



FP6- 027685 MESH

D8.1a

Project assessment metrics and review guidelines

Contractual Date of Delivery:	M4 (June 2006)
Actual Date of Delivery:	28 July 2006
Workpackage:	WP8 Review and Assessment
Dissemination Level:	Public
Nature:	Report
Approval Status:	Approved
Version:	5
Total Number of Pages:	22
Distribution List:	WP8, MC members, European Commission
Filename:	mesh-wp8-d8.1-20070207-evaluation-metrics.doc
Keyword list:	metric, assessment, self-assessment, review

Abstract

This document describes MESH's process for self-assessment. By having regular internal as well as external reviews of the MESH deliverables, steered by a well-defined set of metrics, we will monitor the quality of deliverables, as well as progress and strategic direction of the project. After reviewing processes and metrics of other projects, it has been decided to follow the method of aceMedia, which is reproduced here.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

History

Version	Date	Reason	Revised by
01	2006-06-25	First draft	Franciska de Jong Yiannis Kompatsiaris
02	2006-07-24	Second draft	Yannis Avrithis
03	2006-07-25	Third draft for MC	Yannis Avrithis
04	2006-07-28	Approved version	Yannis Avrithis
05	2007-02-07	Revised version	Yannis Avrithis

Author list

UT	Franciska de Jong
ITI	Yiannis Kompatsiaris
ITI	Yannis Avrithis
Motorola	Paola Hobson

Executive Summary

Workpackage 8 of MESH is about assessment and evaluation. This report describes the metrics for assessment that will be for the review of MESH documents and deliverables that will undergo an internal or an external review. The metrics described in this deliverable have been compiled as part of the work in Task 8.1 (Metrics Assessment), and will be the basis for the review work as part of Task 8.2 (Peer Reviews). It is *not* a goal of this assessment to review the quality of software or benchmark MESH tools, which are dealt with elsewhere.

We have reviewed evaluation processes and metrics of other projects and, based on experience with the method, we decided that the method of aceMedia project [7] was adequate for MESH's purposes. We therefore reproduce below the method as developed for aceMedia.

Internal MESH participants will conduct internal reviews. Members of the MESH Advisory Board will perform external reviews. The results of the review will be used by the MESH management to track the progress and any deviations from the MESH objectives as well as the timing.

This document starts with describing the peer review process in MESH, including follow-up on the review results. Then we discuss the metrics that we will use for the internal and for the external review of MESH deliverables in self-assessment. We do not plan to review the software developed in MESH. The standard set of metrics that we will use falls into three categories:

- ◆ Metrics related to the scientific objectives of MESH;
- ◆ Metrics related to the document itself;
- ◆ Other metrics

Further, we have identified metrics specific for internal and external reviews. We also allow for specific metrics for the deliverable-under-review, to be specified by the authors of the deliverable.

As a final part of this deliverable we are giving an evaluation form to be used by reviewers. This form contains the metrics as well as a scoring mechanism. We request the reviewers to score metrics for the applicability and quality of the deliverable.

The deliverable concludes with a preliminary review schedule which in a more elaborate version will be the basis for the review work as part of Task 8.2.

The metrics described do not form a final set; we see it as a living document, to be updated when metrics appear to be not relevant, or relevant metrics are missing. The same follows for the Advisory Board and the schedule of reviews. As a consequence it should not be reviewed as a final deliverable.

Table of Contents

1. Abbreviations and Acronyms	5
2. Introduction.....	6
3. Description of the peer review process.....	7
3.1. Internal and external reviews.....	7
3.2. Analysis of the review results and further steps	7
3.2.1. Internal reviews.....	7
3.2.2. External reviews.....	8
4. Metrics.....	9
4.1. Background and literature	9
4.2. Standard set of metrics for all deliverables to be reviewed.....	10
4.2.1. Scientific objectives.....	10
4.2.2. Metrics related to the document itself	10
4.2.3. Other metrics	11
4.3. Special metrics related to internal reviews only.....	11
4.4. Special metrics related to Advisory Board reviews only.....	11
4.5. Additional metrics specified by the editor of a deliverable.....	12
5. Reviewer forms.....	13
6. Schedule of expected reviews	20
7. Conclusions / recommendations.....	21
8. References	22

