EVALUATION OF LARGE PUBLIC SECTOR IS PROJECTS: EXPERIENCE FROM THE ICTE-PAN PROJECT

Euripides Loukis[1], *Thom Pick*[2], *Karel Van Isacker*[3]

Abstract. Despite the high ambitions of the e-government programs, which are in progress in many countries, aiming at the development of highly complex information systems in public administration, the reality is quite different: many failures of information systems (IS) projects, and especially of the largest ones, are reported, which result in waste of huge financial resources, loss of opportunities, disappointment of the public and heavy criticism by the press. If these failures are not drastically reduced, the progress towards e-government, and in general towards the 'information society', will be significantly retarded. Appropriate evaluation of these IS projects can highly contribute to the reduction of these failures and to the increase of the benefits from the IS of public sector. However, although extensive theoretical research has been conducted in the area of IS evaluation, which has analyzed its inherent difficulty and complexity and has produced several IS evaluation directions and frameworks, only limited research has been conducted towards their practical application in large and complex 'real-life' projects. In the present paper is analyzed the experience gained from the evaluation of a complex inter-organizational G2G collaboration support environment, which has been developed in the ICTE-PAN project of the IST Program of the European Union. We present a user-driven evaluation methodology, which has been developed for this purpose, based on the ISO/IEC 14589 and ISO/IEC 9126 standards. Also we analyze the application of this methodology for the evaluation of the above large inter-organizational IS by the user-partners of this project, its results and also the conclusions drawn from the whole evaluation process.

1. Introduction

Public and private sector organizations are making big investments for the development of various kinds of information systems (IS) in order to support their internal functions and their communication and transaction with their external environment and finally achieve various kinds of both operational and strategic business benefits. However, many IS development projects fail to deliver the expected technical performance, functionality and business benefits within budget and schedule, or even are abandoned. During the last 25 years high failure rates of IS projects are continuously reported and investigated in the relevant literature. McFarlan [1] in 1981 reported that "Despite businesses' more than 20 years of experience with

[1] University of the Aegean, Department of information and Communication Systems Engineering, Gorgiras Street, 83200 Karlovassi, Island of Samos, Greece, e-mail: eloukis@aegean.gr
[2] Ministry of Environment of Lower Saxony, Archivstraße 2, 30169 Hannover, Germany, e-mail: thom.pick@mu.niedersachsen.de
[3] European Dynamics, 209 Kifissias Av. & Arkadiou Str., 15124 Maroussi, Athens, Greece, e-mail: karel.van.isacker@eurodyn.com

information systems, disasters in that area still occur with surprising regularity". The widely quoted CHAOS Report [2] by the Standish Group reports that 31.1% of the software projects are abandoned during the development cycle, while 52.7% of them, although they are completed and become operational, suffer serious budget overruns and/or schedule slips and/or offer less functionality and features than initially specified, while only 16.2% of the software projects are finally successful. Even though most of the literature about IS projects failure is focused mainly on enterprises of the private sector, Government organizations experience similar problems. It has been estimated that in the UK the IS projects failures in Government organizations have costed the taxpayers £5 billion over twelve years [3]. According to the Cabinet Office [4] of UK "Government IT projects have too often missed delivery dates, run over budget or failed to fulfil requirements" and for this reason the SPRITE (Successful PRojects in an IT Environment) Program (www.ogc.gov.uk) has been established for "improving the management, success and realization of government IT-enabled business change". Similar failures have been also reported in Greece in many relevant reports and official documents, concerning the big IS investments, which have been made in Government organizations since the early 90s, financed from National and European Union funds, e.g. from the Integrated Mediterranean Programs, the European Community Support Frameworks, etc. ([5], [6], [7]). Despite the high ambitions of the e-government programs, which are in progress in many countries, aiming at the development of highly complex both internal systems and "outward-looking" IS in public administration, the reality is quite different: there are many failures of information systems (IS) projects, and especially of the largest ones, resulting in waste of huge financial resources, loss of opportunities, disappointment of the public and heavy criticism by the press. As mentioned by Lyytinen and Robey [8] "information systems development remains a high-risk proposition", while "organizations failed to learn from their experience in systems development". If these failures are not drastically reduced, the progress towards e-government, and in general towards the 'information society', will be significantly retarded.

Appropriate evaluation of the public sector IS projects, particularly of the large ones, in all the stages of their life cycle, can highly contribute to the reduction of these failures and to the increase of the benefits which are offered by the IS to the citizens, the enterprises and the public servants. According to Smithson and Hirschheim [9] as IS evaluation is meant "the assessment or appraisal of the value, worth or usefulness of an information system", being focused "on the potential or actual costs and benefits of introducing a new system", and in a wider sense on the development of "a wider view of the impact of new system". Farbey et al [10] define IS evaluation as "a process that takes place at different points in time or continuously, for searching for and making explicit, quantitatively or qualitatively, all impacts of an IT project". Pre-implementation IS evaluation, usually as part of a feasibility study, can be quite useful for the optimal selection of the appropriate IS investments and for the optimal definition of their scope and functionality. Intermediate "formative" evaluation gives a feedback on the IS under development, while it is being developed, enabling an early diagnosis of problems, malfunctions and deviations from the initial objectives of the IS, budget overruns, etc., so that appropriate corrective actions can be taken in time if necessary, before further deterioration happens; also it assists the subsequent design decisions. Post-implementation "summative" evaluation enables us to investigate to what extent the initial objectives of the IS have been achieved and whether it can serve the purpose it has been built for; it also enables us to take appropriate corrective actions (e.g. improvements, modifications, enhancements, etc.) if necessary before the IS productive operation starts, and avoid possible disruptions of significant organizational business processes. Moreover, it offers to opportunity to investigate whether sufficient value for the taxpayers' money spent in the project has been produced, to evaluate the project management and to facilitate organizational

learning. Especially for the "outward-looking" IS of public administration, e.g. for IS that offer the citizens and the enterprises the capability to perform electronically their transactions with the public administration (such as declarations, applications, etc.) via the Internet and other electronic channels, their post-implementation evaluation is absolutely necessary before the official start of their productive operation, for safeguarding the credibility of the public administration. Extensive theoretical research has been conducted in the area of IS evaluation (reviewed in section 2), which has analyzed its inherent difficulty and complexity and has produced several IS evaluation directions and frameworks; however, limited research has been conducted towards their practical application in large and complex 'real-life' projects.

