A Hierarchical Evaluation Scheme for Pilot-based Research Projects

Thomas Zefferer
Secure Information Technology Center Austria (A-SIT), Inffeldgasse 16a, 8010 Graz, Austria

Keywords: Evaluation Scheme, Project Evaluation, Hierarchical Evaluation Model.

Abstract: Evaluation is an integral part of most research projects in the information-technology domain. This especially applies to pilot-based projects that develop solutions for the public sector. There, responsible stakeholders require profound evaluation results of executed projects to steer future research activities in the right directions. In practice, most projects apply their own project-specific evaluation schemes. This yields evaluation results that are difficult to compare between projects. Consequently, lessons learned from conducted evaluation processes cannot be aggregated to a coherent holistic picture and the overall gain of executed research projects remains limited. To address this issue, we propose a common evaluation scheme for arbitrary pilot-based research projects targeting the public sector. By relying on a hierarchical approach, the proposed evaluation scheme enables in-depth evaluations of research projects and their pilots, and assures at the same time that evaluation results remain comparable. Application of the proposed evaluation scheme in the scope of an international research project confirms its practical applicability and demonstrates its advantages for all stakeholders involved in the project.

1 INTRODUCTION

Leveraging the use of information technology (IT) is on top of the agenda of many public-sector institutions. In Europe, besides various national attempts to push the use of IT in public-sector use cases, the European Union (EU) invests considerable financial resources in bringing forward its digital agenda on pan-European level (European Commission, 2018a). One approach followed by the EU to achieve these goals is the funding of international research projects that develop innovative IT solutions for the public sector. The Large Scale Pilots (LSPs) STORK 2.0[1], PEPPOL[2], or e-SENS[3] are just a few examples of recent research activities funded by the EU.

Most research activities follow a pilot-based approach, i.e. they develop new IT solutions and integrate them into pilot applications. These pilot applications serve a concrete use case and thus evaluate the applicability and the usefulness of the developed solution in practice. For instance, the LSPs STORK and STORK 2.0 developed an interoperability layer for national electronic identity (eID) solutions (Leitold and Posch, 2012). Several pilot applications relying on this interoperability layer enabled EU citizens to use their national eID to authenticate at electronic services provided by other EU member states. Details of these pilots have been discussed by (Knall et al., 2011) and (Tauber et al., 2011).

Obviously, the piloting phase is an integral part of pilot-based projects. However, it is usually restricted to the project's lifetime. As a consequence, even successful pilot applications are typically terminated at the end of a research project. Turning the pilot application into a productive service is usually out of the project's scope. It is thus essential that all stakeholders involved in the project, i.e. researchers, funding bodies, and end users, derive as many lessons learned from the piloting phase as possible. After completion of the project, these lessons learned are crucial to turn a promising pilot into a successful productive service.

Accordingly, a sound evaluation of pilot applications is crucial. Unfortunately, the current situation is often unsatisfying. It can be observed that pilot evaluations are often heterogeneous with regard to approaches followed, methods applied, and hence also results obtained. This heterogeneity can be observed within projects, i.e. between pilot applications, and also between projects. Ultimately, this leads to a situation, in which obtained evaluation results are hardly

[1] https://www.eid-stork2.eu/
[2] https://peppol.eu/
[3] https://www.esens.eu/

Zefferer, T.
A Hierarchical Evaluation Scheme for Pilot-based Research Projects.
DOI: 10.5220/0007228103560363
In *Proceedings of the 14th International Conference on Web Information Systems and Technologies (WEBIST 2018)*, pages 356-363
ISBN: 978-989-758-324-7

comparable. Consequently, lessons learned from conducted pilot evaluations often cannot be aggregated to a coherent holistic picture, which in turn makes it difficult for responsible stakeholders to draw the correct conclusions from obtained evaluation results.

To address this issue, we propose a common evaluation scheme for pilot-based research projects targeting the public sector. The proposed evaluation scheme is project and pilot independent and can hence be applied to a broad range of research projects. The scheme relies on a hierarchical evaluation-criteria model and defines a common procedure to apply criteria based on this model. By providing a common basis for pilot evaluation, the proposed scheme ensures that obtained evaluation results are homogeneous enough to enable comparisons between different pilots within a project as well as between different projects. In this paper, we introduce the proposed evaluation scheme in detail and evaluate it by means of a concrete research project.

