Grant Agreement Number **ECP 2006 DILI 510049**

ENRICH

Methodology for Testing and Validation of e-Applications

Deliverable number	***D-7.1***
Dissemination level	*Public*
Delivery date	*30th May 2008*
Status	*Draft*
Author(s)	*Dr. Nerutė Kligienė, Institute of Mathematics and Informatics, LITHUANIA*

e*Content*plus

This project is funded under the *e*Content*plus* programme,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

Document Version Control

Version	Date	Change Made (and if appropriate reason for change)	Name and Institution of Commentator(s) or Author(s)
0.1 (draft)	February 2008	The main quality criteria and categories identified, illustrative examples presented	Institute of Mathematics and Informatics, Lithuania
1.0 (draft)	April 2008	Evaluation, Testing and Validation specifically for ENRICH WPs added instead of a general set of Criteria and sub criteria	Institute of Mathematics and Informatics, Lithuania, AiP Beroun, Czech Republic
1.1	May 2008	Evaluation, Testing and Validation specifically for ENRICH WP 3-6 revised and synchronized	Tomas Psohlavec, AiP Beroun, Czech Republic; Elsa Sklavounou-Zipstein, SYS, France
1.2	29/05/2008	Finalization, including remarks of partners and technical coordinator	Nerute Kligiene, IMI Jakub Heller, CCP

Document Review

Reviewer	Institution	Date and result of Review
Tomas Psohlavec	AiP Beroun, s r. o. Czech Republic	April 2008, revision of the second draft, specification of criteria for evaluation of results in project WPs.
Jakub Heller	Cross Czech a.s., Czech Republic	May 2008, revision of the third draft, comments on criteria for evaluation in ENRICH. Document approved for submission to EC.

Document Signature/Approval:

Before the table of content each document is to contain an approval signoff form.

Approved By (signature)	**Date**

Accepted by at European Commission (signature)	**Date**

1 Structure Summary

Work Package7 of ENRICH project is devoted to test and validate the tools, platforms and applications developed in the project in order to evaluate their accessibility, usability and adaptability. Existing research results and widely known quality publications from the last few years have been studied by the partners to select those basic principles and evaluation criteria which fulfil ENRICH project needs. The selected studies/research and their results are described in chapter 4 of this document and can be used for detailed reference.

A set of principles and quality criteria has been chosen and adapted for evaluation and testing of e-applications developed in the frame of ENRICH project. Section 3.1 of this document describes the quality criteria and sub-criteria in an example representing expected future estimation results. The developed matrix example helps to understand the general view of ENRICH evaluation methodology. Section 3.3 includes detailed introduction to the evaluation criteria adapted to the work being done in the project. The next section explains the relationship of the quality criteria and W3-WP6 items and how to convert the results of WP3-WP6 evaluation into terms of general quality criteria and categories. The most important part of Chapter 3 is section 3.6, which proposes the evaluation strategy including a detailed plan and partners' responsibilities in WP7.

The additional materials can be found at:

- *Annex 1*: System Usability Score currently applied to ENRICH project web site usability evaluation. Similar Questionnaire for monitoring the general interest of the end-user can be used later on when implementing the Task 7.2. The first results of evaluation are presented in this annex.

- The usability evaluation interactive form for evaluating the usability of project collaborative environment and the results of pilot evaluation already done at www.musicalia.lt/sus/

- Translation evaluation interface developed by SYSTRAN is accessible to all ENRICH users to evaluate the translation quality: http://enrich.systran.fr/enrich/feedbackForm.jsp?loc_lang=en

Partner responsible for WP 7 in ENRICH Consortium:
CR 8– **Institute of Mathematics and Informatics, Lithuania**

Contract Start Date: 01.12.2007. Project duration: 24 months

CONTENT

2 Introduction

The objective of ENRICH project is to create a base for the European digital library of cultural heritage, a real research environment built upon the existing Manuscriptorium platform adapted to the needs of organisations holding repositories of manuscripts. The *Manuscriptorium* Digital Library (www.manuscriptorium.com) represents manuscripts, rare old printed books, and other documentary heritage and is used as a working field of the project ENRICH.
ENRICH enables research activities in a particular type of cultural repositories – manuscripts – and follows a vertical approach to enable more types of cultural organizations to integrate their repositories also into TEL – The European Library. ENRICH will enable real seamless metadata and image data incorporation from dispersed resources onto a single platform with a uniform interface that will co-operate in real time with remote source data including various image banks stored in original locations of content owners. Principles based on multilingual ontologies provided by Systran will be evaluated and their use will be tested in the frame of the prototype running in parallel with the Manuscriptorium platform. This evaluation will lead to the choice of most suitable search technologies fulfilling the end-users' needs. This will enable end users to search for documents in their local languages and retrieve relevant data in all the source languages. As a first step, consortium partners developed the methodology for evaluation, testing and validation of the e-applications fulfilling the above.

The main objectives of this document are:

- To propose a methodology for evaluation of usability and adaptability of tools, platforms and applications developed in ENRICH project, based on the achievements of similar research done in the theory and practice in validating usability of digital spaces.

- To define detailed technical criteria and measures for the evaluation of e-applications developed in the framework of ENRICH, which will be described in the Evaluation report at the end of the project.

As described in the DoW and D1.1 (3.3 Quality Organisation; 4. Assurance of technical quality, its monitoring and evaluation) there are many diverse and interdependent tasks to be resolved and implemented in the project. The features and attributes of the results of these tasks – tools, applications and processes – will be more precisely defined as ENRICH will proceed. The particular features of each object of evaluation will be defined taking into account the end-users reviews performed during the project, their subsequent analyses, their implementation, and also the implementation of preceding interdependent task results.

This is why this document defines methodology for evaluation of the project as a whole, selects tasks suitable for evaluation and proposes evaluation criteria for the tasks and subtasks selected.

The detailed content of proposed criteria and sub-criteria of the methodology defined here will be later designed for each task in a separate subsequent document produced under Task 7.2. The detailed documentation will reflect the level of development and knowledge gained in the project and will include specific survey proposals and questions for task and subtask evaluation.

The evaluation itself will be performed based on the detailed documentation as soon as each task and subtask produces concrete results – so an instant feedback will be provided to the Task and WP leaders and enabling them, in case it is needed, to take correction steps.

The process of creating of the detailed documentation will run as follows:

- The necessary information on subject of evaluation and questions of the survey will be designed in cooperation of particular task leader and the WP7 leader for each task where an evaluationr is equired.
- The task leader will propose particular content and form of the survey.
- The WP7 leader will ensure it is in conformance with methodology defined here.
- Subsequently the survey must be discussed with participating partners and agreed corrections will be implemented.
- Final form of the survey will be deployed and gathering of required responses will be ensured by common work of participating partners.
- The processing of the gathered information according to the described evaluation methodology will undertake the WP7 leader.

In case there are or will be specialised tools available to perform particular task evaluation then these tools will be used in order to increase the efficiency of evaluation.
Following the lines of the Description of Work, the main duties in evaluation, testing and validation are described in Work package 7.

WP-7 Work package Description

Work package number :	**WP - 7**	**Start date:**	**3th month**	**End date:**	**23th month**
Work package title:	**EVALUATION, TESTING AND VALIDATION**				

Objectives

- Evaluation of usability and adaptability of tools, platforms and applications developed while implementing the ENRICH project, such as personalization for contributors and users, multilingual and user friendly access;
- To test the possibilities of application of modern tools for automated translation tools for multilingual search engine over existing data and metadata and the new data sets in order to fix possible shortcomings and improve the results before ending the project.

Description of work

Work package leader: IMI

Task 7.1 – Defining evaluation strategy
Task leader: IMI
Task participants: NKP, AIP, OUCS, KU, BNCF, MICF, VUL, SYS, ULW, SAM, UZK, DSP, NULI, BNE, BUTE, PSNC

The basic principles and evaluation criteria developed by worldwide known teams (such as W3C, Minerva Technical Guidelines) will be studied and adopted for evaluation and testing of e-applications planed to be developed in the frame of the project ENRICH. The respective weighting of criteria for evaluation of the results should be proposed, indicating more and less important categories. Describing positives and negatives of the applications usable for improving the working methods. Evaluating the platform, allowing testing of newly processed data (i.e. automated translation) and their usefulness to new implemented or modified tools.

