



FP6-2004-27020

## **Access-eGov**

### Access to e-Government Services Employing Semantic Technologies

**Instrument:** STREP

**Thematic Priority:**

SO 2.4.13 Strengthening the integration  
of the ICT research effort in an enlarged Europe

## **D8.1 Trial Evaluation Strategy**

**Start date of project:** January 1, 2006      **Duration:** 36 months

**Date of submission:** April 19, 2007

**Lead contractor for this deliverable:** German University in Cairo

**Revision:** Final 1.0

**Dissemination level:** PU

**Acknowledgement:** The Project is funded by European Commission DG INFSO under the IST programme, contract No. FP6-2004-27020.

**Disclaimer:** The content of this publication is the sole responsibility of the authors, and in no way represents the view of the European Commission or its services.

## D8.1 Trial evaluation strategy

<b>Work package:</b>	WP8	<b>Task:</b>	T8.1
<b>Date of submission:</b>	April 19, 2007		
<b>Lead contractor for this deliverable:</b>	German University in Cairo (GUC)		
<b>Authors:</b>	Ralf Klischewski (GUC) Stefan Ukena (GUC)		
<b>Version:</b>	1.0		
<b>Revision:</b>	Final		
<b>Dissemination level:</b>	PU		

### Project partners:

Technical University of Kosice (TUK), Slovakia (Coordinator); University of Regensburg (UR), Germany; German University in Cairo (GUC), Egypt; Intersoft, a.s. (IS), Slovakia; EMAX S.A. (EMA), Poland; Kosice Self-Governing Region (KSR), Slovakia; Cities on Internet Association (COI), Poland; e-ISOTIS (ISO), Greece; Municipality of Michalovce (MI), Kosice; City Hall of Gliwice (GLI), Poland; State Government of Schleswig-Holstein (SHG), Germany.

### Abstract

This document lays out the common trial evaluation strategy in order to get measurable feedback from planned evaluation activities resulting in guidance for further user-oriented research and development within the Access-eGov project.

### Document Sign-off

Nature of Sign-off (Reviewed/Approved/Submitted)	Name	Role	Participant short name	Date
Reviewed	Ralf Klischewski	TL	GUC	17/04/2007
Approved	Gabriela Hajdukova, Dariusz Wozniak	WPL	KSR, COI	18/04/2007
Submitted	Tomas Sabol	PM	TUK	19/04/2007

### Document Change Record

Date (d/m/y)	Version	Contributor(s)	Change Details
12/11/2006		Ralf Klischewski	Initial version
23/11/2006	0.2	Ralf Klischewski, Stefan Ukena	
27/11/2006	0.3	Ralf Klischewski	D8.1 and D8.2 split into two individual reports
10/04/2007	0.9	Ralf Klischewski, Stefan Ukena	Instrument definition based on user partners' pilot specification; defining user test lab
17/04/2007	1.0	Ralf Klischewski, Stefan Ukena	Final version ready for approval

### Files

Software Products	User files / URL
Microsoft WORD	

## Table of Contents

Executive Summary .....	5
1 Introduction .....	6
2 Systematic Approach to Trial Evaluation .....	7
2.1 Information Integration and Information Quality .....	7
2.2 Process Improvement .....	9
2.3 Other issues .....	11
3 Instruments and Measurements for Evaluation .....	13
3.1 Suggested instruments for evaluation.....	13
3.2 Evaluation of improving information integration and quality.....	14
3.3 Evaluation of improving processes .....	16
3.3.1 Slovakian Pilot .....	16
3.3.2 Polish Pilot .....	16
3.3.3 German Field Test.....	17
3.4 Evaluation of improving other issues.....	17
3.5 User test lab (Egypt).....	19
4 Conclusion.....	21

## Executive Summary

Research and development within Access-eGov are following a “user driven” approach, i.e. systematic effort is exerted ensuring that the R&D results are serving the users’ needs. Holding two trials of the pilot applications and the field test is a vital part of this approach as it will provide the context for all implementation and evaluation activities. Both trials go through the phases of preparation, monitoring/documentation, and evaluation according to specific criteria. This document lays out the common trial evaluation strategy in order to get measurable feedback from planned evaluation activities resulting in guidance for further user-oriented research and development within the project.