1. Abbreviations and Acronyms

MC	Management Committee of MESH
WP	Workpackage in MESH

2. Introduction

The purpose of the assessment and evaluation workpackage is to monitor that the scientific and technical objectives of the MESH project are met. Specific objectives are:

- ◆ The development of appropriate metrics for measurement of quality;
- ◆ Delivery of reports on peer-reviewed deliverables;
- ◆ Recommendations for corrective actions arising from significant peer review comments and/or technical quality deviations.

The development of appropriate metrics is the subject of this report. It is important to understand the peer review process in order to understand the classification in metrics that we propose. Therefore section 4 will describe this process.

This deliverable is part of the work in Task 8.1 (Metrics Assessment). Along with the (preliminary) review schedule in section 6, it will be the basis for the review work as part of Task 8.2 (Peer Reviews). Note that it is *not* a goal of self-assessment to review the quality of software, which will be dealt with in WP6. It is also *not* a goal to benchmark MESH tools, which will be the outcome of Task 8.3 (Benchmarking). Instead, we concentrate on the metrics for measuring the quality of documents and deliverables in MESH.

We have reviewed evaluation processes and metrics of other projects and, based on experience with the method, we decided that the aceMedia method was adequate for MESH's purposes. The projects have similar structure, size and type of outcome so the evaluation plan of aceMedia matches very well the MESH project. We therefore reproduce below the method as developed for aceMedia and as presented in aceMedia deliverables D8.1 [8] and D.8.4 [9], authored by Paola Hobson (Motorola) and Hans van Gageldonk (Philips).

3. Description of the peer review process

3.1. Internal and external reviews

This workpackage reviews project performance, both internally by project partners, as well as by external reviewers.

People working on the MESH project itself will perform the *internal reviews*. This internal review of strategic deliverables (e.g. a cross-workpackage deliverable) helps the project to identify deviations from project objectives at an early stage.

External reviews will be done by members of the MESH Advisory Board, or by independent reviewers. The Advisory Board aims to alert the MESH partners to any possible drifting from the stated project and workpackage objectives. The external reviews will be aligned with the annual EC review cycle.

Reports from the reviews will be used by the MESH project management and the MESH partners to identify any required improvement.

To assist with self-evaluation activities in WP8, an independent MESH Advisory Board will be set up, comprising experts who are not partners in MESH (suitable Non-Disclosure Agreements would be in place). The Advisory Board will be asked periodically to review technical progress such that MESH remains innovative, forward looking, and ensures that MESH is producing work of high technical quality. Advisory Board membership will reflect the composition of the MESH consortium, and will therefore include people such as senior industrialists, prominent academics with expertise in key MESH technologies, and in order to maintain ethical standards, at least one expert in ethical issues (for example from a research institute specialising in such matters).

3.2. Analysis of the review results and further steps

3.2.1. Internal reviews

There are two places we expect to use the internal review scores:

1. The subtask responsible for the deliverable needs to carefully read the review and respond to it e.g. say how they will remedy any problems that are highlighted, issue a new version of the deliverable if necessary, replan future tasks etc, all with a specified deadline. When there are comments in the review that are *not* incorporated in the final document, the authors need to communicate these with the reviewers as well with the MESH MC. In this way it is ensured that actions are being taken up, or it is made clear why they are not incorporated. This can be done by light-weight communication such as an e-mail to the people involved. However, if the review results are used to *change the actions in the subtask*, then this needs to be communicated within the project, by means of a document stating the subtask planning changes.

2. By the WP leaders and within the MESH MC as part of the project management. The reviews are a means of verifying that deliverables are being produced of sufficient quality and commensurate with the resources used to produce them. If there are problems that highlight issues within subtasks, WP leaders and the MC should work together to help the subtask partners find remedies. This definitely applies to feedback that leads to re-planning a task or workpackage, or even a re-direction of the project itself. These issues need to be communicated at the workpackage and MC level.