Task 7.2 – Testing and evaluating the accessibility, usability and adaptability of developed applications
Task leader: IMI
Task participants: NKP, AIP, OUCS, KU, BNCF, MICF, VUL, SYS, ULW, SAM, UZK, DSP, NULI, BNE, BUTE, PSNC

Evaluation step by step the technical aspects and usability of the system. Consortium will prepare and assess evaluation tests, following recognized usability procedures. The usability evaluation will cover the assessment of all aspects of the service and language groupings, and be carried out in partnership with the user partners. The evaluation results will be fed back to the technical partners (who will adjust the technology platform). The evaluation process will reflect all strong and weak points of the results derived in the frame of ENRICH project: interactivity, interoperability – developed system being able to share information across databases and other online entities. Tested if similar data models and metadata element sets are used for semantically similar items and concepts.

Results:

- All the e-applications developed in the project will be tested and evaluated on different content
- Problems and bugs of developed tools will be fixed and adjusted by technical partners
- All the strong and weak points of the results of ENRICH will be described in **the evaluation report**

(Inter-) Dependencies, milestones and expected result

Milestone 7.1. Tested and evaluated usability and adaptability of the e-applications (personalization for contributors and users, multilingual and user friendly access, 23rd month of the project work).

The WP depends on the results of all technological WPs – WP3, WP4, WP5 and WP6

Expected results – well tested system, contributor & user-friendly, accessible and usable in

wider European community, reflecting wider cultural requirements, ensuring the content to be perceivable, operable and understandable by the broadest possible range of users and compatible with their wide range of assistive technologies, now and in the future.

Deliverables

D7.1. **Methodology for Testing and Validation of e-Applications** (m5); responsible partner: IMI.

D7.2. The Evaluation Report (m23); responsible partner: IMI.

The five items will be evaluated using the Methodology for Testing and Validation of e-Applications created as D7.1. Those items are:

- **The usability of the collaborative environment in the ENRICH project;**
- **Migration Tool developed in WP 3;**
- **Tools for creation of virtual documents by researchers developed in WP 4;**
- **Possibilities for sharing of large data sets investigated and developed in WP 5;**
- **Personalized Translation Interface developed in WP 6.**

Because we are also aware of the stress given to the area of Preservation and Security in the field of digitisation of historical resources, a subsequent additional evaluation of the most important Preservation and Security aspects will be performed.

It will be prepared in cooperation with Content (and Associated) Partners as an added result of the WP5, where the partners' repositories and practices will have to be examined in a deep detail anyway.

This evaluation will add more value to the whole ENRICH project outcomes, it will provide a comparison and the necessary feedback to the project Content (and Associated) Partners and last but not least: it will comply with the important trends observed by the prominent European projects mentioned below and their results.

Quality evaluation of digital documents provided in Manuscriptorium

In addition to above listed project results to be tested and evaluated, evaluation of quality of digital documents provided in Manuscriptorium is proposed. During the ENRICH project, various digital documents are provided in Manuscriptorium. The documents come from many places, and are created under different local conditions, using various procedures. Very important criteria for user, is usability of image – the visual perception. The metadata, which carry the information about the original documents and also provide the connection to the digital images, are on same level of importance. In the long time perspective of the use of such a data, specification of other features of digital data, and also evaluation of the conditions of their creation is important.

When digitizing in order to create an access to cultural heritage, it is necessary to consider the following aspects:

- unique value of the document
- safety of the digitization process for the original
- quality of digital data (images,…)
- quality of metadata
- safety and reliability of data preservation
- usability, accessibility (formats, metadata)

When specifying these aspects, our goal is to ensure well-balanced digital research environment. The motive of the quality evaluation is not to promote any kind of comparison, or rivalry among the partners. For example it is possible, that the only, and therefore very important image of some document just exists, taken by built in camera in mobile phone, and therefore, this image is acceptable in Manuscriptorium.

Currently draft of methodical documents for quality evaluation are being prepared, other shall follow. These detailed methodologies will be prepared by the technological coordinator AIP and summarized in a document “Technical quality criteria”.

3 The Quality Criteria for Validating ENRICH Work

The evaluation of the ENRICH results is the core activity of the WP7. Existing research results and widely known quality publications from the last few years have been studied by the partners to select those basic principles and evaluation criteria which fulfil ENRICH project needs. The selected studies/research and their results are described in chapter 4 of this document and can be used for detailed reference. The summary of our findings - resulting in a form of a matrix of criteria proposed for the future estimation - is presented in this chapter.

3.1. User-Based Objectives. The Set of Possible Quality Criteria

According to the MINERVA Technical Guidelines (MINERVA, 2005) the important areas for consideration at least have to include:

Interoperability: It is important that content can be accessed seamlessly by users, across projects and across different funding programmes. It should be possible to discover and interact with content in consistent ways, to use content easily without special tools, and to manage it effectively.
Accessibility: It is important that materials are as accessible as possible and are made publicly available using open standards and non-proprietary formats. If material is to be a widely useful resource it will be necessary to consider support for multiple language communities and ensure accessibility for citizens with a range of disabilities.
Preservation: It is important to secure the long-term future of materials, so that the benefit of the investment is maximized, and the cultural record is maintained in its historical continuity and media diversity.
Security: In a network age it is important that the identity of content and projects (and, where required, of users) is established; that intellectual property rights and privacy are protected; and that the integrity and authenticity of resources can be determined. Failure to address these areas effectively may have serious consequences, resulting in the waste of resources by different parties:

- ***Users - the scholar, the citizen, the learner****:* They will waste time and effort as they cannot readily find or use what is most appropriate to their needs, because it is not described adequately, or it is delivered in a particular way, or it requires specialist tools to exploit, or it was not captured in a usable form.
- ***Information providers and managers****:* Their investment may be redundant and wasted as their resources fail to release their value in use, as their products reach a part only of the relevant audience, as they invest in non-standard or outmoded practices.
- ***Funding agencies****:* They have to pay for redundant, fragmented effort, for the unnecessary repetition of learning processes, for projects that operate less efficiently than they should and deploy techniques that are less than optimal, for content that fails to meet user needs or does not meet market requirements.
- ***Creators, authors***: Their legacy to the future may be lost.

ENRICH Consortium will prepare and run evaluation tests, following defined usability procedures. The usability evaluation will cover all aspects of the service and language groupings, and will be carried out in partnership with the user partners. The evaluation process has to reflect all strong and weak points of the results developed in ENRICH project:

interactivity, interoperability – so that the system is able to share information across databases and other online entities. The evaluation process will include the selection of the main criteria and sub-criteria and then their deployment. The respective weighting of criteria for evaluation of the results will be proposed, if needed, also indicating the most important categories. The evaluation procedure will contain also description of positive and negative features of the applications, which will be used to improve the working methods. The evaluation results will serve as feedback to the technical partners (helping to adjust the technology platform accordingly).

An extensive research of evaluation and validation methods was performed as a starting point for project work in the WP7. The different approaches applied worldwide in the last few years for quality evaluation were studied. The range of considered initiatives was wide including: *W3* Consortium initiative – Web Content Accessibility Guidelines; European *Minerva* quality principles for cultural websites; *UsabilityNet* providing the relevant international standards in usability; the *nestor* Catalogue of criteria for trusted digital repository evaluation and certification in Germany – to Open Archival Information System (OAIS) Reference Model, technical criteria applied to Open Access Repositories in New Zealand (OARiNZ), Open Polytechnic NZ technical evaluation of Learning Management System (LMS), technical evaluation of selected open access repositories in New Zealand. All of them have some common features concerning criteria and metrics used for evaluation. The number of principal criteria used varies from 6 to 14, with different scalability, sub criteria and technical focus. New Zealand initiatives were mostly oriented to the evaluation of Learning Management Systems and digital publishing spaces while the *nestor* Catalogue (2006) and DRAMBORA (2007) stress long term preservation aspects and trustworthiness of digital repository, *Minerva* is oriented to better representation of cultural websites, Web Content Accessibility Guidelines focus on general usability, operability in a specific context. The listed approaches are described in details in chapter 4 of this document.
But none of the mentioned methods or results is directly applicable to the e-Applications developed in the framework of ENRICH.