The administrations of the two Eastern European pilot sites describe themselves as being in a transition phase in which the administration’s ICT environment is expected to be enhanced during the project duration. This includes several changes from traditional services to electronic services, inside and outside the scope of the project. All trials have in common that the functionality being tested is to support user activities in the area of seeking personal information and/or documents, seeking administrative approval which may be case-based, and selected life events combining these informational and transactional services.

The improvements of services to be included in the trial will be traced according to: (a) informational output, as well as (b) in the process itself. Evaluation of improvements will build on criteria related to (1) information quality, (2) process automation, rationalization and reengineering (including change to e-service delivery), as well as (3) other issues such as accessibility, ease of use, security and trust. In this document, the above criteria are explained in detail, instruments and measurements for evaluation are introduced, and for each detailed aspect of the evaluation criteria certain instruments are suggested to be applied during the trials and afterwards. Furthermore, the setup of the user test lab (located in Egypt) is described in order to systematically challenge the technology and application for technical feasibility and service quality from an outside view.

The evaluation strategy and framework is the same for all trial sites, and all user partners have elaborated on it to describe their specific trial and the specific functionality according to the characteristic of the services included in pilots and field test. In the deliverable D8.2, each trial is described in detail, including the specific user activities for which the Access-eGov system is expected to provide functional support.

# 1 Introduction

The aim of evaluation is to ensure that the project meets the needs of both user partners and citizens. The overall evaluation process can be seen as three phases:

1. Preparation of trials;
2. Monitoring & documentation;
3. Evaluation according to specific criteria.

Research and development within Access-eGov are following a “user driven” approach, i.e. systematic effort is exerted ensuring that the R&D results are serving users’ needs. Holding two trials of the pilot applications and the field test is a vital part of this approach as it will provide the context for all implementation and evaluation activities. The above phases will be stepped through twice, once for each of the two trials.

This document lays out the trial evaluation strategy in order to get measurable feedback from planned evaluation activities resulting in guidance for further user-oriented research and development within the project. To this end, we develop a common evaluation strategy to be applied at the three test sites as well as at the user test lab (Egypt) to systematically challenge the technology and application for technical feasibility and service quality from an outside view.

The common evaluation strategy then paves the ground for designing the trials: deliverable D8.2 provides a specification of each pilot and the field test, including an activity plan, as well as responsibilities how to run and evaluate the trials. The non-trivial scenarios for the trials will span areas such as

- [1] seeking personal information and/or documents (e.g. birth certificate),
- [2] seeking administrative approval which may be case-based (e.g. work permit, opening new business), and
- [3] selected life events which will make use of [1] and [2].

The results of use case analysis and planning will play the role of a common ground for the development of necessary knowledge models, configuring Access-eGov components to meet the required functionality and carrying out the two trials. To ensure a systematic and coherent approach across all pilots and the field test, the following chapter presents methodological considerations resulting in overarching categories how to describe in detail the existing situation and the (expected) improvements from the user’s point of view. Based on this approach chapter 4 describes the instruments and measurements to be applied during or after the trials, including the setup of the user test lab (located in Egypt).

## 2 Systematic Approach to Trial Evaluation

To ensure a systematic and coherent approach across all pilots and the field test, this chapter presents methodological considerations resulting in overarching categories how to describe in detail the existing situation, the expected improvements from the user point of view and how to set up measurements as well as an evaluation strategy. The scope of the evaluation has already been outlined in D2.2, covering the implementation of the activity scenarios, stakeholder opinions, requirement fulfilment, and application of Semantic Web technology for e-government. The definition of scope and the level of description and specification of the trials must be detailed enough in order to evaluate the extent to which they cover defined user requirements and to foresee the kind of information to be gained from trials for forthcoming research and development.

Within this project, e-government services are considered as performing information processing with the aim of information integration from the user point of view who have a certain goal in mind: a service to be included in the trial shall be briefly described by its informational input and output, the service providers involved, and the set of activities the service consumers have to perform (cf. process models introduced in D2.2). Improvements from the user point of view can be found in: (a) the informational output, as well as in (b) the process itself. Improvement evaluation therefore can build on criteria elaborated in relation to (1) information quality, (2) process automation, rationalization and reengineering (including change to e-service delivery), as well as (3) other issues such as accessibility, ease of use, security and trust.