2 RELATED WORK

During the past years, the EU has funded a series of research project to improve public sector related IT services. To support its strategy of a Digital Single Market (European Commission, 2018c), many of these projects have focused on achieving cross-border interoperability between IT services of different EU Member States. Examples are the Large Scale Pilots STORK and STORK 2.0[4], epSOS[5], or PEPPOL[6]. Results of these LSPs have been consolidated by the research project e-SENS[7]. Leveraging the use of IT in public-sector use cases has also been the main goal of the international research project SUNFISH[8], funded under the EUs Horizon 2020 research and innovation programme, and of the project FutureID[9] funded under the ICT theme of the Cooperation Programme of the 7th Framework Programme of the European Commission. Scientific contributions of these projects have been discussed by (Suzic and Reiter, 2016) and (Rath et al., 2015), respectively.

Having a more detailed look at all these projects' internal structure reveals that all of them follow a similar approach: developed solutions are tested by means of pilot applications. Furthermore, all projects contain some sort of evaluation, where obtained results are assessed against defined criteria. However, the approaches followed by the various projects to carry out evaluations differ considerably from each other. In the worst cases, the same evaluation method is not even applied within a project consistently, e.g. to evaluate different pilot applications of a project. This heterogeneity in applied evaluation methods has an impact on obtained evaluation results. While these results might be sufficient within the scope of a single pilot application, they cannot be assembled to a coherent big picture. This, in turn, renders the conclusive derivation of findings from available evaluation results difficult.

In literature, interesting works exist that focus on the evaluation of research projects. For instance, (Khan et al., 2013) discuss the problem of evaluating a collaborative IT-based research and development project. While this work identifies relevant challenges to overcome, the proposed evaluation method has been tailored to one specific project, rendering its application to arbitrary projects difficult. More generic evaluation methods have been introduced by (Eilat et al., 2008) and by (Asosheh et al., 2010), which both make use of the balanced scorecard (BSC) approach and data envelopment analysis (DEA). However, these proposals do not take into account the specifics of the type of projects targeted in this paper, i.e. pilot-based projects settled in the public sector.

In summary, evaluation schemes proposed in literature usually focus on very specific types of projects or are even tailored to one single project. An evaluation scheme that can be applied to a broad range of pilot-based research projects from the public sector is currently missing. The evaluation scheme proposed in this paper closes this gap.

3 PROJECT MODEL

The challenge in developing an evaluation scheme for a broad set of research projects lies in the trade-off between assuring general applicability and obtaining meaningful evaluation results. On the one hand, an abstract scheme enables a broad applicability to arbitrary research projects. However, obtained evaluation results remain abstract too and often do not yield concrete conclusions. On the other hand, a more specific evaluation scheme can consider peculiarities of a given research project or pilot application. However, such a specific scheme can usually only be applied to a limited set of projects or pilots.

To overcome this challenge, we have based the proposed evaluation scheme on a well-defined project model. The proposed scheme can be applied to

[4]https://www.eid-stork2.eu/
[5]http://www.epsos.eu/
[6]https://peppol.eu/
[7]https://www.esens.eu/
[8]http://www.sunfishproject.eu/
[9]http://www.futureid.eu/

any research project complying with this model. The project model has been based on three basic assumptions, which define the scope of targeted project types. First, the proposed evaluation scheme targets pilots and projects that provide solutions for the public sector. Hence, the evaluation scheme assumes that an IT agenda is in place, from which project goals are derived. Second, the proposed scheme targets pilot-based projects, which test their results by means of one or more pilot applications. Third, the proposed scheme assumes the research project to be carried out in multiple consecutive project phases, each comprising its own pilot evaluation.

From these assumptions, the general project structure shown in Figure 1 is derived. By intention, this structure has been kept rather abstract, in order to assure a broad applicability to concrete research projects. In brief, the shown project structure complies with all projects that are driven by a given IT agenda and that foresee development and operation of one or more pilot applications.