Based on the above listed research results, a new methodology was developed for ENRICH. The main idea was to select a reasonable number of background principles expressed in the Criteria Set, each Criterion having a number of sub criteria (their maximal number was fixed). The sub-criteria were classified into categories, reflecting the different tasks to be tested. Each category has a fixed position in a set of sub-criteria. For example, the first three criteria refer a digital object, the next three the software tools used, another three the processing properties and so on. If a smaller number of sub-criteria is sufficient in some category, the numeration in the next category should be started from the number fixed to that category. Such structured model allows to test and to evaluate against the same universal principles multifaceted items: separate digital objects, processes or tools**.** Similar approach will be used for modification of a Usability test to be done by users having a general interest.

A limited number – five criteria covering the most important aspects of digital repository functionality has been selected and is listed below shortly. If each criterion can have, say, maximum 12 sub-criteria, all together we can consider up to 60 features, reflecting various attributes of multifaceted objects under investigation. At the moment a tentative list of possible sub-criteria is proposed, having in mind future testing of the tools developed in WP3, WP4, WP5, WP6. Therefore, the categories of tools in Adaptability and Usability have a

maximal number of sub-criteria, equal to 9 and 11 correspondingly. Later on when a real content for testing will be available in ENRICH project, the list of sub-criteria has to be updated and improved. But the same structured model and categories will be kept. We have considered the following categories:

i. the separate items – digital **objects** for submission or delivery,
ii. the software **tools** used,
iii. **processing** of assessment, ease of deployment,
iv. User-friendliness and long term preservation properties in a level of whole **repository**.

The set of sub-criteria is classified into categories according to (i) – (iv) statements in each of the 5 main Criteria. The following structure is proposed to use in each Nth CRITERION (N – numerated from 1 to 5).

N. CRITERION (an explanation of proposed structure in each of CRITERIA)

N.1. sub criterion concerning a **digital object** – separate item
N.2. sub criterion concerning a **digital object** – separate item
N.3. sub criterion concerning a **digital object** – separate item
N.4. sub criterion concerning a separate **tool/ software** tested
N.5. sub criterion concerning a separate **tool/ software** tested
N.6. sub criterion concerning a separate **tool/ software** tested
N.7. sub criterion concerning **process/activities** tested
N.8. sub criterion concerning **process/activities** tested
N.9. sub criterion concerning **process/activities** tested
N.10. sub criterion concerning **whole repository**
N.11. sub criterion concerning **whole repository**
N.12. sub criterion concerning **whole repository**

The summary of the possible set of criteria with structured sub-criteria is described below, later on to be provided with metrics for quantities' evaluation. The list of used sub-criteria has to be considered only as a prototype of the real one, which has to be composed with new results developed in ENRICH project. The current sub-criteria are presented now for usability demonstration purposes of the methodology proposed. DR– an abbreviation for Digital Repository is used below.

THE SET OF QUALITY CRITERIA – AN EXAMPLE

1. INTEROPERABILITY (9 sub-criteria)

1.1. Objects and their relationships in the DR are uniquely and permanently identified. **1.2. Adequate metadata for formal and content-based description and identification of the digital objects ingest.** **1.3. Adequate metadata for structural description of the digital objects acquired in DR.**
1.4. OAI-PMH compliant (required)
1.7. The assignment of metadata to the objects is guaranteed at all times. **1.8. Metadata for documenting changes made on the digital objects.**
1.10. The DR acquires adequate metadata for technical description of the digital objects. **1.11. The DR acquires adequate metadata to record the corresponding usage rights and conditions.** **1.12. Compatibility of used metadata with other standard metadata systems.**

2. ADAPTABILITY (8 sub-criteria)

2.1. Ease of working on code base: add/ change digital object
2.4. Software and hardware tools - only common/basic software and hardware required **2.5. Possibility to add other media files (for example audio, video)** **2.6. Documentation of software (*M-tool, M-edit*) is sufficient**
2.7. Ability to customize look and feel - change the header, theme, footer in personalized area **2.8. Ease of publishing - inexperienced users of the repository can easily publish content** **2.9. Installation, dependencies and portability**
2.10. Separate repository and branding for each institution

3. USABILITY (9 sub-criteria)

3.1. Corresponds to Web Content Accessibility (W3) Guidelines (perceivable, operable, understandable, robust) **3.2. Scalable fonts and graphics**
3.4. Reliability of tools: maturity, fault tolerance, recoverability, availability **3.5. Cross-browser support, not dependent critically on Java Script**
3.7. Text only navigation support **3.8. Operable to users with disabilities** **3.9. Efficiency: time behaviour, resource utilization, accessibility of distributed content**
3.10. Transparent, accessible website **3.11. Effectively searchable repository.**

4. SECURITY (5 sub-criteria)

4.1. Ability to restrict access at repository item level (e.g. view metadata but not content)
4.7. The authentication used by the repository to authenticate user. **4.8. The main aspects of application security supported**
4.10. Backup strategies for safety of content

4.11. The infrastructure implements the security demands of the IT security system.

5. MULTILINGUALITY AND LOCALISATION (6 sub-criteria)

5.1. Unicode text editing and storage available
5.4. Ability to present multilingual object description **5.5. Description of software in several languages**
5.7. Multilingual user interface **5.8. Multilingual search ability**
5.10. Localized to relevant languages.

3.2. Metrics for Evaluation of Quality

The next step after defining the main criteria and sub-criteria was to choose the scale and the metrics for evaluation. Minerva Quality Principles (Quality, 2005) propose the Checklists to each of 10 proposed Quality Principles. The outcome of every question in the Checklist is: ***Yes / No / not available.*** Each Quality Principle has the list of Criteria, the Checklist and Practical tests, and different numbers of items used for investigation of properties in each part of Principle under investigation. The principles for each stage of Website life-cycle (9 stages determined: from planning to maintaining) have the rating from 3 – high priority, 2 – mid priority, 1 – low priority. Each Criterion has different number of questions. The 10 principles combined with the 9 life-cycle stages – all this make the results of evaluation too complicated and hardly perceivable.

Another approach was undertaken in the majority of the investigated Technical reports [LMS Evaluation (2004), OARiNZ, Technical Evaluation (2006)] where the scale 0 – 4 was applied for evaluation. Each of the selected criteria was given an importance rating to be used when evaluating the different digital Repository systems. Major criteria were also broken down into sub-criteria with each sub-criteria also having an importance rating. The importance rating range is 0 – 4, with 0 being the lowest and 4 being of the highest score.

Each sub-criteria was then rated using a range of 0 – 4, these ratings defined as:

0 – Failed or feature does not exist,
1 – Has poor support and/or it can be done but with significant effort,
2 – Fair support but needs modification to reach the desired level of support,
3 – Good support and needs a minimal amount of effort,
4 – Excellent support and meets the criteria out of the box, minimal effort.

DRAMBORA (2007) for risk impact score propose to use more complex scaling 0 – 6 from zero impact via: negligible (1), superficial (2), medium (3), high (4), considerable (5) and reaching the 6 – cataclysmic impact.

The System Usability Scale (SUS) has five degrees from „*Strongly disagree*“ to „*Strongly agree*“ which later arranged to have a range from 0 to 4. This is a mature questionnaire, developed by John Brooke. It is very robust and has been extensively used and adapted in usability testing since 1986. This is the most strongly recommended of all the public domain questionnaires, according the *UsabilityNet.org* experts. The questionnaire modified from this

one is added separately in the Annex 1: additionally the classification of sub-criteria according the categories used for the ENRICH purposes were made.

