e. data and software licensing, etc) are limited. Furthermore, in respect of conformity to standards, it has compliance to standards such as ISO 9126 that permits the **targeting of different user communities**.

In the HUMBOLDT context only four evaluation criteria (i.e. functionality, usability, benefits and relevance) were considered for user evaluation, nonetheless the methodology permits the use of a wide variety of evaluation criteria (e.g. non-functional aspects), according to project requirements. This confirms that the methodology is **extendable**. Furthermore, according to evaluation objectives, various evaluation techniques may be adopted e.g. laboratory-tests, behavioural validation using black-box testing, as well as questionnaire approaches. This suggests that the methodology is also **flexible**. The methodology may also be used in an **incremental and iterative** manner, to establish a feedback loop, and thereby improve project outcomes. In this manner, initial criteria, indicators and metrics can be developed from system specification documents, and on the basis of user engagement in subsequent stages can be developed to refine the methodology and outputs.

The CIM methodology benefits from **rigorous planning** at the beginning of the project, with effective user input, in order to prepare appropriate and relevant criteria, indicators and metrics. This process ensures commitment from various stakeholders to provide **access to specific resources** e.g. data which is not always free, in order to facilitate the assessment process. Furthermore, the methodology facilitates the introduction of **specific requirements based indicators** from pre-defined criteria to assess whether or not a particular project requirement has been met. However, the **allocation of weights** to the specific metrics of questions and/or indicators requires careful consideration as the perception and importance of these indicators varies from one user community to another. An **early stage implementation** of the CIM methodology for different software engineering process models such as Agile, Rapid Application Development, Prototyping, etc can be useful in obtaining early user feedback.

Furthermore, generalisation **of this evaluation methodology** to large-scale integrated projects, where many stakeholders are involved in the development of software systems, permits it to effectively perform this type of project evaluation as well. For example, the discovery of common aspects, to be evaluated, for the various HUMBOLDT scenarios makes it easier to perform such user evaluation with limited resources and time.

Finally, in terms of access restrictions, especially for collaborative and integrated projects, it is not always possible to obtain access to data and proprietary softwares to perform the evaluation. In HUMBOLDT, in relation to some scenarios, there were strict data protection and licensing issues which restricted the performance of more rigorous evaluation e.g. white-box testing by independent evaluators. But the CIM approach adopted permitted the deployment of alternative resources to be utilised for user evaluation. For example, as shown in Figure 5, scenario deliverables including training materials for specific application cases, demonstrated the use of screenshots, screen recordings (audio/video), and detailed user oriented documentation, etc that were utilised for the evaluation of certain scenarios.

Based on the application of the CIM methodology, the following disadvantages are identified which could be considered as the basis for future research :

Despite the flexible and extendable nature of the CIM methodology, issues arise in respect of the **readiness of system resources for evaluation**. This is especially the case for large scale integrated projects in which delays in delivering the required system resources can also delay the overall evaluation process. For example, most of the evaluation activities were carried out in the later stages of the HUMBOLDT project due to delays in completion of system development tasks. Such delays further complicate the evaluation process when collaboration is necessary to perform evaluation by independent evaluators.

In the case of HUMBOLDT evaluation, most of the questions were of nominal-polytomous (i.e. more than two unordered answering options) but the some dichotomous type questions (two answering options) appear to be rigid, and make it difficult for the stakeholders to judge and select appropriate answers. This suggests that a flexible metric scale (e.g. bounded-continuous type – continuous scale) or system usability scale (Brooke J 1996; Finstad K 2010) could reveal detailed and more fine-grained insights about the evaluation outcomes and impact assessment.

Finally, as the HUMBOLDT project aimed to harmonise cross-border and cross-thematic spatial data from the GMES related application domains i.e. scenarios using INSPIRE principles, a major concern was to include questions relating to both INSPIRE and GMES domains. However, this questionnaire strategy was undermined to the extent that **knowledge gaps** exist for both INSPIRE and GMES communities in respect of the activities and developments taking place in the other domains. For

example, not all the participants who are active in INSPIRE domain, were fully aware of the GMES initiative and its basic thematic areas and technical developments. Likewise, participants who are active in the GMES domain, were not fully aware of developments in the INSPIRE domain, e.g. basic awareness about data and metadata specifications and different services such as download, visualisation, discovery, etc. Consequently the need was identified, for **more integrated actions** to bridge the gaps between both communities. Furthermore, this diversity of participants from different domains made it difficult to decide to what extent the scenario material to be evaluated should capture evaluation aspects relevant to both INSPIRE and GMES domains. Based on this experience it is suggested that a **more domain/community specific approach** (e.g. techniques presented by Bryson J, et. al. (2011)) should also be accommodated to acquire more constructive feedback and a more effective evaluation methodology.