These criteria are elaborated in the following subsections. They will be used to highlight the differences between the situation “as is” and the expected changes during the trials. Based on the analysis of expected changes, certain measurements will be defined which can be operationalized and practically applied during the trials (chapter 4).

The evaluation following the first trial will most importantly provide an agenda for the developers. The evaluation following the second trial will instead try to gain insights about the general applicability of the Access-eGov technology as well as input for the methodological framework. This is intended to ensure broad acceptance and application of the developed technology.

### 2.1 Information Integration and Information Quality

Technically speaking, information integration aims at uniform data access and uniform information representation across a variety of (distributed) sources. However, from the user point of view information integration shall enhance the “fit for use” of the output of information process. When evaluating the output of an information process, researchers usually refer to the concept of information quality (IQ) and the recurring criteria elaborated during the last two decades. Huang et al.<sup>1</sup> have produced the most extensive list of information quality criteria defining 15 dimensions of grouped into four classes:

- *intrinsic quality*: accuracy, objectivity, believability, and reputation;
- *accessibility quality*: access, and security;

---

<sup>1</sup> Huang, K.-T., Lee, Y.W. and Wang, R.Y. (1999), *Quality Information and Knowledge*, Prentice-Hall, New York, NY.

- *contextual quality*: relevancy, value-added, timeliness, completeness and amount of data;
- *representational quality*: interpretability, ease of understanding, concise representation, and consistent representation.

With reference to the general description of a service in focus (see above), improvements of informational outputs of e-government services from the user point of view shall be described and measured along these dimensions as follows (see table below):

- (1) The output of the existing service is described by indicating the characteristics of the information in each dimension (as far as applicable). This characterisation serves as the baseline for all subsequent steps. To highlight improvements thereafter, the initial description must include shortcomings or insufficiencies in at least some of the dimensions (i.e. reasons for improvement). However, existing high quality of output should be acknowledged as well in order to ensure (or to control for) that this will be preserved throughout the trials.
- (2) The expected output of the improved service is described by indicating the characteristics of the information in the same dimensions. The focus here should be on the difference to the situation as is (see 1) or, for the 2<sup>nd</sup> trial, on the difference to the 1<sup>st</sup> trial. It should be noted that the 1<sup>st</sup> trial is mainly concerned with the evaluation of individual components while the 2<sup>nd</sup> trial extends evaluation to the integrated overall AeG platform.
- (3) The actual output will be measured and evaluated in comparison to the expected output, based on measurements defined for each case (see D 8.2).

<b>Quality dimension</b>	<b>Before trial (as is)</b>	<b>1<sup>st</sup> Trial</b>		<b>2<sup>nd</sup> Trial</b>	
		Expected	Measured	Expected	Measured
intrinsic quality					
– accuracy					
– objectivity					
– believability					
– reputation					
accessibility quality					
– access					
– security					
contextual quality					
– relevancy					
– value-added					
– timeliness					
– completeness					
– amount of data					
representational quality					
– interpretability					
– ease of					



understanding					
– concise representation					
– consistent representation					

However, measuring the information quality is far from trivial. Lillrank<sup>2</sup> (p. 698f) distinguishes two basic types of information quality:

- The quality of **information as an artefact**: “The producer of information has an intention to produce a symbolic representation of certain entities or events, put it into a context, assign it a pre-defined meaning and transmit it to a receiver. The expectation is that the receiver will capture the meaning as intended.” If the transmission is repeated with some regularity, we can define an information process which is subject to modelling and (partly) automation (e.g. order-to-delivery systems).
- The quality of **information as deliverable**: “A deliverable is defined through negotiations between producer and receiver. [...] the receiver selects and priority orders the available deliverables based on subjective criteria or negotiates with the intention to achieve optimal fitness for use. Information as a deliverable emerges from a communicative structure, usually an informal, open system setting, where the truth conditions of data, relevance of contexts, suitability of knowledge and appropriateness of meanings are negotiated.”