The importance rating for sub-criteria in range is 0 – 4, (with 0 being the lowest and 4 being of the highest importance) was used in OARiNZ, Technical Evaluation (2006). Importance Rating is a subtle question: who could assign objectively the correct rate from scale 0 – 4 to each of sub-criteria? Wrong ratings can distort critically a whole procedure, therefore this Methodology do not consider an additional importance rating.

The next step was to deploy the scale selected – to evaluate items in context of usage, involving the direct and indirect users. The different kinds of users from target audience can be identified as:

I. Content, information provider, manager
II. Technical staff, supporting personnel
III. Scholar, researcher in the historical documents area
IV. End-user with a general interest.

In reality the full list of sub-criteria can be hardly assessed for evaluation by each type of target user, therefore it is natural to tolerate omitting a few of them. For example, content or information provider is strong on metadata submission and repository maters, while technical personnel can focus on processing and interoperability of platforms. Evaluation is derived while calculating an average value over users – a sample mean of some sub criteria $\hat{\boldsymbol{e}} = 1/n_1 \sum(estimates)$, where n_1 is the sample size – the number of respondents. The average estimate $\hat{\boldsymbol{e}}$ will give us more reliable result than individual estimates of expert. It is recommended to fix responses in the focus groups (estimates by contributors, technical staff, and researches) separately. Evidently the more users will be involved in each focus group, the more reliable results can be expected. The expected sample size is around 50–70 (users involved in the evaluation). This quantity is satisfactory to get a reliable evaluation of any sub-criteria under study.

It seams unreasonable to use only summing the rates of sub-criteria inside each criteria because usually we have different numbers of them. One of the possibilities to compare results derived in different criteria or WPs is to compare them with a corresponding possible maximal value of item under investigation. This was done in the 3.4 section matrix, introducing the different coefficient to multiply results derived in rating of each separate WP. Then in every WP (in categories as well) the overall value of score obtained will have a range from **0 to 100** and results become comparable with each other. The interpretation of derived result, say the score 77, can be: it is the 77% of **possible maximum,** where the 23 relative units is missing to the "ideal" which is equal to 100**.** In mathematical statistics the Box-plot diagrams are used to compare the ordered data, where the lower and upper quartiles divide all range of data in 3 regions, corresponding to low (up to 25%), satisfactory (26 –75%), and high value (76 –100%). Then a score falling in the interval 0 – 25 means the result is low, 26 – 75 it is rather good (satisfactory), 76 – 100 it is very good – over the upper quartile.

As it was stated in the section 3.1, we agreed to provide each criterion with four characteristics reflecting sample means of categories – object, tool, process, and repository. Then it will be possible to compare average results in each criterion by categories as well.

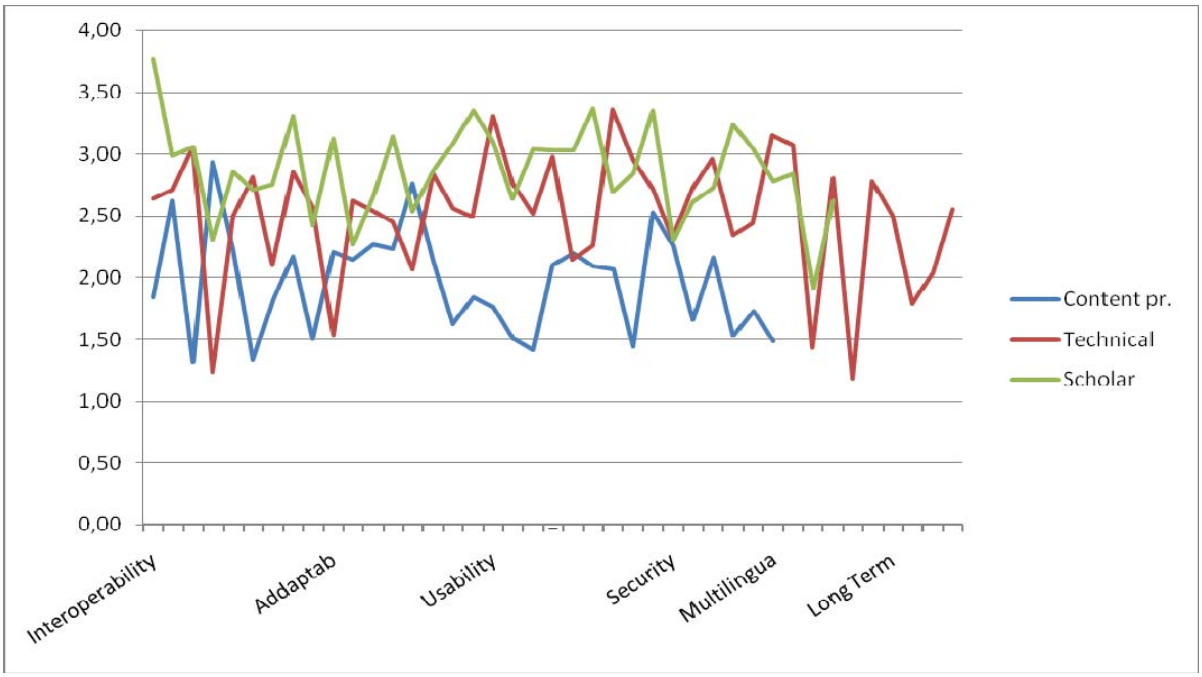


Fig.1. The example of evaluated average quality over the 42 sub-criteria belonging to 6 main criteria. The estimated averages provided by different groups of users are plotted. Rather chaotic behavior of evaluation values anyhow enables to extract the information of interest, see the section 3.8**.**

3.3. Evaluation, Testing and Validation Specifically for ENRICH WPs

The four technological work packages: WP3, WP4, WP5, WP6 - are closely related to each other and interdependent with evaluation, testing and validation. Keeping the same numeration of the Tasks as described in DoW under work packages, here we reformulate the items (assigning each by a triple number: for example the item 4.3.2 is #2 in the Task 4.3) which can be tested with the quality criteria described in section 3.1 and evaluated applying the metrics introduced in section3.2. Section 3.4 includes a table of interdependencies showing the relations of reformulated items with the main quality criteria/sub-criteria, and demonstrating the results of evaluation in a more general context of their conformity to widely accepted quality principles.

3.3.1. Standardization of Shared Metadata – WP 3

Objectives

To ensure interoperability of the metadata used to describe all the shared resources by analyzing the various standards used by different partners and ensuring their mapping to a single common format, which will be expressed in a way conformant with current standards.

Work package leader: OUCS

The aims of WP 7 – to test and evaluate how these objectives attained.

Task 3.1: Conversion between TEI P4 and TEI P5 platforms for description of manuscripts (m2-m15)

3.1.1. Development of migration tools on the basis of the sample data sets and validation with respect to:

a) Interoperability/Tools

b) Adaptability/Tools,

c) Usability/Tools.

Task 3.2: Implementation of OAI-PMH harvester into Manuscriptorium (m8-m18)

3.2.1. Ensured OAI/PMH metadata available for harvesting from the resources (where OAI interface available) managed within Manuscriptorium concerning Interoperability/Tools

3.2.2. Developed appropriate software tools to perform harvesting, evaluate the Usability/Tools.

Task 3.4: Improvement and generalization of UNICODE treatment in Manuscriptorium (m8-m23)

3.4.1. Validation of TEI recommendations for the representation of non-standard scripts, handling of Unicode and non-Unicode data concerning:

a) Adaptability/Tools an evaluation of representation (display) of the non-standard script samples,

b) Usability/Tools.