The present paper analyzes the experience gained from the intermediate (formative) and the post-implementation (summative) evaluation of a complex inter-organizational G2G collaboration support environment, which has been developed in the ICTE-PAN project (Methodologies and Tools for Building Intelligent Collaboration and Transaction Environments for Public Administration Networks) of the Information Society Technology (IST) Program of the European Union. This experience can be very useful for the design and implementation of the evaluation processes of large and complex national or international IS projects. In section 2 a brief review of the literature on IS evaluation is presented. It is followed by a description of the ICTE-PAN project and the G2G collaboration support environment MERMIG, which has been developed in this project. The evaluation methodology is presented in section 4, while the results of the evaluation of MERMIG environment are presented in section 5. Finally the conclusions are discussed in section 6.

2. Background

Extensive research has been conducted in the last 25 years on the evaluation of IS ([9] - [16]), motivated both by the big IS investments being made by most private and public organizations, which necessitate an investigation of the value they produce, and also by the inherent difficulty and complexity of IS evaluation:

- the benefits and in general the value created by most categories of IS are complex and multidimensional, both tangible and intangible, so it is difficult to decide "what to measure" for the evaluation and "how"; only for IS which aim mainly at the automation of complex and labor intensive tasks it is possible to quantify in monetary terms most of the benefits they create (mainly labor cost reductions), while for other IS categories (e.g. for management information and decision support systems, inter-organizational systems, etc.) this is quite difficult or even impossible,

- different IS categories have quite different objectives and produce different types of benefits and value, so they require different kinds of "measurements" and evaluation methods; therefore there is not a generic "best IS evaluation method" suitable for all situations, and the best we can do is to develop IS evaluation directions and frameworks (most of them being suitable for a specific IS category), which can be customized and elaborated for each particular IS evaluation we have to perform,

- moreover, IS usually involve multiple stakeholders (e.g. various levels of management, various groups of users, IS experts, project team, etc.), with different concerns, value systems and agendas, so IS evaluation has to take into account all these different perspectives,

- also the costs of IS are complex and multidimensional, both tangible and intangible; some of them are direct, while some others are indirect, and should be appropriately allocated.

According to Smithson & Hirschheim [9] the extensive research literature on IS evaluation can be grouped into five "levels" of analysis, which correspond to the five levels of IS impact: the macro level (national or international perspective), the sectoral level (impact of IS on the productivity and the distribution of sales and profits in specific sectors), the firm level (impact of IS on the performance of individual firms – 'productivity paradox' literature), the application level (impact and value of particular applications) and the stakeholder level (different impact and value of IS on different stakeholder groups). The present paper focuses on the "application level", as it attempts to evaluate a particular G2G collaboration support IS (described in section 3).

For all the above reasons IS evaluation has long been recognized as a difficult and complex task. F. Land in 1976 wrote one of the first reviews on IS evaluation and its basic characteristics, issues and difficulties [11]. During the 80s increase exponentially not only the penetration and the use of IS in private and public organizations, but also the complexity and the sophistication of IS; new categories of IS are introduced in most organizations in order to support more complex and sophisticated organizational functions. Therefore the theory and the practices of IS evaluation had to evolve as well. Hirschheim & Smithson in 1988 critically review these evolutions in a paper presented at the IFIP WG 8.2 Working Conference on "Information Systems Assessment" [12]; in the same paper they propose a useful framework for classifying the numerous IS evaluation methods and relevant literature. Ten years after, in 1998, the same authors wrote an "updated" paper on the same topic [9], which reflects the further advances and evolutions of IS evaluation, which had taken place, since the whole business and technological context in which IS were developed and used had become much more demanding and complex. In this paper an "updated" form of the above classification framework is proposed, which classifies IS evaluation methods and relevant literature into two basic categories. The first category consists of efficiency-oriented methods, which have been influenced mainly by engineering approaches. They evaluate the performance or quality of an IS with respect to some detailed specifications, being concerned mainly with the question "is it doing things right?" These methods are based on various software reliability inspection or quality assurance methods, metrics and standards. The second category consists of effectiveness-oriented methods, which have been influenced mainly by management science approaches. They evaluate how much an IS supports the execution of business-level tasks or the achievement of business-level objectives, being concerned with the question "is it doing the right things?" as well. A typical example of such an effectiveness-oriented IS evaluation method is the "Service Quality Gap Analysis", which identifies and measures the gaps between the actual performance and the expectations of users and developers [17].