In practice, services such as in e-government are most often combinations of artefacts and negotiated results. However, for describing the trials and setting up measurements this distinction is crucial because in the first case it presupposes a stable context with predefined meanings of data accepted by all stakeholder involved – whereas in the second case the frame for valuing the information may differ between information provider and consumers, or needs to be established, clarified etc. For example, when a citizen is looking online for a specific tax form, providing her with the URL for download is exactly what she expected within the already established context. But (at the same time) alerting her that this form is applicable only in certain cases and that in her case most likely another form and/or procedure is valid – this kind of information was not expected, and in this case the information quality with respect to customer expectation maybe much higher but is also much more difficult to measure.

This distinction must be kept in mind during evaluation of information quality improvement: It should be clarified if the e-government service users already share the context of information value in focus, or this needs to be established. Any user survey must control for this aspect, and also all other instruments suggested must take it into account when determining the effectiveness of the newly developed components.

## 2.2 Process Improvement

In general, information processing can be the changing (processing) of information in any manner detectable by an observer (Wikipedia). Here, we focus on the information processing from the users’ point of view, i.e. the activities the user has to perform in order to have an

<sup>2</sup> Lillrank, P. (2003), The quality of information, International Journal of Quality & Reliability Management, Vol. 20 (6), 2003, pp. 691-703

informational input provided returned with a desired informational output. The scope of the tasks and activities has been defined in D2.2 and was the basis for the requirement elicitation summarized in section 2.4 of that deliverable. Important criteria for the process evaluation can be derived from the area of business process automation & reengineering, the most important are considered process cycle time (i.e. time between providing informational input and receiving desired informational output), number and the complexity of steps within user activities, integration/reduction of various media (e.g. paper-based vs. electronic)

With reference to the general description of a service in focus (see above), improvements of informational process of e-government services from the user point of view shall be described and measured along these dimensions as follows (similar to 3.1; see table below):

- (1) The process of using the existing service is described by indicating the characteristics of the process (as far as applicable). This characterisation serves as the baseline for all subsequent steps. To highlight improvements thereafter, the initial description must include shortcomings or insufficiencies in at least some of the dimensions (i.e. reasons for improvement). However, existing high quality of output should be acknowledged as well in order to ensure (or to control for) that this will be preserved throughout the trials.
- (2) The expected output of the improved service is described by indicating the characteristics of the process in the same dimensions. The focus here should be on the difference to the situation as is (see 1) or, for the 2<sup>nd</sup> trial, on the difference to the 1<sup>st</sup> trial.
- (3) The actual output will be measured and evaluated in comparison to the expected output, based on measurements defined for each case (see also chapter 4).

<b>Quality dimension</b>	<b>Before trial (as is)</b>	<b>1<sup>st</sup> Trial</b>		<b>2<sup>nd</sup> Trial</b>	
		Expected	Measured	Expected	Measured
Cycle time					
User activities (number & complexity of steps)					
– <activity 1>					
– <activity 1>					
– ...					
Media integration					
– Paper-based vs. electronic					

Information processing also can be viewed as consisting of creation, publication, organization, access and maintenance. This view may apply to the annotation process to be performed by administration employees (as users of the annotation component) and shall lead to a separate evaluation approach.

## 2.3 Other issues

Besides the information process and its informational output, it is important to investigate other non-functional issues such as accessibility, usability of interfaces, security and trust. The quality dimensions taken into account here are based on the relevant categories that were already used for the elicitation of user requirements (cf. D2.2 “User requirement analysis & development/test recommendations”, section 2.4 “Summary of requirements for developers”). The application of the criteria follows the same procedures as in indicated above (2.1., 2.2).

Within the context of this R&D project, accessibility and ease of use are the most important issues from the user’s point of view. However, the evaluation strategy mainly focuses on relative improvements (before vs. after) whereas ‘ease of use’ should apply to any developed Access-eGov component even if there has been no equivalent before. Whenever feasible in terms of improvement, aspects of ‘ease of use’ (e.g. relevance, completeness, media integration) are specifically addressed in the above mentioned criteria (see section 3.1 & 3.2); in addition accessibility and ease of use will be explicitly tested in terms of overall user perception (see also section 4.4).