Description of Evaluation Activities, Expected Outputs

- As soon as the Tasks T 3.1 – T 3.4 will be accomplished, the partners will be asked to evaluate the results using the rating from **0 to 4** to each of the above enumerated items from 3.1.1 to 3.4.1 (including (a) - (c) aspects where they are available).
- All together the 7 items for evaluating the results in WP3 are in use, each of them can be classified (explained in 3.4) as reflecting one or more of the main Criteria: Interoperability, Adaptability, and Usability with their corresponding categories.
- The each item's score contribution will range from **0 to 4** and the maximal possible score for the WP3 is 7 x 4 = 28. If we prefer to have a result easily comparable with others, let us multiply the sum of the scores by 3,57 (i.e. 100/28) and the overall value of score obtained will have a range of **0 to 100**.
- The WP score falling in the interval 0 – 25 means that the result is low, 26 – 75 it is rather good (satisfactory), 76 – 100 it is very good.
- The visualization of the achieved results in respect of the main principles of quality such as Interoperability, Adaptability, Usability and their categories will be done by adding these results to the evaluation of other WPs.
- **The result – developed Migration Tools evaluated and validated as it was foreseen in the WP 7.**

3.3.2. User Personalization – WP 4

Objectives

Enable to subdivide the contents of Manuscriptorium into thematic collections. To satisfy the needs of all Manuscriptorium end-users, thematic collections will be created and maintained by authorized experts.

Furthermore, end-users will be able to construct their own individual collections and virtual documents via usage of newly developed tools – this will create opportunities to build individual user virtual libraries according to their personal needs (such as study,

teaching etc.). The tools will enable to decompose the digitized documents into necessary chunks/analytical digital objects and recompose them in new virtual documents following special teaching or learning goals, e.g. showing all illuminations from one scriptorium in a virtual document in spite of the fact that they are from various originals owned by different institutions in different countries.

Work package leader: MICF .

The aims of WP 7 – to test and evaluate how these objectives attained.

Task 4.1: Analysis of needs of typical end-users (mostly researchers and teachers) in view to implementation of common collections into Manuscriptorium (m3-m12)

4.1.1. Definition and **creation of basic set of tools** for creation and maintenance of common collections in Manuscriptorium; evaluate its impact as Usability/Effective tools

4.1.2. Implementation of the usual ways to work with common collections in Manuscriptorium; evaluate them concerning Adaptability/Processes – to evaluate implementation of the usual ways to work with common collections **as a process** more or less adaptable to a targeted user.

Task 4.2: Creation of individual collections for end-users (m3-m15)

4.2.1. Definition of requirements for the system for creation, maintenance and sharing of individual users' collections/libraries; Evaluate the usefulness of the system by filling SUS – the Questionnaire applicable to evaluate USABILITY of any System – in general expressing the feelings of users.

Task 4.3: Creation of virtual documents for research and teaching purposes (m6-m18)

4.3.1. Pilot implementation creating virtual documents for research and teaching purposes with existing technical resources – creation of sample virtual documents relating to the results of standardization (WP3) and publication of tools for free download; evaluate in respect of Adaptability /Tools.

4.3.2. Approving of possibility to import documents into special collections in Manuscriptorium; evaluate in respect of:

a) Adaptability to digital objects,

b) Security of processing.

Task 4.4: Evaluation of bibliographic searching capabilities of Manuscriptorium (m3-m21)

4.4.1. Implementation of tracking of searching behaviour of Manuscriptorium users to gather information for analysis and continuous improvement of research tools for better Usability/Tools.

Task 4.5: Implementation of deep searching possibilities above all metadata and textual data: the quality enhancement of search opportunities both above metadata and full texts. (m12-m23)

4.5.1. The access to the digital content through intelligent operators, performing a semantic search of metadata; evaluate in respect of:

a) Adaptability for search tools,

b) Usability for search tools.

Description of Evaluation Activities, Expected Outputs

- As soon as the Tasks T 4.1 – T 4.5 will be accomplished, the partners will be asked to evaluate the results using the rating from **0 to 4** to each of the above enumerated items from 4.1.1 to 4.5.1 (including (a) - (b) aspects where they are available).
- All together the 12 items for evaluating the results in WP 4 are in use, each of them can be classified (explained in 3.4) as reflecting one or more of the main Criteria: Interoperability, Adaptability, Usability Security, and their categories.
- The each item's score contribution will range from **0 to 4** and the maximal possible score for the WP 4 is 12 x 4 = 48. If we prefer to have a result easily comparable with others, let us multiply the sum of the scores by 2,08 (i.e. 100/48) and the overall value of score obtained will have a range of **0 to 100**.
- The WP score falling in the interval 0 – 25 means that the result is low, 26 – 75 it is rather good (satisfactory), 76 – 100 it is very good.
- The visualization of the achieved results in respect of the main principles of quality such as Interoperability, Adaptability, Usability, and their categories will be done by adding these results to the evaluation of other WPs.
- **The result – tools for creation of virtual documents by researchers evaluated and validated as it is foreseen in the WP 7.**

3.3.3. Personalization for Contributors – WP 5

Objectives

Development and implementation of the next generation of tools for structuring of existing metadata and newly created digitized documents and their implementation into Manuscriptorium structures, adjusted according to the results of WP3, accompanied by the use of large external data sets.

Providing on-line tools for verification, authentication and implementation of metadata and data into Manuscriptorium while respecting the specific characteristics of original data structures.

Allowing the accessibility of partners' data via Manuscriptorium without the need of changing their original presentation and use.
Work package leader: AIP.
The aims of WP 7 – to test and evaluate how these objectives attained.

Task 5.1: On-line tools for structuring of existing metadata and data related to manuscripts (m3-m16)

5.1.1. Set of tools will be based on existing tool provided in the frame of Manuscriptorium; evaluate including on-line versions. Tools to be used by the content partners for structuring of existing data, are they well usable? Evaluate in respect of:

a) Adaptability/Tools,

b) Usability/Objects, Usability/Tools.

Task 5.2: Analysis and development of tools for use of large external data sets (m3-m11)

5.2.1. To produce the evaluation of partner's repository and its objects and processes based on the analyses made in T5.2, evaluate concerning:

a) Adaptability/Tools,

b) Adaptability/Repository.

5.2.3. Establishing basic requirements for the tools for modification of digitized manuscripts into MASTER structure updated according to WP3 results extended by the use of large contributors' datasets; evaluate concerning Interoperability/Tools.

Task 5.3: Pilot implementation of large data sets provided by the selected partners into Manuscriptorium structure (m6-m16)

5.3.1. Development and validation of tools for large-scale zero-loss conversion of external data including the use of large data sets of pilot provider; evaluate in respect of:

a) Adaptability/Tools,

b) Usability/Processes.

5.3.2. Validation of routine data actualization in respect of:

a) Adaptability/Object,

b) Usability/Object.

Task 5.4: Integration of external data into Manuscriptorium (m10-m24)

5.4.1. Integration of external data into Manuscriptorium: evaluate with respect of:

a) Interoperability/Objects,

b) Adaptability/Tools,

c) Usability/Tools.

Description of Evaluation Activities, Expected Outputs

- As soon as the Tasks T 5.1 – T 5.4 will be accomplished, the partners will be asked to evaluate the results using the rating from **0 to 4** to each of the above enumerated steps from 5.1.1 to 5.4.1(including (a) and (c) aspects where they are available).
- All together the 13 items for evaluating the results in WP 5 are in use, each of them can be classified (explained in 3.4) as reflecting one or more of the main Criteria: Interoperability, Adaptability, Usability and their categories.
- The each item's score contribution will range from **0 to 4** and the maximal possible score for the WP 5 is 13 x 4 = 52. If we prefer to have a result easily comparable with others, let us multiply the sum of the scores by 1,92 (i.e. 100/52) and the overall value of score obtained will have a range of **0 to 100**.
- The WP score falling in the interval 0 – 25 means that the result is low, 26 – 75 it is rather good (satisfactory), 76 – 100 it is very good.
- The visualization of the achieved results in respect of the main principles of quality such as Interoperability, Adaptability, Usability and their categories will be done by adding these results to the evaluation of other WPs.

- **The result – basic conditions for sharing of large data sets evaluated and validated as it is foreseen in the WP 7.**

3.3.4. Multilingual and User Friendly Sophisticated Access – WP 6

Objectives

This work package aims at integrating a multilingual module via a user friendly sophisticated access: multilingual search application, multilingual forums, and multilingual ontology editor.