Farbey et al [13] provide a framework, named the "benefits evaluation ladder", for classifying IS according to the method required for evaluating of the benefits they offer. It consists of the following eight IS categories, named "ladder rungs": mandatory IS, automation IS, direct value added IS, management information and decision support systems (MIS - DSS), infrastructure IS, inter-organizational IS, strategic IS and business transformation enabling IS. Moving up the ladder increase the potential benefits, but also increase the uncertainty of outcomes, the risk of failure and the difficulty-complexity of benefits evaluation. For each of the above rungs a different evaluation method is proposed: while in the lower rungs (e.g. for mandatory or automation IS) the evaluation is based on the precise quantification of benefits and costs, in the higher rungs (e.g. for strategic or business transformation enabling IS) the evaluation (especially the initial pre-implementation evaluation) is mainly judgemental and

requires the involvement of the higher management. Subsequent research literature in this area ([14] – [16]) keep emphasizing the need for IS evaluation methods specialized to specific types of IS (e.g. IT infrastructure) or even to specific industries, which take into account their particular objectives; however, it is also reported that due to these difficulties and complexities IS evaluation is often neglected or carried out incorrectly and ineffectively in many organizations, as managers believe that it takes too long, costs a significant amount of money, involves too many people with departmental or individual agendas and very often does not create significant value.

3. The ICTE-PAN project

The project ICTE-PAN (Methodologies and Tools for Building Intelligent Collaboration and Transaction Environments for Public Administration Networks) aimed at developing an electronic collaboration environment for supporting the various G2G collaboration typologies of modern public administration. Its total budget was 3,191,023 Euro, partially funded by the Information Society Technologies (IST) Program of the European Union (contract IST-2001-35120); it started on 1st March 2002 and had a duration of three years. The project has been implemented by a consortium of technology providers and public organizations-users, consisting of European Dynamics SA (Greece), University of the Aegean (Greece), TXT e–Solutions SpA (Italy), National Environment Research Institute (Denmark), Ministry of Environment of Lower Saxony (Germany) and Province of Genoa (Italy). It included nine workpackages shown in Table 1.

WP0	Project Management
WP1	State of the Art Analysis
WP2	Users Requirements Analysis
WP3	Public Administration Modeling
WP4	System Design & Development
WP5	System Integration
WP6	Pilot Implementation
WP7	Evaluation
WP8	Dissemination – Exploitation Plan

Table 1: Workpackages of the ICTE-PAN project

As we can see from the above Table the ICTE-PAN project included the design and development of a G2G collaboration support environment (WP4, WP5), which was named MERMIG (it is the Greek word for "ant", as ants are characterized by intensive collaboration), and the implementation of the following four pilots (WP6) in order to test it:

i) Quality assurance of the German Environmental Data Catalogue (Ministry of Environment of Lower Saxony, Germany)

ii) Tendering and contracts (Province of Genoa, Italy)

iii) Documents Review of the European Environment Information and Observation Network (National Environment Research Institute, Denmark)

iv) Career Offices Network (University of the Aegean, Greece)

In order to design the MERMIG environment the requirements of public administration for electronic support of G2G collaboration were analyzed (WP2). For this purpose initially were constructed and analyzed the 40 use cases of the above four pilots, and based of them were analyzed the corresponding requirements; also, were analyzed, in less detail, the relevant requirements of 150 collaboration processes from various public administrations of European Union member states, and the relevant requirements of the 'Interchange of Data between Administrations' (IDA) Program of the European Union. Based on these requirements was designed the MERMIG environment and were determined its necessary modules and their basic functionality. These modules are: **(a)** workflow management system; **(b)** document repository; **(c)** content management system; **(d)** groupware (including a structured argumentation tool); **(e)** e-mail client/server application (with members directory); **(f)** web form design and validation tool; **(g)** advanced profiling mechanism; **(h)** interface module; and **(i)** intelligent agent module (with search, e-mail, and meeting/planning/scheduling capabilities) **(j)** modeling tool. More information on the ICTE-PAN project and the architecture and modules of MERMIG are given in other publications ([18] – [22]).

4. Evaluation Methodology

In accordance with the conclusions of the literature review presented in section 2 it was decided to evaluate the G2G collaboration support environment MERMIG in two stages: initially in a first stage to perform an "efficiency-oriented" evaluation based mainly on the user requirements defined in WP2, and then in a second stage to perform an "effectiveness-oriented" evaluation for assessing to what extent MERMIG supports the execution of typical collaborative tasks in public administration. In the present paper is analyzed the first stage of "efficiency-oriented" formative and summative evaluation, which has been performed as part of WP7. In order to select an appropriate evaluation methodology for this purpose initially were defined the selection criteria and also a number of critical issues that should be taken into account. The selection criteria were:

I. Range of activities: the evaluation methodology should include the appropriate combination of activities in order to produce a useful mix of qualitative and quantitative data, and record views, impressions and assessments from various members of the project partner institutions; also in order to use effectively the project resources, the evaluation methodology should consist of activities that complement each other, rather than activities that provide the same type of information.

II. Relevance of activities: the evaluation methodology should include activities which are directly relevant to the project and have the potential to provide usable outcomes

III. Feasibility: another constraint on evaluation activities is the available resources allocated to this task of the project and the skills of the users; therefore the evaluation methodology should not include activities that require excessive resources or quite high users' skills, which are not available; this criterion is very important since in public administration usually there are limited resources (e.g. financial, human, etc.), especially for IS evaluation purposes.

Also for the selection of the evaluation methodology the following issues should be taken into account:
- if we define as evaluation stakeholders the members of the Project Coordinating Board (PCB), which is composed by representatives from all the project partners, both technology providers and public organizations-users, the evaluation methodology should enable and facilitate taking into account all these different perspectives and concerns,
- in order to achieve consistent and comparable evaluation results across all partners, evaluation activities should be structured, clear and follow a pre-defiend format,
- it is important to be aware of the differences between the partner institutions in terms of administrative environment, subject content, procedures, resources, etc.,

Using the above selection criteria and taking into account the above issues it was finally selected to use the international standard ISO/IEC 14598 (Software Product Evaluation), which provides guidance and requirements for the software evaluation process ([23] – [28]), in combination with the ISO/IEC 9126 (Software Product Quality), which provides software quality characteristics and metrics ([29] – [31]), as basis of our evaluation methodology. This series of standards has numerous advantages against other known evaluation methodologies, since it defines a generic enough evaluation framework, it uses a wide range of software quality characteristics covering a variety of aspects (and not only the functional ones), and it is adaptable to all software development technologies. Furthermore, taking into account that the potential users (both in the private and in the public sector) have more confidence in products and services conforming to international standards, this selection can significantly contribute to the wide acceptance of the MERMIG environment. Also, since the development partners of ICTE–PAN project, European Dynamics SA and TXT e–Solutions SpA, are both ISO 9001 certified, it was very important the software product of the project to be as well in line with the ISO guidelines.