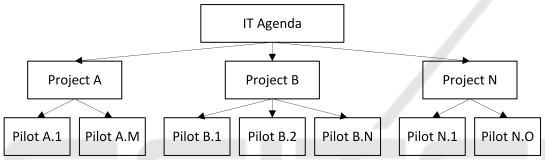

Figure 1: The proposed evaluation scheme can be applied to all projects complying with the shown general structure.

In addition to this project structure, general project-execution phases and related evaluation phases can also be derived from the assumptions made. This is shown in Figure 2. Again, also project phases and related evaluation phases have been defined on a rather abstract level to ensure applicability to a broad range of projects.

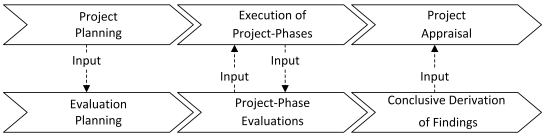

Figure 2: The proposed evaluation scheme can be applied to all projects implementing the shown general project-execution steps.

During *Project Planning / Evaluation Planning*, the project's overall structure, contents, goals, and setup are defined and the project's evaluation is planned. Accordingly, relevant input from project planning needs to be considered during evaluation planning. The *Project Planning / Evaluation Planning* phase is followed by the *Execution of Project Phases / Project-Phase Evaluations* phase. According to the assumptions made above, the project is executed in consecutive project phases. Figure 2 shows that each project phase is accompanied by a corresponding evaluation phase. Corresponding project-execution phases and evaluation phases influence each other. While conducted evaluations depend on the respective project-execution phase and its goals and contents, project phases should take into account available evaluation results (e.g. from previous phases) to continuously improve the project. Finally, *Project Appraisal / Conclusive Derivation of Findings* constitutes the third and final phase as shown in Figure 2. It is done at the end of the project to collect all lessons learned, draw the correct conclusions from these lessons, and to bring the project to a round figure. In the corresponding evaluation phase, findings are derived from conducted evaluations and serve as direct input to project appraisal.

Together, the project structure (Figure 1) and the general project-execution and evaluation phases (Figure 2) define the general project model used as basis for the proposed evaluation scheme introduced in the next section. This scheme can be applied to any pilot-based project that complies with this general project model. Given the deliberately abstract nature of the model, this should be the case for the majority of pilot-based projects targeting the public sector. Thus, the project model reasonably handles the trade-off between assuring general applicability and obtaining meaningful evaluation results.

4 PROPOSED EVALUATION SCHEME

Based on the defined project model, we propose a generic evaluation scheme for the systematic evaluation of pilot-based public-sector research projects. The proposed evaluation scheme aims for two goals. On the one hand, the scheme aims to be sufficiently abstract to be applicable to a broad range of pilots and projects, in order to ensure comparability within projects (i.e. between the project's pilots) and also between different projects. On the other hand, the evaluation scheme should still enable in-depth evaluations that take into account specifics of pilots and projects. The proposed scheme deals with this obvious trade-off by following a hierarchical approach.

Details of the proposed evaluation scheme are introduced in this section. Subsection 4.1 focuses on the scheme's evaluation-criteria model, which provides a framework for the definition of concrete evaluation criteria. Subsequently, Subsection 4.2 introduces the proposed scheme's evaluation process, which specifies a step-by-step procedure to evaluate pilots using defined evaluation criteria.

4.1 Evaluation-criteria Model

The evaluation-criteria model is the first relevant building block of the proposed evaluation scheme. In principle, evaluation criteria used for pilot evaluation need to meet the same requirements as the overall evaluation scheme. Concretely, evaluation criteria should ideally be the same for all pilots in all evaluated projects. Only in this case, direct comparisons between different pilots and even between different projects are feasible. At the same time, evaluation criteria should be concrete enough to consider specifics of pilots. Obviously, these are contradictory requirements, which cannot be met by a simple list of evaluation criteria. Therefore, the proposed evaluation scheme relies on a hierarchical evaluation-criteria model that defines multiple layers of evaluation criteria as shown in Figure 3.