Each internal review will have a deadline of at most two weeks from the availability of the review report. After the deadline, the final version of the deliverable should be ready and any changes in (sub)task plans need to have been communicated within the project.

3.2.2. External reviews

The external review scores have a more critical project direction implication. External reviews must be read by the deliverable editors, WP leaders, MC. If changes in project direction are implied by the review, this must be discussed by the MC. MC should make a recommendation for action within 3 weeks of receiving the review report (and response from the subtask/task/WP), which may mean holding a special MC meeting by phone conference if required.

4. Metrics

4.1. Background and literature

Note that the goal of self-assessment in MESH is *not* to review the quality of software. This will be dealt with in WP6. Instead we concentrate on the metrics for measuring the quality of documents and deliverables in MESH. Still it is interesting to get to know the lessons-learned in establishing sets of metrics for measuring software quality. Important aspects that are reported frequently [1][2][3][4]:

- ◆ The metrics should be well-defined;
- ◆ The set of metrics should be limited for people to work with;
- ◆ People working with the metrics should be convinced that the metrics help them in improving the quality of the software (code and documentation).
- ◆ The set of metrics is a “living document” in the sense that during projects some metrics prove to be useful, while other metrics appear to be missing.

For inspection of documents, the Fagan inspection (IBM, 1970s) was promoted and made available by T. Gilb [5]. One of the main goals is to improve by not injecting defects into a document, but by preventing them. Metrics include

- ◆ Number of defects per page.
- ◆ The time the reviewer invested into reviewing the document.

The defects are classified according to their seriousness: one could have “strongly advised to repair” defects as well as “advised to repair”. For the latter kind of defect, it is up to the author to decide whether to repair it. For the strongly-advised-to-repair defect, one should think of a significant deterioration of the fitness for use of the deliverable (“significant” to be judged by the reviewer). “Fitness for use” comprises:

- ◆ Usability;
- ◆ Performance;
- ◆ Any other factors affecting user acceptance;
- ◆ Maintainability;
- ◆ Etc.

and / or:

- ◆ consider loss of time (repair effort and associated effort) will result later on during the project or after delivery if not repaired (“considerable” means: about 1 hour or more).

In IST FP5 and FP6 projects, an example of metrics used for self-assessment was developed in the BUSMAN project [6]. This project produced an extensive document (D7.2) on project assessment and evaluation criteria. In this project an objective tree was constructed, consisting of about 40 objectives. For each objective a set of metrics was defined. This resulted in a huge set of metrics.

The situation was simplified in the aceMedia project [6], where two different set of metrics were developed, one for internal and one for external reviews. Both sets contained metrics related to scientific outcome of the work presented, metrics related to document itself, specific metrics for internal/external review, as well as custom metrics specified by authors. Thus, metrics remained generic, and any specific metrics required for each deliverable were specified at the time of the review by its authors.

The situation is similar in MESH. Compared to BUSMAN, MESH is a larger project with many goals, which would result in a much larger set of metrics. Second, a reviewer, and especially an external reviewer, should be served with a limited and well-defined set of

metrics to which the document should be reviewed. This is confirmed by the above study on metrics in software quality. Thus, the approach of aceMedia is followed, with some slight modifications/simplifications in evaluation metrics. The projects have similar structure, size and type of outcome so the evaluation plan of aceMedia matches very well the MESH project.

In order to reach an understanding of the comparison between deliverables, numerical scores will be applied to metrics, by the reviewers, such that the degree of satisfaction of each quality criterion can be assessed. Reviewers will be asked to rate each category in a review form (example shown in Section 6) and apply both a numerical score and some comments. A 4-point scale will be used:

- 4 - strongly agree
- 3 - agree
- 2 - disagree
- 1 - strongly disagree

The reviewer will also have an option “n/a” to indicate that the metric is not applicable to the deliverable-under-review. The reviewer is asked to leave the choice blank in case a question cannot be answered due to insufficient knowledge/lack of expertise (adding comments if possible).