<b>Quality dimension</b>	<b>Before trial (as is)</b>	<b>1<sup>st</sup> Trial</b>		<b>2<sup>nd</sup> Trial</b>	
		Expected	Measured	Expected	Measured
Accessibility / Ease of use					
Accessibility for impaired citizens according to W3C-guidelines like WAI (see "Web Accessibility check list")					
Multi channel support					
Support for Authentication and Authorization Infrastructure functionality					
Search facilities					
Openness to external partners					
Value added services <sup>3</sup>					
User support for consumers / providers					
Multi-Lingual Support					
Security					
– digital rights management for annotated content					

<sup>3</sup> In D2.2 this was called „Quality of Service“.

- transmission (data encryption)					
- support for authentication and authorization infra-structure functionality					
- handling of user data (privacy)					

### 3 Instruments and Measurements for Evaluation

The evaluation employs a qualitative approach, i.e. for the criteria defined in the previous section the evaluation results do not aim at statistical relevance but provide qualitative input for developers' agenda. The following section introduces the instruments to be applied during the trials or afterwards. The sections thereafter suggest the instruments for evaluation of improving information integration and quality, improving processes and improving other issues, along with the recommended frequency of their application during each trial at each site. The last section of this chapter introduces the setup of the user test lab located in Egypt.

#### 3.1 Suggested instruments for evaluation

The following instruments can be used for measuring the quality dimensions.<sup>4</sup> Each instrument is described in one sentence followed by its strength and weaknesses.

1. User survey: After the user has used the system to perform a specific task, she answers a number of questions from a questionnaire (either by being asked them or by filling out the questionnaire).
  - Strengths: High number of users can be considered
  - Weaknesses: Results depend on the quality of the questionnaire
2. User interview: After the user has used the system to perform a specific task, she is interviewed in the form of a guided interview (as opposed to a questionnaire).
  - Strengths: Potentially very rich set of results even with low number of observations; results may be used to evaluate more than one quality criteria.
  - Weaknesses: Costly; quality of results depend on interviewer's skill; difficult to evaluate results
3. User observation: After asking a user to perform a specific task by using the system while thinking aloud, the user's actions are recorded on film for later evaluation.
  - Strengths: Potentially very rich set of results even with low number of observations; results may be used to evaluate more than one quality criteria.
  - Weaknesses: Very costly; very difficult to evaluate results; evaluation of results needs some expertise; need for technical infrastructure to record user's actions
4. Assessment by expert: An expert assesses the quality in question and gives her expert opinion.
  - Strengths: Possibly measurement results of very high quality even for difficult measures.
  - Weaknesses: Highly dependent on expert's skill; possibly difficult to find suitable expert; potentially costly.
5. Round table discussion of experts and/or users: A group of experts and/or users is asked to discuss the quality in question during a round table discussion.

---

<sup>4</sup> cf. Flick, Uwe: *Qualitative Forschung – Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften*. Rowohlt Taschenbuch Verlag, 2000

- Strengths: Possibly comprehensive measurement results of high quality even for difficult measures; discussing results from other instruments and receiving feedback from users about these results; in case of experts: less dependent on single expert's skill than assessment by single expert.
- Weaknesses: Possibly difficult to find suitable number of participants; potentially very costly; may be difficult to obtain a single, clear result

When applying these instruments at the site of each trials, the user partners will select users (of each component in use) and experts (matching the area of concern) to be included in the evaluation.

Furthermore, the technical assessment of the running components is also expected to point to qualitative use aspects; available and relevant material (from monitoring by developers and/or user test lab; see below) should be included in expert assessment and round table discussions.

### 3.2 Evaluation of improving information integration and quality

This section suggests instruments for evaluation of improving information integration and quality and the frequency of their application during each trial at each site.