In general, evaluation criteria are closely related to project and pilot goals. Concretely, evaluation criteria are used to assess a pilot's or project's compliance with defined goals. This close relation is also reflected by the evaluation-criteria model depicted in Figure 3. The left pyramid shows the different layers, on which relevant goals can be defined. Note that the pyramid's structure complies with the general project model defined in Section 3. The pyramid's topmost level represents the relevant IT agenda defining the very basic goals to consider. This agenda yields project-specific goals for concrete projects executed under the given agenda. Within a concrete project, pilot-specific goals can be derived for each of the project's pilots from the overall project goals. Finally, pilot-specific goals can be further detailed by defining pilot-specific goals separately for each project phase. Overall, the proposed evaluation-criteria model defines relevant goals to be defined on four different layers of abstraction.

Once all goals are defined according to the four layers, relevant evaluation criteria can be derived. The proposed model foresees evaluation criteria to be defined on four layers as well, yielding the right pyramid shown in Figure 3. When deriving evaluation criteria for the four layers, two aspects need to be considered. First, defined evaluation criteria must cover relevant goals defined before. This applies to all layers and is indicated in Figure 3 by horizontal arrows. Second, evaluation criteria defined in neighboring layers must be related. In particular, given the fulfillment degree of criteria in a layer, it must be possible to derive the fulfillment degree of criteria in the superior layer.

Note that the proposed evaluation-criteria model deliberately does not define concrete evaluation criteria, in order to ensure a broad applicability of the proposed evaluation model. Instead, the proposed evaluation-criteria model merely defines a framework for the definition of relevant goals and related evaluation criteria. This framework enforces a systematic derivation process for relevant goals and evaluation criteria on different layers of abstraction, ensures adequate relations between goals and criteria defined on different layers, and assures a precise mapping between goals and evaluation criteria. This enables a systematic evaluation process, where the fulfillment of higher-level criteria can be derived automatically from the fulfillment of lower-level requirements.

4.2 Evaluation Process

Once relevant goals have been defined and evaluation criteria have been derived, the actual evaluation process can be carried out. The evaluation process, which constitutes the second building block of the proposed evaluation scheme, is illustrated in Figure 4. In principle, the shown process can be regarded as a detailing of the general project and evaluation phases as introduced in Section 3. Thus, the proposed evaluation process implicitly complies with the defined general project model. Figure 4 illustrating the evaluation process is subdivided into six areas. First, the entire figure is split into two halves. The left half describes process steps to be carried out during project execution (corresponding to the upper part of Figure 2). The right half describes necessary steps to be carried out during evaluation (corresponding to the lower part of Figure 2). Second, the two halves of the flow chart are further split into three horizontal areas, corresponding to the three general execution phases defined in Section 3.

As shown in Figure 4, the proposed evaluation scheme comprises 18 steps to be carried out in total. In Step (1), which is the first step in project planning, general goals from relevant IT agendas, under which the project is executed, are identified first. These are the most high-level goals to be considered during project execution. In the end, the success of the project is assessed against these goals. From these general goals, concrete project goals are defined next by Step (2). Project goals detail higher-level goals by applying the project's specific context. Hence, the project's context and its defined contents are a relevant input to this step. Once the general project goals have been fixed, pilot-specific goals can be derived in Step (3). Pilot-specific goals need to comply with the higher-level project goals, but additionally take into account the specifics of the project's pilot applications. Accordingly, the definition of pilots and their foreseen role in the project are relevant inputs to this processing step. Derived pilot-specific goals can

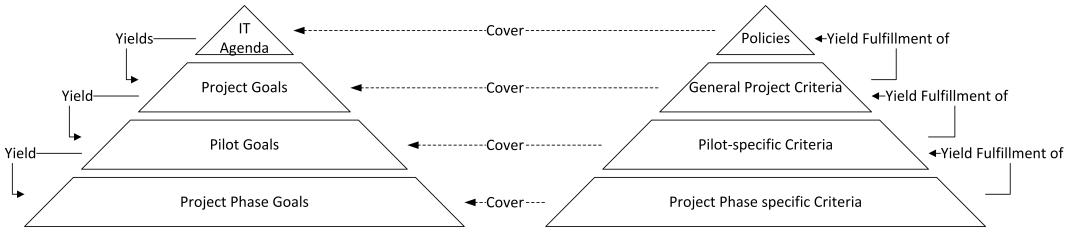

Figure 3: The proposed evaluation-criteria model enables the definition of goals and related criteria on multiple layers.