4.2. Standard set of metrics for all deliverables to be reviewed

Some of the metrics that we will present refer to “MESH objectives”. The authors of the document-under-review should specify which specific objectives they want to have special attention paid to. The questions will be posed such that the reviewers can apply the numerical scores.

4.2.1. Scientific objectives

This deliverable:

- ◆ states its objectives, specific technical areas, related (sub)task(s) and dependencies, as specified in the Technical Annex
- ◆ meets the objectives as specified in the TA
- ◆ closely addresses the specific technical areas that the TA describes for this deliverable
- ◆ represents a suitable outcome for the resources applied to the (sub)task(s) originating the deliverable
- ◆ can be used by dependent deliverables as stated in the TA
- ◆ is suitable for use by its target audience (internal, EC, standards, public technical, public non-technical)
- ◆ is expected to have a high degree of success of intended impact (e.g. in standards, internal to the consortium etc)
- ◆ will lead to further outputs (such as papers, standards contributions etc.)
- ◆ represents a significant technical innovation
- ◆ significantly advances the state-of-the-art at the beginning of the project
- ◆ represents a large step compared with the state of the art

4.2.2. Metrics related to the document itself

This deliverable:

- ◆ is clearly written
- ◆ is concise

- ◆ is complete (there are no significant omissions)
- ◆ is technically correct
- ◆ is easy to read by different types of public
- ◆ is timely (it met its due date)
- ◆ contains a good executive summary such that the reader can understand what is contained in the document without necessarily having to read it in its entirety
- ◆ contains a clear and concise abstract
- ◆ contains suitable conclusions
- ◆ contains appropriate references

4.2.3. Other metrics

- ◆ The reviewer could read and adequately review the document within a reasonable time period
- ◆ The deliverable has been written to adequately target the right audience
- ◆ The documents describes what it is expected to be reported according to the TA description of the (sub)task(s)
- ◆ The document provides the right input to the intended users in MESH
- ◆ Free form questions that the reviewer can add

4.3. *Special metrics related to internal reviews only*

This deliverable:

- ◆ will be extensively used by other WPs in MESH
- ◆ represents a good opportunity for intellectual property generation (patents, copyright etc.)
- ◆ provides a significant degree of competitive advantage for the consortium
- ◆ solved some specific project issues
- ◆ raised new project issues which need to be addressed in other subtasks or deliverables

After the review we should make sure that the authors of the document are happy with the feedback from the review. This will not be part of the metrics themselves, as they are meant for the reviewers.

4.4. *Special metrics related to Advisory Board reviews only*

Here we will ask the Advisory Board reviewers to address special strategic issues, and additional questions relating to the external impact of the deliverable, its comparison to other similar work the reviewer may have seen, and some metric relating to how “world class” the work is.

Some possible statements that could guide the assessment via scored agreement are:

This deliverable:

- ◆ represents a significant step forward in its domain
- ◆ is one of the best I have reviewed in this technical area
- ◆ has strategic benefit within European research into this topic
- ◆ contains material suitable for world-class journals

4.5. Additional metrics specified by the editor of a deliverable

This allows the deliverable editor, authors, and/or their WP leader, to add metrics that would be relevant to certain special aspects e.g. well-known external benchmarks specific to the deliverable. An example might be MIPS and memory usage of the current standard for a certain codec compared to the one developed in MESH.

These metrics can be very specific for a workpackage and deliverable. For example, WP5 can have special questions to be answered, like the performance of the developed tools in a mobile environment. The WP leaders and/or document editor will be asked to formulate questions following the examples above at the time their deliverable is to be reviewed. This approach leaves the metrics specified in this document generic and easier to maintain.