Quality dimension	Statement Evaluation scale: Fully agree (1) - fully disagree (5)	Suggested Instruments
<b>intrinsic quality</b>		
accuracy	a) The information is accurate with regard to legislation, current procedures etc.	Either 2 assessments by experts <b>or</b> 1 round table discussion of experts.
objectivity	n/a	
believability	a) The source (author) of every piece of information can be easily identified b) Links to content of external parties are clearly marked as such c) The information is believable	At least 10 users surveyed and <b>optionally</b> 2 user interviews.
reputation	a) The system is a good source for information about services b) The information is trustworthy	At least 10 users surveyed and <b>optionally</b> 2 user interviews.
<b>accessibility quality</b>		
access	a) It is easy to locate the necessary information	At least 10 users surveyed and <b>optionally</b> and optionally 1 assessment by expert
security	a) Storage, processing, and transmission of user data is secure	Either 1 assessment by expert <b>or</b> 1 round table discussion of experts.

<b>contextual quality</b>		
relevancy	<p>a) The information is relevant for the given task</p> <p>b) The information provided is specific for the given user context</p>	At least 10 users surveyed and <b>optionally</b> 1 assessment by expert.
value-added	<p>a) The needs of different groups of users are well supported</p> <p>b) Information is provided for a high number of different user groups</p> <p>c) The information from different sources is well integrated</p>	Either 1 assessment by expert <b>or</b> 1 round table discussion of experts.
timeliness	<p>a) The information is up-to-date with respect to the current status (opening hours etc.)</p> <p>b) The system supports annotation editors to keep the information up-to-date</p> <p><b>Note: In this case, users must be annotation editors who have used the service annotation component.</b></p>	At least 10 users surveyed.
completeness	<p>a) The user was able to determine whether the information was complete to the given task</p> <p>b) The information is complete with respect to the given task</p> <p>c) The level of detail of the information was sufficient for given task</p>	At least 10 users surveyed and <b>optionally</b> either 1 assessment by expert or 1 round table discussion of experts.
amount of data	a) The amount of data is adequate with respect to the complexity of the supported process	At least 10 users surveyed and <b>optionally</b> either 1 assessment by expert or 1 round table discussion of experts.
<b>representational quality</b>		
interpretability	n/a (see next, "ease of understanding")	
ease of understanding	<p>a) The user was able to comprehend the information</p> <p>b) It is easy to identify the relevant parts of the content</p>	At least 10 users surveyed and <b>optionally</b> either 1 assessment by expert or 1 round table discussion of experts.
concise representation	a) The information was presented in a short and concise way	At least 10 users surveyed.

consistent representation	a) The information is presented in a consistent way	At least 10 users surveyed and <b>optionally</b> either 1 assessment by expert or 1 round table discussion of experts.
---------------------------	---	--

### 3.3 Evaluation of improving processes

This section suggests instruments for evaluation of improving processes and the frequency of their application during each trial at each site.

#### 3.3.1 Slovakian Pilot

<i>Quality dimension</i>	<b>Before trial (as is)</b>	<b>Expected Improvement</b>
<b>Cycle time</b>		
User activities (number & complexity of steps)		
Information providing - Semantic annotating of resources - Putting resources on relevant web pages	At present there are none of the steps done on web-pages of the public administrations.	<ul style="list-style-type: none"> <li>▪ There will be more relevant information available for users</li> <li>▪ Information will be more easily searchable - better structured information</li> <li>▪ Information relevant to particular user's case</li> <li>▪ Links to original sources of information (legal regulations for example)</li> </ul>
...	...	▪ ...

#### 3.3.2 Polish Pilot

<i>Quality dimension</i>	<b>Before trial (as is)</b>	<b>Expected Improvement</b>
Cycle time		
User activities (number & complexity of steps)		
- Information providing (two steps: resources annotating and putting them on web page)	No semantic annotations, information available on web page in a descriptive way.	<ul style="list-style-type: none"> <li>▪ More relevant information</li> <li>▪ Better organised information (most important issues and additional ones)</li> <li>▪ Information relevant to particular user's case</li> <li>▪ Links to original sources of information (legal regulations for example)</li> </ul>
- ...	...	▪ ...