Initially, based on the ISO/IEC 14598 standard, was designed the evaluation process, which is shown in Figure 1.

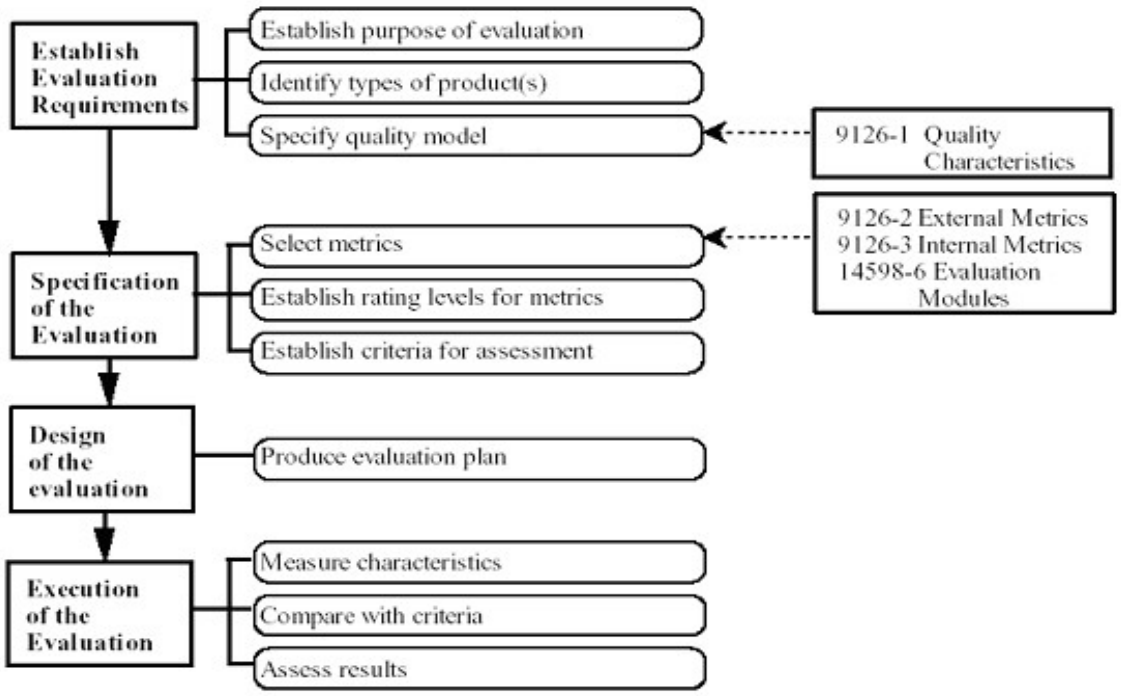

Figure 1: Evaluation Process

Also 135 test scenarios were designed in total:

- 46 pilots test scenarios, which were based on the 40 use cases of the pilots: for each use case were designed one or more test scenarios, aiming at testing whether the corresponding functionality has been correctly implemented,

- and 89 modules test scenarios: for each module were designed several test scenarios, which corresponded to typical "use cases" of the module (i.e. typical tasks to be executed using it), aiming at testing whether the corresponding functionality has been correctly implemented.

Next, based on the ISO/IEC 9126 Quality Model, which is shown in Figure 2, was designed the specific quality model of our evaluation. As we can see in this Figure the ISO/IEC 9126 defines six basic quality characteristics (Functionality, Reliability, Usability, Efficiency, Maintainability and Portability) and their corresponding sub–characteristics. However, the requirements expressed by the user-partners in WP2, and also the ones collected from other sources, such as IDA, etc., are classified within the first three basic quality characteristics of the ISO/IEC 9126 Quality Model (Functionality, Reliability, Usability); therefore for the evaluation of MERMIG we had to focus on these three characteristics and their corresponding sub-characteristics. For each sub-characteristic of these first three quality characteristics, external and internal metrics were defined, the external ones to be used by the actual pilot users from the various participating public administrations, and the internal ones to be used by the developers. For example, for the sub-characteristic Suitability of the characteristic Functionality, we defined the important metric "functional implementation rate", whose calculation formula is: X = A/B, where A = number of functions implemented in the development, and B = number of functions described in specifications (i.e. use cases). Finally the quality model of our evaluation included 3 quality characteristics, 14 sub-characteristics and 39 metrics. Also, for each module the 14 sub-characteristics were prioritised, in accordance with the directions of the ISO/IEC 14598 standard, in a three levels scale (high – medium – low priority), based on the relevant conclusions of the users' requirements analysis (WP2).