5. Reviewer forms

The reviewers will be asked to fill in a form that contains all the metrics. Further, to get a better picture of the quality of the document, we do not ask the reviewer to put in a “yes” or a “no” but to give a score for each metric. This score will be 1-4, where 1 is the lowest and 4 is the highest. An extra column “n/a” has been added for the reviewer to indicate that she/he finds the metric not applicable to the document-under-review. The reviewer is asked to leave the choice blank in case a question cannot be answered due to insufficient knowledge/lack of expertise (adding comments if possible).

Following are two review forms: one for internal and one for external reviews.

MESH Internal review form						
Document no.						
Document title						
Reviewer name						
Date						
<p>Instructions: This review form lists a set of evaluation criteria concerning the document-under-review. Please score each metric and provide comments when necessary (1 - strongly disagree ; 2 – disagree; 3 – agree; 4 - strongly agree; n/a - the metric is not applicable for this document). If you feel that you cannot answer the question due to insufficient knowledge/lack of expertise, please leave the choice blank (adding comments if possible). In case different scores should apply to different parts of the document, please give an average score and provide comments.</p>						
Scientific objectives	1	2	3	4	n/a	Comments
<i>This deliverable:</i>						
states its objectives, specific technical areas, related (sub)task(s) and dependencies, as specified in the Technical Annex						
meets the objectives as specified in the TA						
closely addresses the specific technical areas that the TA describes for this deliverable						
represents a suitable outcome for the resources applied to the (sub)task(s) originating the deliverable						
can be used by dependent deliverables as stated in the TA						
is suitable for use by its target audience (internal, EC, standards, public technical, public non-technical)						
is expected to have a high degree of success of intended impact (e.g. in standards, internal to the consortium etc)						
will lead to further outputs (papers, standards contributions etc.)						
represents a significant technical innovation						
significantly advances the state-of-the-art at the beginning of the project						
represents a large step compared with the state of the art						

Metrics concerning document	1	2	3	4	n/a	Comments
<i>This deliverable:</i>						
is clearly written						
is concise						
is complete (there are no significant omissions)						
is technically correct						
is easy to read by different types of public						
is timely (it met its due date)						
contains a good executive summary such that the reader can understand what is contained in the document without necessarily having to read it in its entirety						
contains a clear and concise abstract						
contains suitable conclusions						
contains appropriate references						
Other metrics	1	2	3	4	n/a	Comments
The reviewer could read and adequately review the document within a reasonable time period						
The deliverable has been written to adequately target the right audience						
The documents describes what it is expected to be reported according to the TA description of the (sub)task(s)						
The document provides the right input to the intended users in MESH						
<i>Free form questions that the reviewer can add</i>						

Specific metrics for internal review	1	2	3	4	n/a	Comments
<i>This deliverable:</i>						
will be extensively used by other WPs in MESH						
represents a good opportunity for intellectual property generation (patents, copyright etc)						
provides a significant degree of competitive advantage for the consortium						
solved some specific project issues						
raised new project issues which need to be addressed in other subtasks or deliverables						
Metrics specified by the editor / author(s)	1	2	3	4	n/a	Comments
<i>Metric 1</i>						
<i>Metric 2</i>						
...						

MESH External review form						
Document no.						
Document title						
Reviewer name						
Date						
<p><i>Instructions: This review form lists a set of evaluation criteria concerning the document-under-review. Please score each metric and provide comments when necessary (1 - strongly disagree ; 2 – disagree; 3 – agree; 4 - strongly agree; n/a - the metric is not applicable for this document). If you feel that you cannot answer the question due to insufficient knowledge/lack of expertise, please leave the choice blank (adding comments if possible). In case different scores should apply to different parts of the document, please give an average score and provide comments.</i></p>						
Scientific objectives	1	2	3	4	n/a	Comments
<i>This deliverable:</i>						
states its objectives, specific technical areas, related (sub)task(s) and dependencies, as specified in the Technical Annex						
meets the objectives as specified in the TA						
closely addresses the specific technical areas that the TA describes for this deliverable						
represents a suitable outcome for the resources applied to the (sub)task(s) originating the deliverable						
can be used by dependent deliverables as stated in the TA						
is suitable for use by its target audience (internal, EC, standards, public technical, public non-technical)						
is expected to have a high degree of success of intended impact (e.g. in standards, internal to the consortium etc)						
will lead to further outputs (papers, standards contributions etc.)						
represents a significant technical innovation						
significantly advances the state-of-the-art at the beginning of the project						
represents a large step compared with the state of the art						