### 3.3.3 German Field Test

<i>Quality dimension</i>	<b>Before trial (as is)</b>	<b>Expected Improvement</b>
Cycle time	Citizens are asked not to register earlier than 6 months in advance of the marriage – therefore the cycle time is shorter than 6 months and may only be a few days	
User activities (number & complexity of steps)		
– <marriage>	<p>Required are at least 3 complex steps (if the couple already knows which documents are required and does not have to produce any new documents):</p> <ul style="list-style-type: none"> <li>• Hand in documents</li> <li>• Registration (formal application)</li> <li>• Marriage ceremony</li> </ul> <p>If the couple does not know which documents are required and has to request some of them:</p> <ol style="list-style-type: none"> <li>1. Gather information on requirements</li> <li>2. Request required documents (depending on the situation this may involve several other steps)</li> <li>4. Receive documents from different administrations</li> <li>3. Hand in documents</li> <li>4. Registration</li> <li>5. Marriage ceremony</li> </ol>	<p>A user has to do the same steps as before but can do several of these online so that he may gather information on requirements and request the required documents all online in one complex step (without having to go (physically) to different locations)</p> <ol style="list-style-type: none"> <li>1. Gather information and request required documents online</li> <li>2. Receive and hand in documents (online)</li> <li>3. Registration (online)</li> <li>4. Marriage ceremony</li> </ol>
– ...	...	...

### 3.4 Evaluation of improving other issues

This section suggests instruments for evaluation of improving other issues and the frequency of their application during each trial at each site.

<i>Quality dimension</i>	<b>Statement</b> (Scale: Fully agree (1) - fully disagree (5))	<b>Suggested Instruments</b>
--------------------------	--	------------------------------

Accessibility / Ease of use	<ul style="list-style-type: none"> <li>a) All relevant information is available from a single point of access</li> <li>b) All electronically available information is made available</li> <li>c) The personal assistant is easy to use</li> <li>d) The application is easy to use</li> </ul>	At least 10 users surveyed. Plus either assessment by 2 experts <b>or</b> 1 round table of experts.
Accessibility for impaired citizens according to W3C-guidelines like WAI (see "Web Accessibility check list")	<ul style="list-style-type: none"> <li>a) The personal assistant meets common accessibility criteria (like WAI, BIK)</li> </ul> <p>Note: In addition, the level of conformance to WAI will be assessed and documented by e-ISOTIS.</p>	Either assessment by 1 experts <b>or</b> 1 round table of experts.
Multi channel support	<ul style="list-style-type: none"> <li>a) The service and information about services are provided through various channels (e.g. Web, email, fax)</li> </ul>	Either assessment by 1 experts <b>or</b> 1 round table of experts.
Search facilities	<ul style="list-style-type: none"> <li>a) The search facility was easy to use</li> <li>b) The search facility delivered good results</li> </ul>	At least 10 users surveyed.
	<ul style="list-style-type: none"> <li>a) The personal assistant is easy to use</li> <li>b) The application is easy to use</li> </ul>	At least 10 users surveyed.

Some more features (from the table “Expected improvement of other issues”) can be simply evaluated by a yes/no checklist (see below). Again, it is up to the user partners to determine who is best suited to complete each item.

Is the feature available?	Yes	No	Comment
Support for authentication and authorization infra-structure functionality			
Openness to external partners			
Value added services <sup>5</sup>			
User support for consumers / providers			
Multi-Lingual Support			
<b>Security</b>			

<sup>5</sup> In D2.2 this was called „Quality of Service“.

– digital rights management for annotated content			
– transmission (data encryption)			
– support for authentication and authorisation infra-structure functionality			
– handling of user data (privacy)			

### 3.5 User test lab (Egypt)

A user test lab is planned at GUC (Egypt) to systematically challenge the technology and application for technical feasibility and service quality from an outside view. The objective of the test lab is to carry out aspects of evaluation that do not repeat but complement the evaluation approach implemented at each trial site, with the aim of providing additional valuable feedback for subsequent technical development as well as for preparing guidelines for future real life use of the Access-eGov components.

The test lab is planned to carry out a series of “experiments”, each testing hypotheses about the role of one independent variable on another dependent variable. The hypotheses are taken from the user requirements analysis, the system design and the pilot specifications all of which were translated into specifications and therefore guiding the implementation (e.g. requirements 1.1.1 “Access to services of Access-eGov is enabled through a single point of entry”, 1.2.1 “Access-eGov Personal Assistant adheres to WAI specification”, etc.). However, the test will only focus on the information consumer perspective (not information or infrastructure provider), i.e. how the trial applications appear to the web users with the role of concerned a citizen or business.