Based on SYSTRAN's machine translation technology, this module will provide also terminology extraction and machine translation customization tools for the construction and retrieval of personalized metadata within the aim to create new multilingual digital documents and multilingual ontologies in Czech, Polish, Spanish, Portuguese, German, Italian, English, French, Danish, Hungarian, Russian and Serbo-Croatian.

Work package leader: SYS

The aims of WP 7 – to test and to evaluate how these objectives attained.

Task 6.1: Multilingual access development (m0-m12)

6.1.1. Multilingual access via the API integration in the data retrieval interface associated or independent of a multilingual search. Evaluate in respect of:

a) Multilinguality in object level,

b) Multilinguality in processing.

6.1.2. Multilingual access dedicated translation interface where ENRICH expert users can fine-tune dynamically the machine translation tools thanks to adapted linguistic tools for terminology extraction and translation post-editing and customization. Evaluate in respect of:

a) Multilinguality / Tools,

b) Multilinguality /Repository.

For translation evaluation purposes SYSTRAN developed a translation evaluation interface which is accessible to all Enrich users so as to evaluate the translation quality. Details can be found at:

http://enrich.systran.fr/enrich/feedbackForm.jsp?loc_lang=en

SYSTRAN Enterprise Server 6 enables users to submit feedback on translation quality, and thus raise issues and propose alternative translations. Once submitted, this feedback can be

reviewed by a Dictionary Manager, who will thereafter be responsible for following its life cycle.

The Dictionary Manager determines the validity of submitted feedback. They can then quickly update the linguistic resources with proposed terminology entries and/or forward the feedback to SYSTRAN.

Following the submission, the feedback is entered into a database and an email is sent to the sender if the "Send me" notifications check-box was ticked at submission. The Search Feedback utility can be granted to users by a system administrator. Once in place, users access this tool via a Search Feedback command present in the left-hand Feedback menu. The Search Feedback page will display, offering a wide variety of search parameters.

The Feedback Search Results page will display, offering a list of feedback matches, as well as a number of tools that can be used to take action on the results selected.

A number of different actions can be performed on a feedback received, including:

- Exporting the selected feedback items
- Replay the translation of the selected feedback items
- Sending the feedback to SYSTRAN support
- Providing additional comments
- Changing the feedback status

All modifications will occur following submittal. This will enable WP6 to react effectively to all users' remarks.

Task 6.2: Translation Stylesheet design and use (m3-m24)

6.2.1. SYSTRAN Translation Stylesheets (STS) implementation exploiting metadata information; evaluate in respect of:

a) Usability/Object,

b) Multilinguality in object level,

6.2.2. Cross-language validation of STS, optimization of translation parameters; evaluate in respect of:

a) Interoperability/Tools,

b) Adaptability/Tools,

c) Usability/ Tools

d) Multilinguality of tools.

6.2.3. Metadata translation module implementation using a fully customized Translation Stylesheet provided by SYSTRAN; evaluate in respect of:

a) Adaptability/Process,

b) Usability/Process,

c) Multilinguality of process.

6.2.4. SYSTRAN Translation Stylesheets (STS) use XSLT to drive and control the machine translation of XML documents (native XML document formats or XML representations — such as XLIFF — of other kinds of document formats); evaluate in respect of:

a) Interoperability/Process,

b) Adaptability/Process,

c) Usability/Process

d) Multilinguality of process.

Remark. STS will provide a simple way to indicate which part of the document text is to be translated, and will enable the fine-tuning of translation, especially by using the structure of the document to help disambiguate natural language semantics and determine proper context. Thanks to STS machine translation is considered as part of the authoring and publishing process: source documents can be annotated with natural language mark-up produced by the author, a mark-up which will be processed by STS to improve the quality of translation, the gateway to the automatic publishing of a multilingual website from a monolingual (annotated) source. The mechanism is implemented through XSLT extension functions for consulting and for setting linguistics options in the translation engine. SYSTRAN will deliver this xslt file in order to fine-tune the system according to the ENRICH xml data elements.

Task 6.3: VICODI implementation (m6-m24)

6.3.1. Definition and homogenization of initial ontology applicable for this project. Evaluate with respect of Usability/object and multilinguality of object

6.3.2. Specification of user-friendly web-interface for visualization of multilingual ontology – special interface for modification. Evaluate with respect of Usability/Tools

Remark. Based on previous experience in the visualization and contextualization of digital content (IST project VICODI) SYSTRAN technology has been implemented for the construction of multilingual ontologies. The Research Center for Information Technologies (FZI) constructed multilingual ontologies available under GNU Free Documentation License (FDL) thanks to the EU-funded IST project Vicodi (http://www.vicodi.org/). Enrich will implement and use VICODI ontologies for the contextualization of the digital content.

Description of Evaluation Activities, Expected Outputs

- As soon as the Tasks T 6.1 – T 6.3 will be accomplished, the partners will be asked to evaluate the results using the rating from **0 to 4** to each of the above enumerated steps from 6.1.1 to 6.3.2 (including (a) - (d) aspects where they are available).
- All together the 20 items for evaluating the results in WP6 are in use, each of them can be classified (explained in 3.4) as reflecting one or more of the main Criteria: Interoperability, Adaptability, Usability, Multilingualism and others.
- The each item's score contribution will range from **0 to 4** and the maximal possible score for the WP 6 is 20 x 4 = 80. If we prefer to have a result easily comparable with others, let us multiply the sum of the scores by 1,25 (i.e. 100/80) and the overall value of score obtained will have a range of **0 to 100**.
- The WP score falling in the interval 0 – 25 means that the result is low, 26 – 75 it is rather good (satisfactory), 76 – 100 it is very good.
- The developed translation evaluation software and interface enables all ENRICH users to evaluate the translation quality. The visualization of the achieved results in respect of the main principles of quality such as Interoperability, Adaptability, Usability and their categories will be done by adding these results to the evaluation of other WPs.
- **The result – Personalized Translation Interface evaluated and validated as it is foreseen in the WP 7.**

3.4. Relationships of WP 3 – WP 6 Items and Quality Criteria

PRINCIPLE	Interoperability				Adaptability				Usability				Security				Multilinguality			
Categories: / Items	Object	Tools	Process	Repository	Object	Tools	Process	Repository	Object	Tools	Process	Repository	Object	Tools	Process	Repository	Object	Tools	Process	Repository
3.1.1		■				■				■										
3.2.1		■																		
3.2.2										■										
3.4.1						■				■										
Total in WP 3	7 items to evaluate, the coefficient to multiply = 3,57																			
4.1.1										■										
4.1.2							■													
4.2.1									SUS											
4.3.1						■														
4.3.2					■										■					
4.4.1										■										
4.5.1						■				■										
Total in WP 4	12 items to evaluate, the coefficient to multiply = 2,08																			
5.1.1						■			■	■										
5.2.1						■		■												
5.2.3		■																		
5.3.1						■					■									
5.3.2					■				■											
5.4.1	■					■				■										
Total in WP 5	13 items to evaluate, the coefficient to multiply = 1,92																			
6.1.1																	■		■	
6.1.2																		■		■
6.2.1									■								■			
6.2.2		■				■				■								■		
6.2.3							■				■								■	
6.2.4			■				■				■								■	
6.3.1									■								■			
6.3.2										■										
Total in WP 6	20 items to evaluate, the coefficient to multiply = 1,25																			
Total in each category	1	4	1	0	2	9	3	1	5	11	4	1	0	0	1	0	3	2	3	1
Total principle	**6**				**15**				**21**				**1**				**9**			
% of Total	11,32				28,30				41,51				1,89				16,98			
TOTAL	**52** items to evaluate as results in WP 3, WP 4, WP 5, and WP 6																			
Total in categories	**Objects: 11; Tools: 26; Process: 11; Repository: 4**																			

This matrix demonstrates evidently the relations of the Tasks accomplished during ENRICH work with general quality principles and categories. Each evaluation of an item will range from 0 to 4 without an additional importance rating, which often is too subjective in our opinion. The average of estimates will be used in the context of the target groups of users. Using the different coefficient to multiply in each WP, the score derived in each WP will range from 0 to 100 and it can be interpreted as a percentage of maximum value theoretically possible in this WP.