It is suggested that more effective feedback from the questionnaire could be secured by the greater contextualisation of questions. For example, provision of additional context related information targeting respondents from one domain, where questions concern issues arising from another domain. In particular, there is an opportunity to secure this contextualisation at evaluation planning, design and implementation stages where the CIM methodology requires in-depth engagement with stakeholders. At these stages awareness raising training could be provided, targeting state-of-the-art developments in related domains. This could be facilitated by the HUMBOLDT training platform (2006 – 2011) which is designed to promote such awareness, and deliver detailed knowledge of the HUMBOLDT framework tools via online training for both basic and advanced user groups.

7. Discussion on the HUMBOLDT Evaluation

The application of the CIM evaluation methodology to the HUMBOLDT project has highlighted a number of issues in respect of both the HUMBOLDT framework tools as well as the scenarios. For the HUMBOLDT Framework tools, and in particular HUMBOLDT Alignment Editor, the average response was positive with useful feedback concerning the introduction of new features as well as the improvement of the system by correcting existing functionality. However, the main emphasis of the discussion concerns the results of the various user evaluation questionnaires for the different scenarios, concerning usability, functionality, benefits and relevance as follows:

-Usability: there were varying results for the different scenarios, but most reflect high levels of usability and some partial usability. None of the scenarios were considered to be of low usability, especially for the criteria of understandability, learnability, usefulness, relevance, etc. It should be noted that these results reflect an ongoing process of improvement throughout the project, which on the basis of feedback provided at various times, for example, at quarterly project meetings, in

total enhanced the usability aspects of the scenarios. The experience of using the HUMBOLDT Framework in solving various harmonisation problems in the scenario demonstrators has been captured in the form of training materials including images, videos/audios, screen recordings, etc, and accordingly the high degree of understandability of this material can be viewed as beneficial for the general geospatial community in understanding, adapting, and adopting the HUMBOLDT solutions. In general, it can be concluded that the usability aspects of the HUMBOLDT scenarios are well addressed by the scenario demonstrators and their associated documentation.

-Functionality: the functional aspects of the HUMBOLDT scenarios were relatively difficult to evaluate due to accessibility issues e.g. due to the use of various existing proprietary legacy software, licensing issues, data privacy and the delayed completion of the specified and implied tasks for the development of scenario demonstrators. As a result the production of scenario training material and scenario demonstrations with hands-on workshops, which took place at the above mentioned events, provided a suitable solution for the issue of accessibility of various resources in performing the functional assessment of the scenarios. As a substantial proportion of questions were based on functional assessment, the responses to these questions reflect average results, especially in the first two evaluation exercises, and this may be attributed to the ongoing demonstrator development activity which was not fully completed at the time of the evaluation exercises. The higher scores for these criteria were evident in the third evaluation exercise as outlined in the analysis section, above. In general, and as identified from the functional assessment, the development of scenario services (in a prototype) and data modelling, from the scenario specifications, is compliant with existing standards (as indicated in the HUMBOLDT handbook of standards (2010)). This suggests that it would be possible for the external geospatial community to reuse the existing components and/or extend them for their customised purposes, for example introducing new services, adding new data sources, etc.

-Benefits and Relevance: A major part of the scenario evaluation questionnaires were dedicated to the assessment of aspects of benefit and relevance. In all the evaluation exercises the results were very promising and demonstrate and then the substantial relevance and benefits of the scenario demonstrators in achieving HUMBOLDT objectives, in solving specific harmonisation problems, and in demonstrating the usefulness and relevance to the INSPIRE and GMES communities. In general, the experience of using HUMBOLDT components in various scenarios, shared with other SDI (or geospatial) communities via the training materials, provides a unique opportunity to acquire state-of-the art knowledge on various spatial harmonisation issues concerning risk management, the ocean, air quality, protected areas, trans-boundary catchments, forest, urban planning, urban atlas and border security. Overall the HUMBOLDT scenario demonstrators provide a proof of concept in solving a subset of identified harmonisation problems, by using the HUMBOLDT Framework. Furthermore, the HUMBOLDT Framework and the scenario demonstrators have established a platform for solving other major geospatial data harmonisation problems discovered during the lifetime of the project, or not considered within the scope of the project.

8. Conclusion

Evaluating the large scale HUMBOLDT project was not straightforward as a consequence of accessibility issues, including stakeholders at remote locations and resource constraints e.g. data licensing issues, coverage of diverse application domains, and different technical and application objectives. The CIM methodology has permitted the performance of user evaluation based on well-defined criteria, indicators and metrics, which identified the strengths and weaknesses of the HUMBOLDT Framework tools and its associated scenarios. Furthermore, the CIM evaluation methodology has been shown to present many advantages including simplicity, ease of development, and application in a collaborative environment where access to resources are limited, as well as flexibility and extendibility as the methodology permits the use of a wide variety of evaluation criteria, evaluation techniques, and deployment of alternative evaluation resources, for example associated training materials. Furthermore, the methodology permits rigorous planning from the outset of the project, and it may be used in an incremental and iterative manner permitting feedback from user communities to be incorporated within subsequent technical specifications, thereby improving the overall development outcomes.