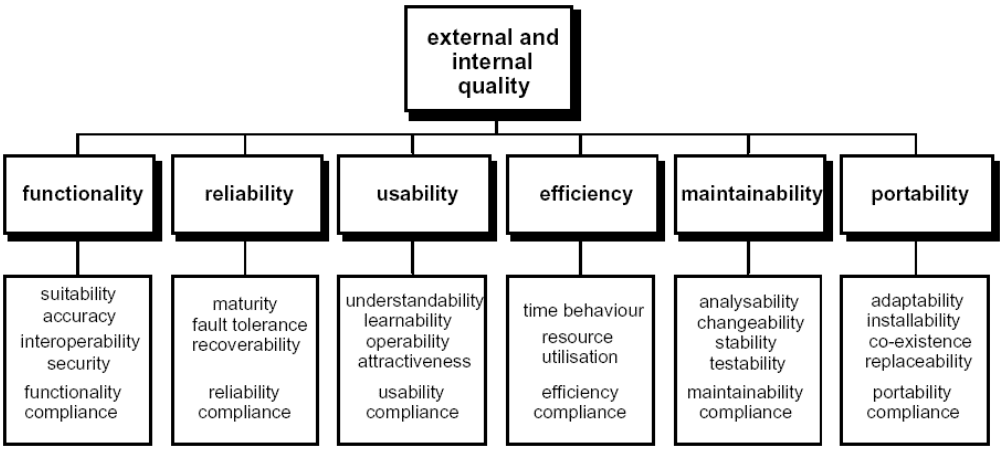

Figure 2: ISO/IEC 9126 Quality Model

The evaluation process included both formative and summative evaluation. The formative evaluation was performed at several points of the lifecycle of the project both by technical experts and users; its basic objective was to identify problems/malfunctions and to suggest improvements of the MERMIG environment during its development, and also to assist various design decisions. The summative evaluation was performed mainly by the users

during the operation of the pilots, when the basic functionality of the environment had been developed; its basic objective was to assess the quality of the pilots and the modules. As a result of the iterative development approach adopted in this project, it was not always easy to draw a clear line between summative and formative evaluation. Also, while the summative evaluation was performed in a high structured way, according to the initial evaluation plan, the formative evaluation was performed in a more flexible way.

In order to have a complete and reliable assessment of the MERMIG environment a "dual" evaluation was performed: i) evaluation of the modules and ii) evaluation of the pilots (each pilot used most of the modules and also required some more development and customization), based on the corresponding test scenarios (as described previously in this section). The evaluation of each pilot consisted of the following steps:

i) each user partner initially performed the test scenarios of their pilot; for each of these tests were registered in detail in an electronic form the results and the comments,

ii) based on these tests, for each pilot initially was calculated the "functional implementation rate" (according to the definition we gave previously in this section), which constitutes a highly objective assessment,

iii) then followed two more subjective assessments: first, for each test scenario were assessed all the 14 quality sub-characteristics; for this purpose "evaluation matrices" were designed, which had the quality sub-characteristics along the horizontal dimension and the test scenarios along the vertical dimension, and then filled by the user-partners,

iv) and second, for each pilot was made a total assessment (based on all its test scenarios) of all the 39 quality metrics; for this purpose a questionnaire was designed, which included 39 questions (one for each quality metric), and then filled by the user-partners,

v) finally, for each pilot a consolidated calculation of the 14 quality sub-characteristics was made, based on both the assessments of step (iii) in the evaluation matrices and the assessments of step (iv) in the questionnaire.

The same steps were followed as well for the evaluation of each module of the MERMIG environment: initially the test scenarios of the module were performed, then the functional implementation rate was calculated for the module, followed by the assessment of the 14 quality sub-characteristics for each of the test scenarios of the module using the evaluation matrices, the total assessment the 39 quality metrics (based on all its test scenarios) using the questionnaire, and finally the consolidated calculation of the 14 quality sub-characteristics for the module (using the data from both the evaluation matrices and the questionnaires).

5. Evaluation Results

Using the evaluation methodology described in the previous section a large quantity of data was accumulated and from processing them a large amount of results were produced. In this section are presented and discussed some typical results. Concerning the evaluation of the pilots two different approaches were followed: for each metric, sub-characteristic and characteristic the results of the different pilots were compared, and then aggregated in order to calculate the corresponding average value. In Figure 3 we can see the functional implementation rate of each pilot with respect to its use cases, and the overall functional implementation rate over all pilots. We remark that there is a high variance among the four

pilots concerning the implementation of functionalities, which varies from 91,4% for the Province of Genoa (PROGE) pilot, to 84% for the AEGEAN pilot and 75% for the Ministry of Environment of Lower Saxony (MELS) pilot, down to 56,4% for the National Environment Research Institute (NERI) pilot, resulting to an overall implementation of 76,8%. The main reason of the low implementation rate in some of the pilots (especially in the NERI pilot) is that the evaluation was performed some months before the end of the project (according to its initial schedule), while some development and customisation work was still in progress. These implementation rate values (which can also be "drilled down") show us where more effort is required and more resources should be allocated.