Evaluation focuses on technical, semantic, as well as pragmatic aspects. During the first trial, the test lab will be targeting mainly technical and only some semantic and pragmatic aspects because it is expected that only core user interfaces will be available in English at that time. During the second trial, emphasis will be put more on pragmatic and semantic aspects while maintaining the most important technical test criteria:

Testing of *technical aspects* will be based on a list of test items mainly provided by the project partners involved in development (e.g. performance indicators like network accessibility, session duration, response time, but also possible system malfunctions and failures). Test results will be used to alert developers about bottlenecks and shortcomings of their implementations so that remedial action can be taken.

Testing of *semantic aspects* will focus on operation and usage of the semantic layer seeking to find out e.g. which parts of an ontology were used, how semantic constructs from various ontologies were connected with each other, what kind of semantic matches were performed etc. This evaluation can be performed in two basic ways: (a) test users will be given specific application tasks of which the system’s responses will provide adequate indicators to be documented and evaluated (this option is restricted for the second trial when full English interface is available); (b) the Access-eGov framework provides special inspection and monitoring options that can be utilized by test lab users (based on English translation of the ontologies in use). In both cases, the evaluation will be based on a set of “competence questions” that test the performance concerning processing of meanings relevant to the domain and to

users' cases. Test results should indicate in which aspects the semantic constructs developed sufficiently support the Access-eGov applications (or not) in order to find out the needs for improvement.

Testing of *pragmatic aspects* will focus aspects how well the system is able to fulfil some of the users' needs that are not systematically controlled for during onsite evaluation. Test criteria will include mainly accessibility aspects and use of applications across language barriers. Test results should indicate adequacy of developed interfaces and of modelling administrative issues as the basis for component functions. Note: the trials have been set up mainly for local users; therefore the test results should not be mistaken as giving indication about actual user satisfaction – they only provide *additional* input to improve research and development.

The series of tests will be set up as follows:

- GUC will provide one of its computer labs for this purpose in which 5-10 test users can easily work in parallel. Test users will be recruited from interested GUC students of Business Informatics, Computer Science and Management with IS major. Again, for this kind of experiments interest in the subject matter as well as sufficient computer and Internet skills are more important than choosing a sample with a greater variety of characteristics (or even a “representative” sample, but representative of what?). Only for certain aspects of accessibility some test maybe carried out at other places with selected users (to be specified in cooperation with ISOTIS).
- For both trials a number of experiments will be designed (probably between 5 and ten for each trial, depending on complexity and possible combination of test objectives). Each experiment is defined by test objective(s), test user tasks, number of test users and/or repetitions of the particular experiment as well as measurements of system and user behaviour. Each experiment will be carried out on the basis of a script to be handed to the assigned test users.
- Test user will use the Access-eGov components and record system behaviour as well as their own experience according to the instructions of the scripts provided. Each test user session based on one script lasts 60-120 minutes. An estimated number of hundred test user sessions shall be held during each trial period, involving an overall number of 20-30 test users.
- After completing all sessions, the script-based session records will be evaluated in order to draft test reports for each experiment to be shared with project partners.

In addition, it is expected that technical performance will be monitored and documented by the component operators (using standard reporting tools) and shared with the test lab in order to indicate if the tests truly challenge the critical aspects and to update test strategies if necessary.

## 4 Conclusion

The evaluation strategy and framework is the same for all trial sites, and all user partners have elaborated on it to describe their specific trial and the specific functionality according to the characteristic of the services included in pilots and field test. In the deliverable D8.2, each trial is described in detail, including the specific user activities for which the Access-eGov system is expected to provide functional support.

This document provides a solid basis for trial setup and evaluation as well as for the cooperation of all partners involved. However, not every detail can be foreseen, and changes to the evaluation strategy and framework may be necessary, in particular after the experience of the first trials. If so, a change management report will be included in the deliverable D8.3.