Metrics concerning document	1	2	3	4	n/a	Comments
<i>This deliverable:</i>						
is clearly written						
is concise						
is complete (there are no significant omissions)						
is technically correct						
is easy to read by different types of public						
is timely (it met its due date)						
contains a good executive summary such that the reader can understand what is contained in the document without necessarily having to read it in its entirety						
contains a clear and concise abstract						
contains suitable conclusions						
contains appropriate references						
Other metrics	1	2	3	4	n/a	Comments
The reviewer could read and adequately review the document within a reasonable time period						
The deliverable has been written to adequately target the right audience						
The documents describes what it is expected to be reported according to the TA description of the (sub)task(s)						
The document provides the right input to the intended users in MESH						
<i>Free form questions that the reviewer can add</i>						

Specific metrics for external review	1	2	3	4	n/a	Comments
<i>This deliverable:</i>						
represents a significant step forward in its domain						
is one of the best I have reviewed in this technical area						
has strategic benefit within European research into this topic						
contains material suitable for world-class journals						
Metrics specified by the editor / author(s)	1	2	3	4	n/a	Comments
<i>Metric 1</i>						
<i>Metric 2</i>						
...						

6. Schedule of expected reviews

The following table shows an indicative MESH internal and external reviews schedule. It defines the due month and type (internal/external) of scheduled reviews; the exact document/deliverable to be reviewed will be decided by the MC at the time of the review (1 month before the due date), depending on project progress. The actual review schedule will be maintained online at the MESH wiki.

Month	Review type	Deliverable to be reviewed
6	Internal	
12	External	
18	Internal	

7. Conclusions / recommendations

This document describes metrics that will be used for reviewing MESH documents. These reviews serve as input to the self-assessment in MESH. The metrics are defined both for the internal and for the external reviews for self-assessment. Further, we have described the review process itself.

The metrics described do not form a final set; we see it as a living document, to be updated when metrics appear to be not relevant, or relevant metrics are missing. The same follows for the Advisory Board and the schedule of reviews. As a consequence it should not be reviewed as a final deliverable.

8. References

- [1] Stan Rifkin, Charles Kox, “Measurement in Practice (CMU Software Engineering Institute)”, Carnegie Mellon University, report CMU/SEI-91-TR-16, 1991.
- [2] Watts S. Humphrey, “Managing the Software Process”, SEI Series in Software Engineering, 1989.
- [3] D.N. Card, R.L. Glass, “Measuring Software Design Quality”, Prentice Hall, 1990.
- [4] L. Rosenberg, L.E. Hyatt (NASA Software Assurance Technology Center) “Software Metrics Process for Risk Assessment”, 47th International Astronautical Congress & Exhibition, 29th Safety and Rescue Symposium, Risk management and Assessment Session, Beijing, China, October 1996.
- [5] Michael Fagan, “A History of Software Inspections”, SD&M Conference, Springer, 2001.
- [6] BUSMAN, “Bringing User Satisfaction to Media Access Networks”, IST-FP5, <http://busman.elec.qmul.ac.uk/>
- [7] aceMedia Integrated Project, “Integrating knowledge, semantics and content for user centred intelligent media services”, FP6-001765, <http://www.acemedia.org/>
- [8] aceMedia Deliverable D8.1, “Report on evaluation plan and evaluation criteria”, April 2004, http://www.acemedia.org/aceMedia/download.html?file_id=8
- [9] aceMedia Deliverable D8.4, “Report on evaluation plan and evaluation criteria”, June 2005, http://www.acemedia.org/aceMedia/download.html?file_id=20