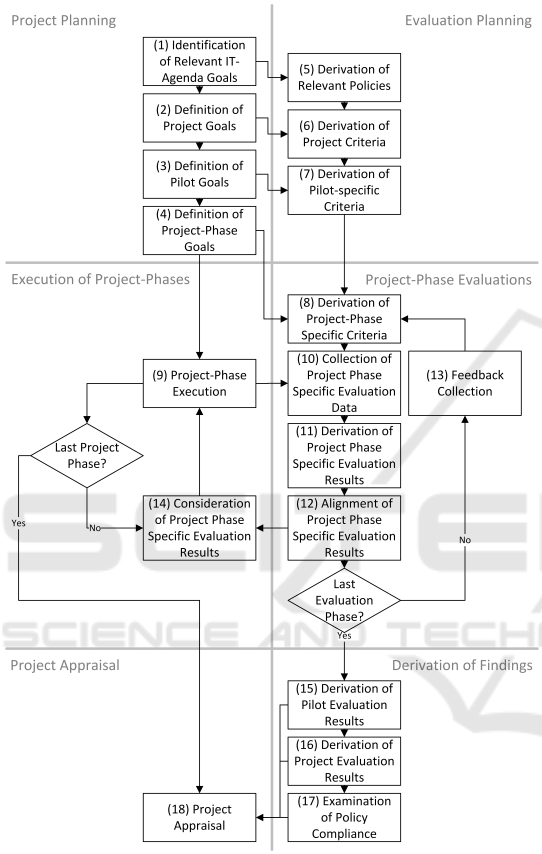

Figure 4: The evaluation scheme specifies process steps to be carried out in the shown order.

again vary between different project phases. To consider these phase-specific variations, pilot-specific goals are further detailed to project phase specific goals in Step (4).

Step (5), i.e. derivation of relevant policies, is the very first step to be taken in evaluation planning. Relevant policies to be considered are derived from general goals extracted from relevant IT agendas. In the end, the overall evaluation process will tell whether the project complies with these policies. By taking into account the policies derived and the general project goals obtained in Step (2), evaluation criteria are defined on project level in Step (6). These criteria must be suitable to assess whether the project meets its goals. From these project criteria, pilot-specific criteria are then derived in Step (7). Pilot-specific goals obtained in Step (3) are a relevant input to this step.

After completing the process steps described so far, all goals and (almost) all evaluation criteria are defined and set in relation according to the evaluation-criteria model introduced in Section 4.1. What is left to be done is the derivation of project phase specific evaluation criteria (Step (8)). This task is intentionally shifted to the subsequent project-execution/evaluation phase (i.e. the next horizontal area), as the proposed evaluation model foresees a dynamic adaption of these criteria during the entire project life-cycle. In Step (8), project phase specific evaluation criteria are derived for each pilot, taking into account the respective pilot's pilot-specific criteria from Step (7) and project phase specific goals derived in Step (4). For all but the first project phase, feedback collected during the previous project phase is considered for the definition of criteria for the current project phase. This way, criteria can be dynamically adapted during the entire project life-cycle to consider changing circumstances.

The execution of a project phase according to the project setup is covered by Step (9). Step (10) is executed in parallel to Step (9) and collects evaluation data. This can be achieved e.g. through interviews, questionnaires, or the automated measuring or logging of data. The proposed evaluation scheme leaves the choice of the most suitable method to the respective project evaluators. Step (11) analyzes collected evaluation data in order to derive evaluation results for the current project phase. The concrete analysis process depends on the type of evaluation data collected. Hence, the proposed evaluation scheme does not apply any restrictions here. Evaluation results of the current project phase are subsequently aligned with other relevant stakeholders involved in the project in Step (12). Depending on the project setup, this can be e.g. pilot developers or pilot operators. This step gives involved stakeholders the chance to comment on results obtained, in order to ensure a consensual overall evaluation process. In case there is at least one more project phase to come, feedback

on the applied evaluation process is collected from all involved stakeholders (Step (13)). Again, the method applied to collect feedback is left open to the respective evaluators. Independent of the chosen method, collected feedback serves as input for the definition of project phase specific criteria for the next project phase. After completion of a project phase, and if there is another project phase to come, Step (14) is carried out by project executors. In this step, evaluation results of the just completed project phase are analyzed. If possible, lessons learned are derived that serve as input for the execution of the next project phase. This way, the project is continuously improved based on intermediate evaluation results.