At the same time analysis of the application of the CIM evaluation methodology identified a number of challenges and potential disadvantages. These include the need to apply the methodology at an early stage in the evaluation process, which particularly for large-scale integrated projects may not be feasible, especially, given the characteristic delays associated with large-scale and complex project activity. In addition, the allocation of weights to specific metrics may be subjective, and the perceptions of the importance of these metrics and other indicators to different user communities raise substantial challenges in the agreement of such weights.

Clearly these issues concerning the disadvantages and limitations of the CIM evaluation methodology will form the basis for further research to improve the effectiveness of the methodology and its wider application. However, a central consideration, beyond the specific challenges identified above, concerns the question of the extent to which the CIM evaluation methodology is generalisable to a wide range of application domains and associated user communities. There is some evidence from the application of the evaluation methodology in the HUMBOLDT project indicates that a more targeted domain specific approach to the application of the methodology would enhance its effectiveness, particularly in relation to the ability to effectively engage with the diverse user communities that is so essential to the proper application of the CIM evaluation methodology.

Conflicting Interests: There are no conflicting interests.

Funding and Acknowledgements:

This work was performed in the HUMBOLDT project (Duration: October 2006-March 2011). The HUMBOLDT project was funded by the European Commission Framework Programme Six (EC project grant reference: 30962). We also acknowledge contributions of HUMBOLDT partners at various stages of the HUMBOLDT project evaluation.

References

- Babar MA, Zhu L, Jeffery R. 2004, 'A Framework for Classifying and Comparing Software Architecture Evaluation Methods', *Proceedings of the 15th Australian Software Engineering Conference (ASWEC)*, IEEE Computer Society, pp. 309-319, Melbourne, Australia.
- Basili V, Donzelli P, Asgari S. 2004, 'A unified model of dependability: capturing dependability in context', *IEEE Software*, 21/6: 19–25.
- Basili V, Caldiera G, Rombach D. 1994, 'Goal Question Metric Approach, Encyclopedia of Software Engineering', pp. 528-532, John Wiley & Sons, Inc.
- Briand L, Morasca S, Basili V. 2002, 'An Operational Process for Goal-Driven Definition of Measures', *IEEE Transactions on Software Engineering*, 28/12: 1106-1125.
- Brooke, J. 1996 'SUS – A quick and Dirty Usability Scale'. In: Jordan, P., Thomas, B. and Weerdmeester, B. & A. L. McClelland (eds) *Usability Evaluation in Industry*, pp. 189–94. Taylor and Francis. London, UK.
- Bryson J, Patton M, Bowman R, 2011, 'Working with evaluation stakeholders: A rationale, step-wise approach and toolkit', *Evaluation and Program Planning*, 34/1, pp. 1-12, ISSN 0149-7189, 10.1016/j.evalprogplan.2010.07.001.
- CASCADOSS Consortium (ed.) 2007: 'Evaluation criteria for Open Source GIS and RS software' [Online], URL Access: http://www.cascadoss.eu/en/index.php?option=com_content&task=view&id=9&Itemid=11, Last Access: 4 December 2012.
- Dobrica L, Niemela E. 2002, 'A Survey on Software Architecture Analysis Methods', *IEEE Transactions on Software Engineering*, 28/7: 638-653.
- Falessi D, Babar MA, Cantone G, Kruchten P. 2009, 'Applying empirical software engineering to software architecture: challenges and lessons learned', *Empirical Software Engineering*, 15/3: 250-276.
- Fichtinger, A., Rix, J., Schäffler, U., Michi, I., Gone, M. and Th. Reitz, 2011, Data Harmonisation Put into Practice by the HUMBOLDT Project, *International Journal of Spatial Data Infrastructures Research*, 2011, Vol.6, 234-260.
- Finstad K, 2010, The Usability Metric for User Experience, *Interacting with Computers*, 22(2010), pp. 232-327.
- GMES, 2012, Global Monitoring for Environment and Security, URL Access: <http://www.gmes.info/> Last Access: 4 December 2012.
- Gustedt J, Jeannot E, Quinson M. 2009, 'Experimental Validation in Large-Scale Systems: a Survey of Methodologies', INRIA NANCY - GRAND EST Report: inria-00364180, version 1, 25 Feb 2009. p.16.

HUDSON, 2010, Extensible continuous integration server, URL Access: <http://hudson-ci.org/>, Last Access: 4 December 2012.