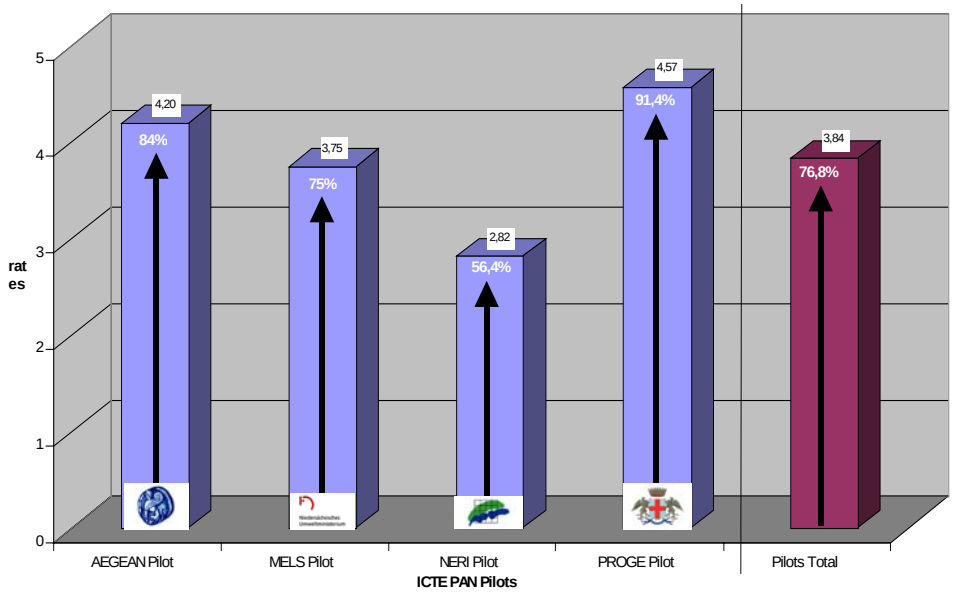

Figure 3: Pilots functional implementation rates

In Figure 4 we can see for each pilot the average quality assessment over all the 39 metrics of the quality model, based on the questionnaires filled by the user partners.

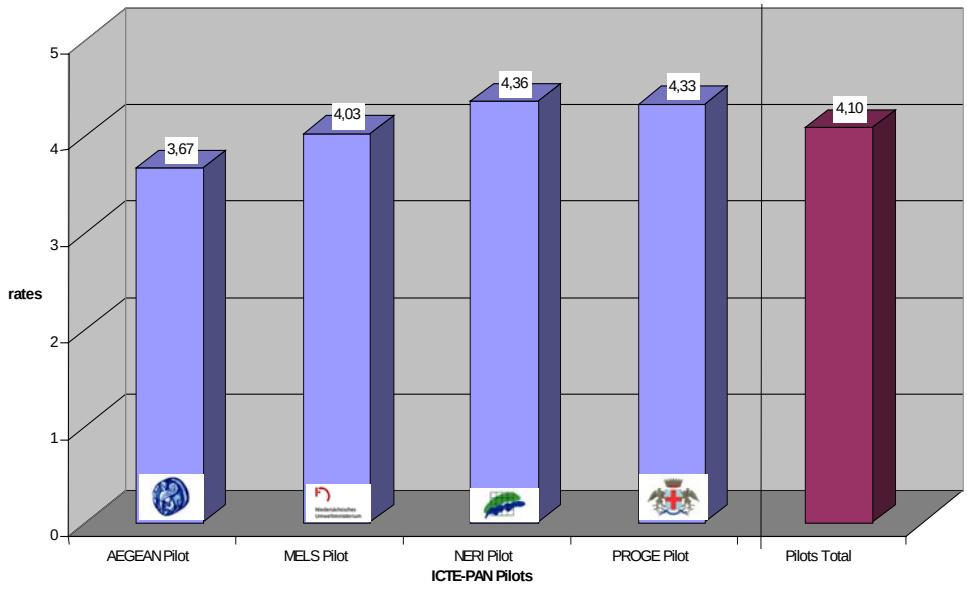

Figure 4: Pilots quality assessment based on all metrics

From Figure 4 we remark that the variance in total quality assessment among the pilots is lower in comparison with the variance in functional implementation rate shown in Figure 3. This difference in the variance among the pilots, which is quite pronounced for the pilot of NERI, is attributed to the differences between these two evaluation approaches. Missing functionalities have a highly negative impact on the functional implementation rate. However, the quality model metrics approach takes into account missing functionalities in only 3 out of the 39 metrics; in all the other metrics missing functionalities are not taken into account, since we assess them based only on the existing functionalities, e.g. we do not assess usability for a function that has not been implemented. Therefore it is concluded that missing functionalities have a very low impact on the total quality assessment. For this reason it is recommended to use simultaneously both these evaluation approaches in order to have a complete picture.

In Figure 5 we can see in a net-graph for all the characteristics and sub-characteristics of the quality model the average assessment over all pilots, based on the consolidation of the assessments from the evaluation matrices (taking into account all the test scenarios of each pilot) and the questionnaires (one total assessment per pilot). We remark that for all sub-characteristics the assessments are above 3,5 and for all the characteristics even above 4 (for functionality 4,02, reliability 4,14 and usability 4,24). The net-graph is almost "round", with small "indentations" for Fault Tolerance (3,54), Attractiveness (3,57) and Security (3,68), which have slightly lower assessments than the other sub-characteristics and require improvements. In general this net-graph illustrates a satisfactory overall quality assessment, and at the same time shows areas for improvement.