After completion of the final project phase, the last two overall phases, i.e. Project Appraisal and Derivation of Findings are carried out. Step (15) is executed after completing the last iterative project and evaluation phases. In this step, project phase specific evaluation results are combined for each pilot to derive the overall pilot-specific evaluation results. If evaluation criteria have been defined according to the proposed evaluation-criteria model, this process step can be carried out efficiently, as pilot-specific evaluation results can be derived directly from project phase specific results. Once all pilot-specific evaluation results have been derived, Step (16) combines them to overall project evaluation results. Again, this is an efficient process, if evaluation criteria have been defined such that the degree of fulfillment can be derived from the lower layer of the evaluation-criteria model. Step (17) finally checks derived project evaluation results against relevant policies identified at the beginning of the project. This way, stakeholders can assess whether the project and its results comply with IT agendas, from which these policies have been derived. This process step concludes evaluation-related activities within the project. All evaluation results (pilot evaluation results, project evaluation results, and policy compliance) serve as input for the final project-appraisal phase. In the final Step (18), all evaluation results stemming from different layers of abstraction are combined to derive the most valuable lessons learned. Due to the multi-layered approach followed, detailed analyses of evaluation results can be conducted, including specific results as well as comparisons between pilots and even projects.

Although the above-described evaluation process specifies in detail necessary steps to be carried out, it deliberately remains abstract in certain aspects and gives evaluators, who apply the propose evaluation scheme in practice, room for parametrization. This makes the proposed scheme more flexible and applicable to a broader range of projects. In particular, it ensures that the evaluation scheme can be applied to all pilot-based research projects complying with the general project model introduced in Section 3.

Of course, leaving certain aspects unspecified imposes an additional task on the evaluators of a project. They need to parametrize the proposed evaluation scheme, in order to adapt it to the specifics of the respective project. In the following section, we show one possible parametrization by discussing the application of the proposed evaluation scheme in the EU-funded research project FutureTrust.

5 EVALUATION

Future Trust Services for Trustworthy Global Transactions (FutureTrust) is an international research project funded by the EU under the programme *H2020-EU.3.7. - Secure societies - Protecting freedom and security of Europe and its citizens*. The project consortium consists of 16 partners from 10 countries, including EU member states as well as third-party countries. The overall aim of FutureTrust is to support the practical implementation of the EU eIDAS Regulation (European Union, 2018) in Europe and beyond. Software components developed by FutureTrust are applied to real-world use cases by means of several pilots and demonstrators.

FutureTrust fully complies with the general project model described in Section 3. The project has a focus on the public-sector domain and is motivated by an EU agenda, as described in the project's funding programme (European Commission, 2018b). Furthermore, FutureTrust develops and operates a series of demonstrators and pilots. Thus, the evaluation scheme proposed in this paper is well suited for carrying out pilot evaluations in FutureTrust.

5.1 Parametrization

The proposed evaluation scheme intentionally remains generic in several aspects and hence needs to be parametrized before being applied to a concrete research project. In the case of FutureTrust, the following parameters have been chosen:

- **Number of Iterative Project Phases:** The evaluation scheme supports an arbitrary number of iterative project phases and corresponding evaluation phases. For FutureTrust, three phases have been defined. Evaluations are carried out before piloting (ex-ante evaluation), during piloting (mid-term evaluation), and after piloting (ex-post evaluation). This complies with FutureTrust's

overall project setup as defined in the project description.