HUMBOLDT Framework Tools, 2010, URL Access: http://www.esdi-humboldt.eu/files/0982-a5_2-d3_3_0_a_lightweight_introduction-fhg-igd-004-final.pdf Last Access: 4 December 2012.

HUMBOLDT Handbook of Standards, 2010, URL Access: http://www.esdi-humboldt.eu/files/0996-a3_6-d4_handbook_of_standards_-_final_version-ethz-002-final.pdf Last Access: 4 December 2012.

HUMBOLDT Scenarios, 2010, URL Access: <http://www.esdi-humboldt.eu/scenarios.html> Last Access: 4 December 2012.

HUMBOLDT Technical Evaluation, 2010, URL Access: http://www.esdi-humboldt.eu/files/1109-a10_4-d1_report_on_end-user_or_developer_software-fomi-001-final.pdf Last Access: 4 December 2012.

INSPIRE, 2007, URL Access: <http://inspire.jrc.ec.europa.eu/> Last Access: 4 December 2012.

ISO/IEC 9126-1:2001 Software engineering - Product quality - Part 1: Quality model.

Juristo N, Morant J. 1998, 'Common framework for the evaluation process of KBS and conventional software', *Elsevier Knowledge-Based Systems*, 11/2: 145-159.

Juristo N. 1997, 'Structuring the Knowledge-Based Systems Evaluation Process', *Proceedings of Intelligent Information Systems'97*, pp. 266-270.

Kotonya G, Sommerville I. 2002, *Requirements Engineering: Processes and Techniques*, ISBN # 0-471-97208-8, John Wiley & Sons Inc, Chichester, UK

Khan, Z., Ludlow, D., Rix, J., 2012, 'Applying the criteria indicators and metrics evaluation method on ICT research: The HUMBOLDT project', *Research Evaluation*, DOI: 10.1093/reseval/rvs037. .

Khan, Z., Odeh, M. and McClatchey, R. 2011, 'Bridging the gap between business process models and service-oriented architectures with reference to the grid environment'. *International Journal of Grid and Utility Computing*, 2/4: 253-283.

Lindvall M et al. 2007, 'Experimenting with software testbeds for evaluating new technologies', *Empirical Software Engineering*, 12: 417-444, DOI 10.1007/s10664-006-9034-0.

OGC, 2012, Open Geospatial Consortium, URL Access: <http://www.opengeospatial.org/> Last Access: 4 December 2012.

Ostreika A, 2010, Evaluation of Benefit Analysis – The HUMBOLDT Project Deliverable A10.5, URL Access: http://www.esdi-humboldt.eu/files/0885-a10_5_evaluation_of_benefit_analysis-ktc-004-final.pdf Last Access: 4 December 2012.

RESPECT Consortium (ed.) 1998: *User-Centred Requirements Handbook. Version 3.3.* URL Access: <http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/UserCenteredRequirementsHandbook.pdf>, Last Access: 4 December 2012.

RISE (ed.) 2007: *Report on the Operation of the RISE Testing System.* URL Access: http://www.eurogeographics.org/eng/documents/RISE36_Testing_System_Report_V1.0.pdf, Last Access: 4 December 2012.

Sommerville, I. 2006, Software Engineering, ISBN # 9780321313799, Addison-Wesley, Wokingham, UK.

Sonar: Code Analysis server, URL Access: <http://docs.codehaus.org/display/SONAR/Documentation>, Last Access: 1 November 2012.

The HUMBOLDT Project, 2006-2011, Sixth Framework Programme, European Commission, URL Access: <http://www.esdi-humboldt.eu/home.html> Last Access: 4 December 2012.

The HUMBOLDT Project Community, 2006-2011, URL Access: <http://community.esdi-humboldt.eu> Last Access: 1 November 2012.

The HUMBOLDT Training Platform, 2006-2011, URL Access: <http://www.esdi-humboldt.eu/training.html> or <http://www.gisig.it/humboldt/training/> Last Access: 1 November 2012.

Tiits, K. 2003, Usability of Geographic Information Systems in Internet. A case study of Journey Planners. Master thesis at Tartu University, Institute of Geography. URL Access: http://www.hot.ee/kyllitiits/Kylli_Tiits_MSc_thesis.pdf, Last Access: 4 December 2012.

Villa P, Molina R, Gomasasca M, and Roccatagliata E, 2012 'Harmonisation Requirements and Capabilities Towards a European Spatial Data Infrastructure (ESDI): The HUMBOLDT Protected Areas Scenario', International Journal of Digital Earth, 5/5: 417:38.

WebPark consortium (ed.) 2002: Usability Testing Template, D2.2.3, V05. URL: http://mp.mountaintrip.eu/uploads/media/deliverable/webpark_d3.pdf, Last Access: 4 December 2012.