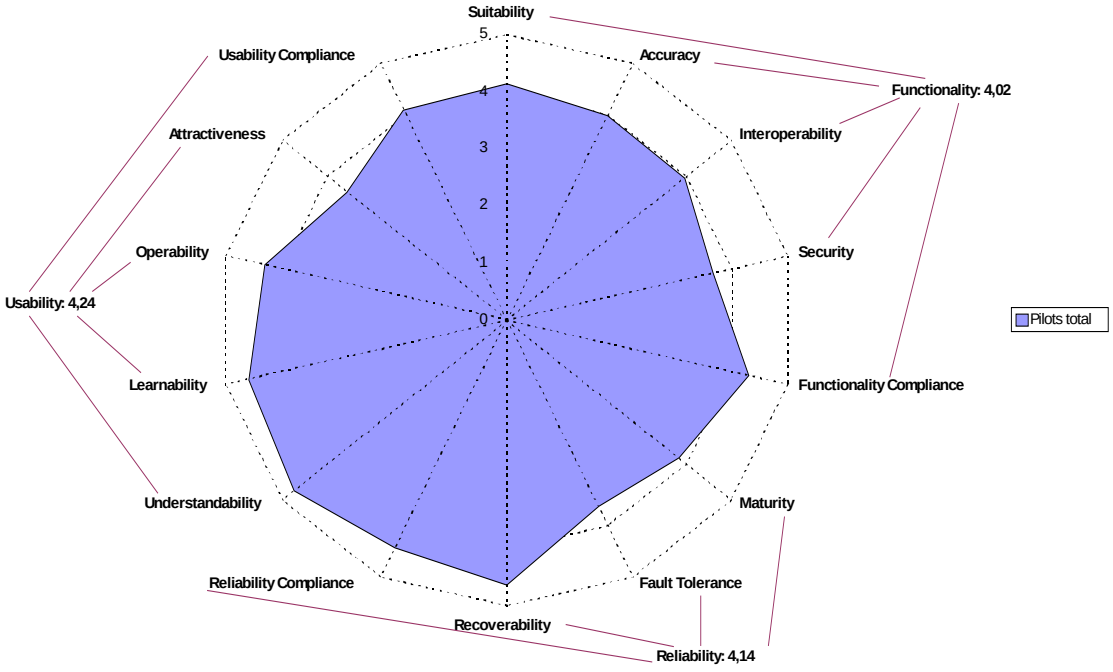

Figure 5: Assessment of all the characteristics and sub-characteristics of the quality model over all pilots

The results of the evaluation of the pilots, with the exception of the NERI pilot, were satisfactory, especially if we take into account that the evaluation was performed some months before the finalisation of the developments; until the end of the project all four pilots were completed and most of the weaknesses diagnosed in the evaluation were addressed. In all pilots we had high assessments for most characteristics and sub-characteristics (higher than

3,5), and only in some pilots we had lower assessments for some sub-characteristics, such as the maturity and the fault tolerance in the PROGE pilot, the attractiveness in the AEGEAN and PROGE pilots, the security in the AEGEAN, PROGE and MELS pilots, the suitability in the NERI pilot and the interoperability in the AEGEAN pilot.

Similarly satisfactory were the results from the evaluation of most of the modules of the MERMIG environment; however for some modules weaknesses were diagnosed. For one of them, the modelling tool, in Figure 6 we can see in a net-graph the assessment of all the characteristics and sub-characteristics of the quality model. We remark that the assessments are low concerning the usability sub-characteristics (e.g. 2,9 for understandability, 2,5 for learnability and 2,8 for operability) and some functionality sub-characteristics (e.g. 3,5 for usability and 3,25 for accuracy). It is therefore concluded that the users had serious problems in understanding and using the modelling tool, so improvements are required.

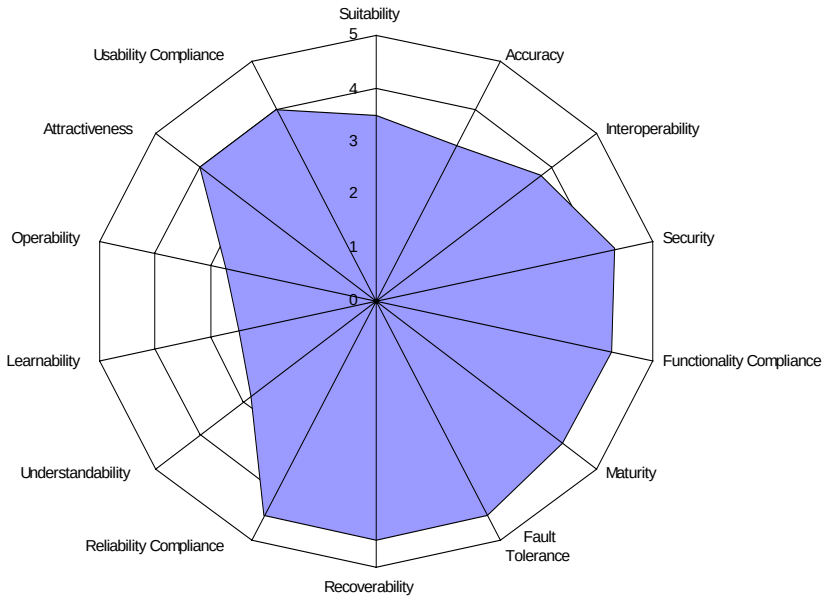

Figure 6: Assessment of all characteristics and sub-characteristics of the quality model for the modelling tool

In general based on the findings of the whole evaluation process, it can be concluded that the basic ICTE-PAN project outcome is a valuable G2G collaboration environment with high reliability and rich functionality, which has proven to be appreciated by the end users.

6. Conclusions

In this paper we have used two generic international standards, ISO/IEC 14589 and ISO/IEC 9126, in order to create and apply a multi-level, multi-dimensional and multi-view evaluation methodology for a specific context: for a large IST project, implemented by a consortium of technology providers and public organizations-users, aiming at the development of a complex G2G collaboration support environment. Although the proposed evaluation methodology is quite demanding for the users (they have to fill lengthy "evaluation matrices" concerning numerous test scenarios which are evaluated in many characteristics and sub-characteristics,

and also lengthy questionnaires with numerous questions – one or more questions for each metric), it offers significant advantages:

- it is a multi-level approach, including various levels of evaluation throughout the lifecycle of the project, putting emphasis both on the formative and the summative view

- it is a multi-dimensional approach, as it is based on a wide range of quality characteristics and sub-characteristics,

- it is a multi-view approach, as its employs a combination and consolidation of mutually complementary subjective and objective assessments,

- it generates a large amount of information, which allows a big and heterogeneous development consortium to diagnose problems and weaknesses and finally produce a high quality software,

- it is based on the combination of pilots evaluation and modules evaluation ('dual' evaluation), in order to give a more complete and reliable picture

- it is based on ISO/IEC international standards, adapted to and enhanced for the needs of the specific context (of the ICTE-PAN projects and the European Public Administration)

However, it is important to note that the evaluation results are heavily dependent on the quality of the use cases, the evaluation matrices and questionnaires. Also the use cases approach (which forms the basis for formulating the test scenarios) suffers from the fact that use cases completely leave out the infrastructure needed to support this software (e.g. server reboot problems will not be diagnosed through this approach). In addition, due to the nature of the use cases they are formulated at the beginning of a project and thus do not take into account the iterative aspects of the project development. Further research is in progress towards an "effectiveness" evaluation of the same collaboration support environment.