3.5. System Usability Score Applied to ENRICH Project Web Site

Evaluation of quality of e-Applications developed in the Framework of ENRICH is not simple as underlined in the sections3.1– 3.4. It can be performed mostly by experts and validated only when the corresponding e-Application/service/tool will be accessible to users.
An alternative simple method can be used already now to evaluate usability of collaborative environment in ENRICH project web site. http://enrich.manuscriptorium.com. The questionnaire, developed by John Brooke has been modified and applied to evaluate usability of ENRICH project digital space. The System Usability Scale (SUS) has five degrees from „*Strongly disagree*" to „*Strongly agree*" this non-numerical scale later has been arranged to have a range from 0 to 4. It is a very robust method and has been extensively used by many organisations. According the *UsabilityNet.org*, of all the public domain questionnaires, this is the most strongly recommended. The modified questionnaire is added separately in the Annex 1: additionally the classification of sub-criteria according the categories used for ENRICH were made.
The pilot version of the questionnaire has been published online www.musicalia.lt/sus/ and partners were asked first to fill it in when accessing the project web site.

The users were asked to tick the appropriate box indicating that they have previous experiences in the area as:

- Content provider, information manager ☐
- Technical personnel, supporting staff ☐
- Scholar, researcher in historical documents ☐
- End-user having general interests ☐

Later on this classification will enable to investigate the usability aspects in context of separate target groups.

Using System Usability Scale (SUS)
The system usability scale is generally used after the respondent has had an opportunity to use the system being evaluated, but before any debriefing or discussion takes place. Respondents should be asked to record their immediate response to each item, rather than thinking about items for a long time.
SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own. Statistical inference on the system usability will be made only after the statistical data collected and processed.
Using SUS all items should be checked. If a respondent feels that he/she cannot respond to a particular item, they should mark the centre point of the scale. In the alternative, the more complicated evaluation, developed in 3.3, the respondent has possibility to answer questions selectively, according the areas of his best knowledge.

The First Results and a Pilot Evaluation
From the first 34 respondents evaluating the quality of ENRICH working environment – answering 10 questions provided in the form online www.musicalia.lt/sus/ (the results received up to 13 May 2008), we can conclude that in average different kind of users (content providers, information managers, technical personnel, supporting staff, scholars, researchers in historical documents, end-users having general interests) expressed rather similar opinion about usability of this web site (Fig.2). Processing was ranked highest by all users and the

tools received the lower evaluation (Fig.3). Surprisingly, all kinds of users have rather evenly distributed an average score of usability of this site (Fig.4).

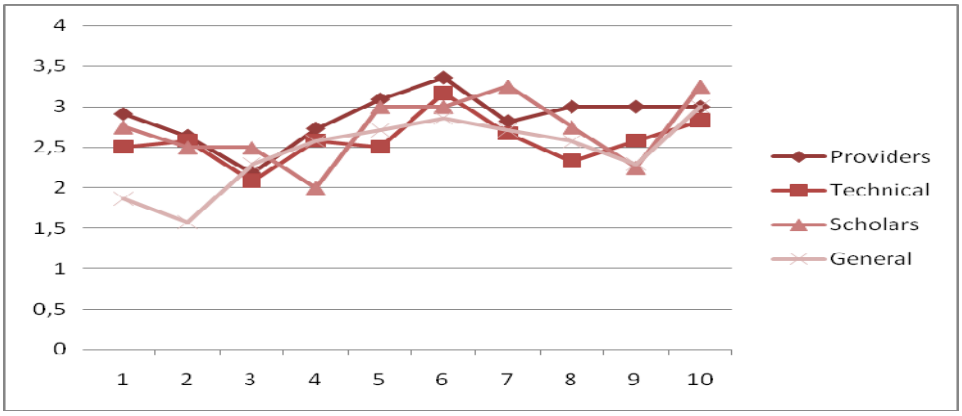


Fig.2. The average evaluation of 10 questions by different kinds of users.

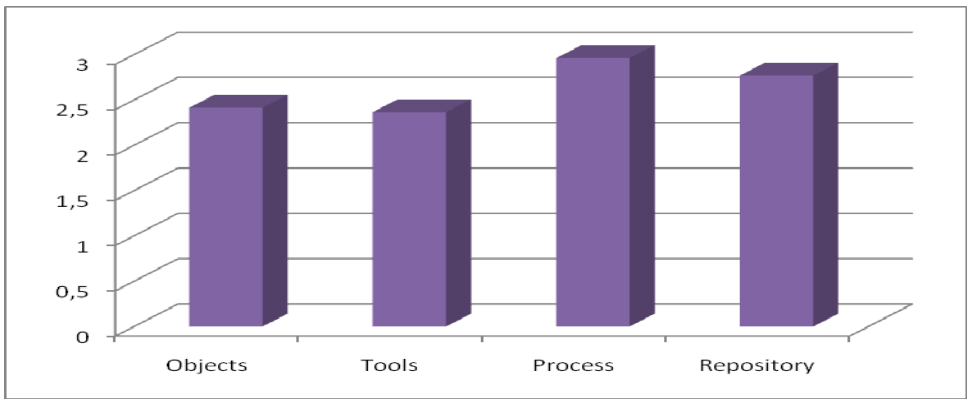


Fig.3. The average of points assigned by respondents to the categories.

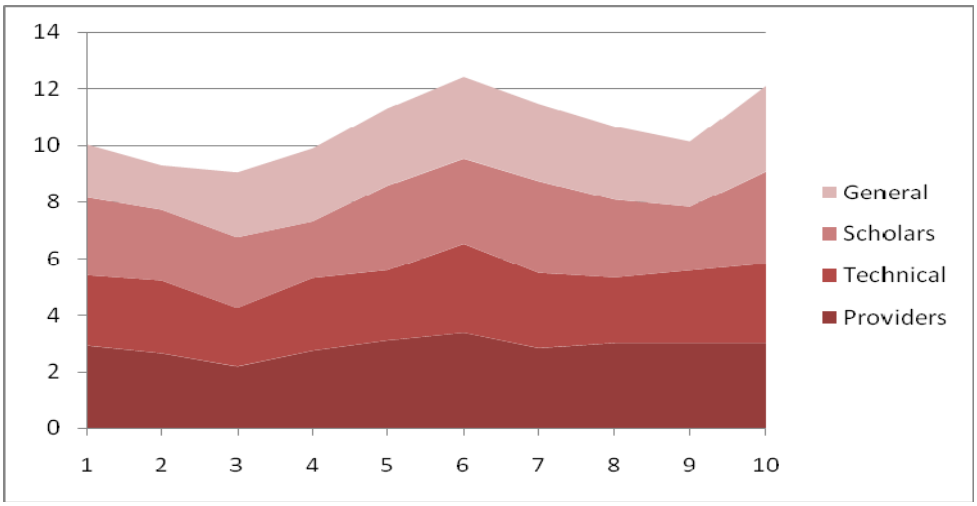


Fig.4. All kinds of users had rather similar opinions on usability of the ENRICH web site.