- **Relation between Evaluation Criteria:** The proposed evaluation-criteria model enables the definition of evaluation criteria on multiple layers of abstraction. Criteria on neighboring layers should be set in relation to each other. This way, the fulfillment degree of higher-level criteria can be derived systematically from the fulfillment degree of lower-level criteria. The model does not impose any restrictions regarding the definition of relations between criteria. For the sake of simplicity, we have followed a rather simple approach, which assumes each criterion to be equally important.
- **Method for Evaluation-data Collection:** In each iterative project phase, evaluation data must be collected by evaluators. As FutureTrust piloting partners are distributed all over Europe, questionnaires have been prepared and sent out to piloting partners to collect necessary evaluation data.
- **Method for Alignment of Project Phase Specific Evaluation Results:** The proposed evaluation scheme gives all involved stakeholders the opportunity to comment on evaluation results in each project and evaluation phase. Due to the local dispersion of stakeholders within the project, FutureTrust follows again a document-based approach. Evaluation results are compiled into intermediary evaluation reports. These reports are sent out to all involved stakeholders in order to provide them the opportunity to give feedback.
- **Provision of Feedback:** Finally, the proposed evaluation scheme also defines a feedback channel from involved stakeholders to the evaluators. FutureTrust organizes regular meetings (online and face-to-face) that bring together involved stakeholders. These meetings can be used to bilaterally provide feedback as defined by the proposed evaluation scheme.

5.2 First Results and Lessons Learned

After applying the parametrization as described in Section 5.1, the proposed evaluation scheme has been applied to the FutureTrust project. In particular, the scheme has been applied twice, once for the evaluation of pilots and demonstrators, and once for analyzing their impact. The applied evaluation process was exactly the same, however, different evaluation criteria have been used for pilot evaluation and impact analysis. Based on EU agendas relevant for FutureTrust, the following general project criteria have been defined for pilot evaluation: *Security and Data Protection*, *Functionality*, *Usability*, *Interoperability*, *Reusability and Sustainability*, *Legal Compliance*, and *Compliance with Project Goals*. Accordingly, the following general project criteria have been defined for impact analysis: *Demonstration of Positive Business Case*, *Empowerment and Protection of Users*, *Increase of Use of Trust Services*, *Reduction of Administrative Overhead*, and *Adherence to Sufficient Technology Readiness Level (TRL)*.

From these general project criteria, pilot-specific evaluation criteria have been derived for each FutureTrust pilot and demonstrator. For each pilot and demonstrator, 24 pilot-specific criteria have been defined for pilot evaluation. In addition, 9 pilot-specific criteria have been defined for impact analysis for each pilot and demonstrator. Finally, project phase specific evaluation criteria have been derived (33 criteria for pilot evaluation and 9 criteria for impact analysis).

Based on the resulting project phase specific evaluation criteria, questionnaires have been prepared and sent out to pilot developers and operators. By analyzing returned filled questionnaires, the fulfillment degree of project phase specific evaluation criteria could be determined. Furthermore, the fulfillment of higher-layer criteria could be derived automatically, based on their relation to project phase specific criteria. As an illustrative example, Figure 5 and Figure 6 show first results of the conducted pilot evaluation and impact analysis. Obtained results demonstrate the applied evaluation scheme's capability to yield evaluation results that are comparable among pilots.

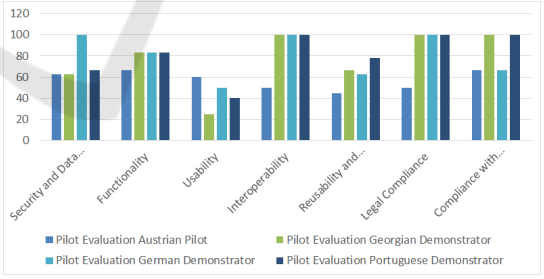

Figure 5: Results of FutureTrust pilot evaluation.

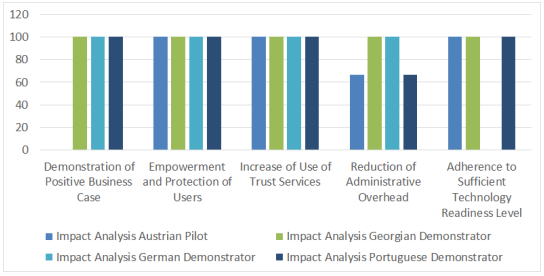

Figure 6: Results of FutureTrust impact analysis.