References

[1] McFarlan F. W., Portfolio approach to information systems. Harvard Business Review, Vol. 59, 1981, pp.142-150.

[2] Standish Group, The CHAOS Report, 1995, available at www.standishgroup.com

[3] Poulymenakou A., Holmes A, A contingency framework for the investigation of information systems failure, European Journal of Information Systems, Vol. 5, 1996, pp. 34-46.

[4] Cabinet Office, Review of Major Government IT Projects – Successful IT: Modernising Government in Action, 2000, available at http://www.ogc.gov.uk

[5] Athens Technology Center, Annual Evaluation Report of the Integrated Mediterranean Programs on Information Technology, 1994, Athens.

[6] Ministry of National Economy, Final Report of Integrated Mediterranean Programs on Information Technology, 1994, Athens.

[7] Ministry of National Economy, Operational Programme 'Information Society'- European Union Support Framework III, 2001, Athens.

[8] Lyytinen K., Robey D., Learning failure in information systems development, Information Systems Journal, Vol. 9, 1999, pp. 85-101.

[9] Smithson S., Hirschheim R., Analysing information systems evaluation: another look at an old problem, European Journal of Information Systems, Vol. 7, 1998, pp. 158-174.

[10] Farbey B., Land D., Targett D., IS evaluation: a process of bringing together benefits, costs and risks, in: W. Currie, R. Galliers (Eds.), Rethinking Management Information Systems, Oxford University Press, New York, 1999, pp. 204-228.

[11] Land F., Evaluation of systems goals in determining a design strategy for a computer-based information system, Computer Journal, Vol. 19(4), 1976, pp. 290-294.

[12] Hirschheim R., Smithson S., A critical analysis of information systems evaluation, in: N. Bjorn-Andersen, G. Davis (Eds), Information Systems Assessment: Issues and Challenges, North Holland, Amsterdam, 1988, pp. 17-37.

[13] Farbey B., Land F., Targett C., A taxonomy of information systems applications: the benefits' evaluation ladder, European Journal of Information Systems, Vol. 4, 1995, pp. 41-50.

[14] Irani Z., Love P., Information systems evaluation: past, present and future, Editorial, European Journal of Information Systems, Vol. 10, 2001, pp. 183-188.

[15] Irani Z., Information systems evaluation: navigating through the problem domain, Information & Management, Vol. 40, 2002, pp. 11-24.

[16] Love P., Irani Z., Edwards D., Researching the investment of information technology in construction: An examination of evaluation practices, Automation in Construction, Vol. 14, 2005, pp. 569-582.

[17] Pitt L. F., Watson R. T., Kaven C. B., Service quality: A measure of information systems effectiveness, MIS Quarterly, Vol. 19(2), 1995, pp. 173-187.

[18] Loukis E., Kokolakis S., Computer supported collaboration in the Public Sector: the ICTE-PAN Project, in: Proceedings of the EGOV-2003 International Conference: From E-Government to E-Governance, Prague, Czech Republic, September 1-5, 2003, Springer Lecture Notes in Computer Science, Vol. 2739, pp. 181-186.

[19] Loukis E., Kokolakis S., An Architecture for a Flexible Public Sector Collaborative Environment based on Business Process Modelling, Electronic Journal of e-Commerce Tools and Applications, Vol. 1(3), 2004.

[20] Karacapilidis N., Loukis E., Dimopoulos St., A Web-based System for Supporting Structured Collaboration in the Public Sector, in: Proceedings of the EGOV-2004 International Conference: Electronic Government, Zaragoza, Spain, August 30 – September 3, 2004, pp. 218-225.

[21] Karacapilidis N., Loukis E., Dimopoulos, S., Computer-supported G2G collaboration for public policy and decision making, Journal of Enterprise Information Management, Special Issue on e-Government (accepted for publication).

[22] Web-site of the ICTE-PAN project: http://www.eurodyn.com/icte-pan/

[23] International Organization for Standardization (ISO), Information technology – software product evaluation – part 1: general overview, Draft International Standard 14598–1, October 1996.

[24] International Organization for Standardization (ISO) Information technology – software product evaluation – part 2: planning and management, Committee Draft 14598–2, August 1995.

[25] International Organization for Standardization (ISO), Information technology – software product evaluation – part 3: process for developers, Committee Draft 14598–3, August 1996.

[26] International Organization for Standardization (ISO), Information technology – software product evaluation – part 4: process for acquirers, Committee Draft 14598–4, October 1996.

[27] International Organization for Standardization (ISO), Information technology – software product evaluation – part 5: process for evaluators, Draft International Standard 14598–5, July 1996.

[28] International Organization for Standardization (ISO), Information technology – software product evaluation – part 6: evaluation modules, Committee Draft 14598–6, August 1995.

[29] International Organization for Standardization (ISO), Software engineering – product quality – Part 1: quality model, First Draft International Standard 9126–1, November 1999.

[30] International Organization for Standardization (ISO), Software engineering – product quality – Part 2: external metrics, Draft Technical Report 9126–2, January 2001.

[31] International Organization for Standardization (ISO), Software engineering – product quality – Part 3: internal metrics," Preliminary Draft Technical Report 9126–3, January 2000.