Overall, the successful completion of the described evaluation steps in the context of FutureTrust demonstrates the practical applicability of the proposed evaluation scheme. As the FutureTrust project is still ongoing, the overall evaluation process is not yet finished. However, most process steps (including the most challenging ones like the definition of evaluation criteria) as defined by the proposed scheme have already been applied successfully in practice.

6 CONCLUSIONS AND FUTURE WORK

Its successful application within a concrete international research project shows that the proposed evaluation scheme meets its goals. By relying on a hierarchical evaluation-criteria model, the scheme enables an in-depth evaluation of specific pilots, while still guaranteeing comparability of obtained evaluation results. Ultimately, the proposed scheme yields more valuable evaluation results, from which all involved stakeholders can benefit in the end. Especially funding bodies can take advantage of more homogenous evaluation results, which supports them in steering research activities into the right directions, and in complying with relevant IT agendas.

Lessons learned from applying the proposed evaluation scheme to the research project FutureTrust are currently used to apply final minor improvements to the scheme. For the future, we also plan to apply the scheme to other projects, to further test its project-independent applicability. At the same time, we aim to extend the applicability of the proposed scheme. While its current focus lies on pilot-based projects from the public sector, we plan to make it applicable to other project types as well.

ACKNOWLEDGEMENTS

This work has been supported by the FutureTrust project (N.700542) funded under the programme H2020-EU.3.7. - Secure societies - Protecting freedom and security of Europe and its citizens (2013/743/EU of 2013-12-03).

REFERENCES

Asosheh, A., Nalchigar, S., and Jamporazmey, M. (2010). Information technology project evaluation: An integrated data envelopment analysis and balanced scorecard approach. *Expert Systems with Applications*, 37(8):5931 – 5938.

Eilat, H., Golany, B., and Shtub, A. (2008). R&d project evaluation: An integrated dea and balanced scorecard approach. *Omega*, 36(5):895 – 912.

European Commission (2018a). Digital Agenda for Europe. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:si0016&from=DE. Online; accessed 05 July 2018.

European Commission (2018b). H2020-EU.3.7. - Secure societies - Protecting freedom and security of Europe and its citizens. https://cordis.europa.eu/programme/rcn/664463_en.html. Online; accessed 05 July 2018.

European Commission (2018c). Shaping the Digital Single Market. https://ec.europa.eu/digital-single-market/en/policies/shaping-digital-single-market. Online; accessed 05 July 2018.

European Union (2018). Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014R0910. Online; accessed 05 July 2018.

Khan, Z., Ludlow, D., and Caceres, S. (2013). Evaluating a collaborative it based research and development project. *Evaluation and Program Planning*, 40:27 – 41.

Knall, T., Tauber, A., Zefferer, T., Zwattendorfer, B., Axfjord, A., and Bjarnason, H. (2011). Secure and privacy-preserving cross-border authentication: the stork pilot "saferchat". In Andersen, K. N., editor, *Proceedings of the Conference on Electronic Government and the Information Systems Perspective (EGOVIS 2011)*, LNCS 6866, pages 94 – 106. Springer.

Leitold, H. and Posch, R. (2012). *STORK - Technical Approach and Privacy book title: Digital Enlightenment Yearbook 2012*, pages 289 – 306. Bus, J. , Crompton, M. , Hildebrandt, M. , Metakides, G.

Rath, C., Roth, S., Bratko, H., and Zefferer, T. (2015). Encryption-based second authentication factor solutions for qualified server-side signature creation. In Francesconi, A. K. E., editor, *Electronic Government and the Information Systems Perspective*, volume 9265 of *Lecture Notes in Computer Science*, pages 71 – 85. Springer.

Suzic, B. and Reiter, A. (2016). Towards secure collaboration in federated cloud environments. In *2016 11th International Conference on Availability, Reliability and Security (ARES)*, pages 750–759.

Tauber, A., Zwattendorfer, B., and Zefferer, T. (2011). Stork: Pilot 4 towards cross-border electronic delivery. In Verlag, T., editor, *Electronic Government and Electronic Participation - Joint Proceedings of Ongoing Research and Projects of IFIP EGOV and ePart 2011*, pages 295 – 301.