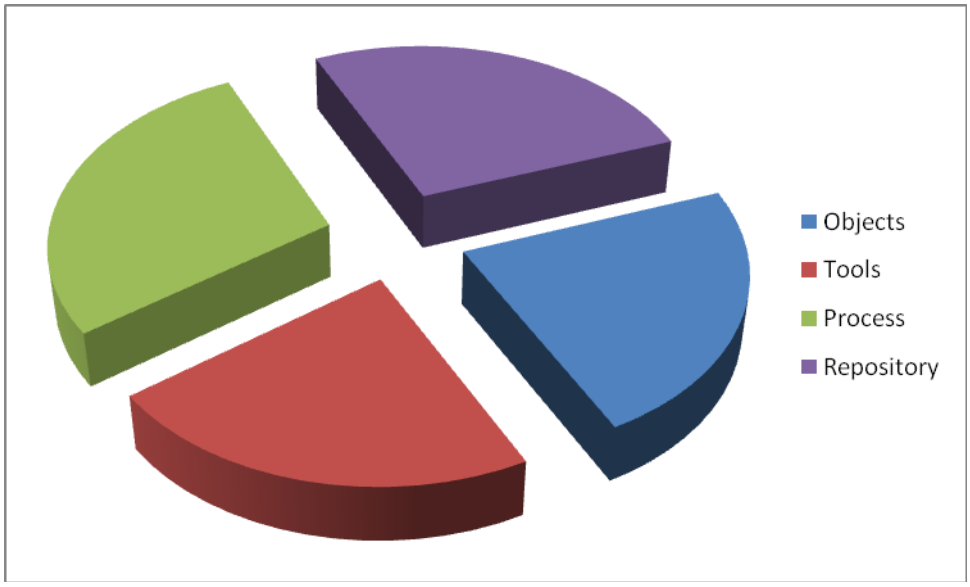


Fig.5. The usability evaluated by respondents reached value 10,51 (from the maximum possible 16 score). The components of averaged categories have rather equal impacts to the total as seen from the numerical values provided below.

Objects	2,41375
Tools	2,36625
Process	2,96125
Repository	2,76875
Usability	10,5100

Using the agreed orientation to the maximal possible value, the standard **Usability** score can be evaluated by (10,51/16) 100 = 65,69 i.e. it is falling into interval 26 – 75, that is rather good. By standardized scores of categories we would have the following scores:

Objects	60,28
Tools	59,16
Process	74,03
Repository	69,22

The conclusion from the first and not large sample of respondents is positive: the usability of web site as collaboration environment of partners **is rather good** (the score 65). The same is applicable to all categories: objects, tools, process, and repository. Although the difference in scores of categories less than 3 points are not significant due to moderate sample size (equal to 34) – this inference come from more sophisticated statistical investigation of sample means properties. Therefore the difference between objects and tools is negligible, but it is significant in a case of process and repository. This pilot evaluation was made in order to demonstrate that proposed method works on a system investigated and we are able to derive a numerical evaluation and interpret it correctly.

3.6. Organisation of Evaluation, Testing and Validation Activities

Work package WP-7 leader: IMI
Task 7.1 – Defining Evaluation Strategy (the start Feb 2008, the end Apr 2008)
Task 7.2 – Testing and evaluating the accessibility, usability and adaptability of developed applications

Usability Validation activities and Tasks **leader – the IMI** is responsible for maintaining coherence of usability validation activities. The steps defined as the sub-tasks of the **Task 7.1**

Sub-task 1	**Output**
Research and study the latest results published in the area of quality evaluation and validation: concepts, principles, methods.	The overview of this investigation is given in the chapter 4 of this document

Sub-task 2	**Output**
Adapting findings for evaluation and testing of e-applications to be developed in the frame of the project ENRICH.	Developed and described in the sections 3.1-3.7.

Sub-task 3	**Output www.musicalia.lt/sus/**
Developing and implementing a simple SUS (System Usability Scale) in order to make pilot trials of evaluation for project website usability at first steps.	The Questionnaire online with 10 simple questions proposed. The algorithm leading to the rate 0–4 applied. It deals with the evaluation of Usability (where the four categories identified). The first results of evaluation described in section 3.5.

Partners who are resources for Usability Validation in the project **(**all except CCP**)**

Task participants: NKP, AIP, OUCS, KU, BNCF, MICF, VUL, SYS, ULW, SAM, UZK, DSP, NULI, BNE, BUTE, PSNC
The identified all partners are responsible for maintaining coherence of usability validation activities in the following:

- Expressing own opinion
  - The rating 0 - 4 for the evaluation.
  - The categorization done as: “object, tool, process, repository”
- All partners will supply the technical human factors know-how in order to improve the set of criteria and proposed metrics for validation
- Select a few samples of documents to be submitted to DR and do several row evaluation according the finally agreed criteria and metrics in context of the different kind of users:
    - Content provider

- Technical staff, supporting personnel
- Scholar, researcher in the historical documents area
- End-user with a general interest.

Partners who are strongly associated with Usability Validation activities

There are partners who are associated with the usability validation activities more than validation resources. WP7 depends on the results of all technological WPs: WP3, WP4, WP5 and WP6, therefore the leaders of those WPs: **OUCS, MICF, AIP, SYS are more involved.** Their tasks are:

- Problems and bugs of developed tools must be fixed and adjusted by technical partners
- Reviewers of usability validation reports
- Consumers of the reports.

Output: Readiness of all partners to the **Task 7.2 – Testing and Evaluating the Accessibility, Usability and Adaptability of Developed Applications.** Start in August 2008.

3.7. Remarks on Timetable, Milestones

Recommendation: it is important that this document is revised at least once a year during the project lifetime, and when a series of usability validations has been already carried out.

The focus of validation at this stage

Usability validation by all partners is being conducted at this stage in order to be better prepared and effectively accomplish the **Task 7.2**

Quality criteria and metrics

It is expected that defining such metrics for evaluation will highlight the weak and strong points in respect of: (i) applied CRITERIA and sub-criteria; (ii) categories (object, tools, process, and repository); (iii) focus groups consisting of indirect and direct users. General user will have to evaluate usability by filling the questionnaire System Usability Scale (SUS). SUS is used for evaluating the usability of collaborative environment of project partners in ENRICH website. Later in the project, more quantitative measurements will be possible. There should be a summary, overall evaluation of the application towards the end of the project, which will focus on the project achievements in terms of the satisfaction of the overall usability objectives outlined at the very beginning.

Test scenarios – to think about

Purpose: to describe the setting in which usability validation will take place. Will this be in a number of focus groups, for instance, or in a laboratory, or in semi- or totally natural environments? Where will the users be collected? What tasks will be carried out by the users?
Recommendation: The Context of Use (see BASELINE project) should always be used as a guide to ensure that users, tasks, and environments in the validations always accord with the intended final realities, and are not 'pretend' or 'simulated' scenarios of doubtful value.

3.8. The Examples of Possible Use of Proposed Structured Model

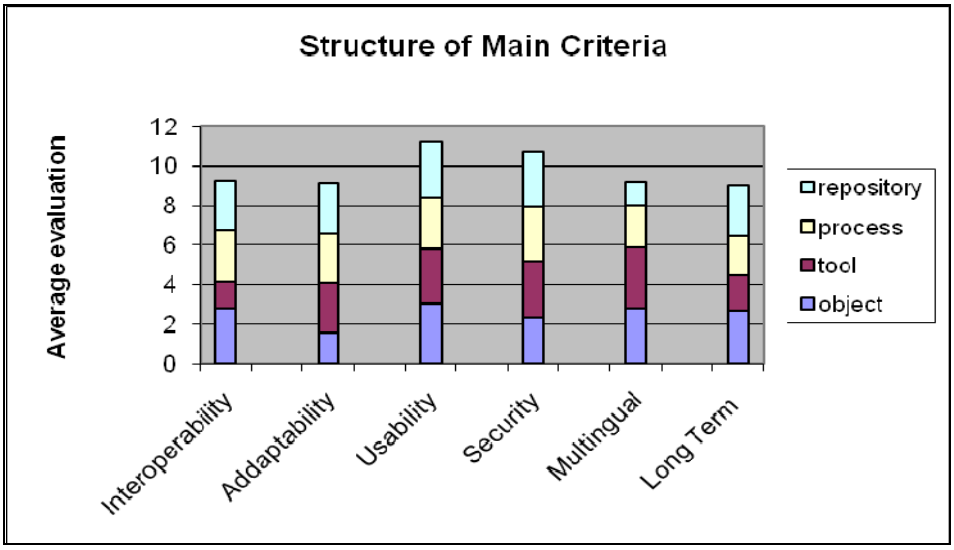


Fig.6. The average scores of the main Criteria structured by 4 categories.

This example illustrates how proposed structured model can be used for statistically based inference. Let us say we are investigating the quality of system using 6 main criteria and 42 sub-criteria belonging to them as it was in example Fig.1. The each sub criterion is asigned to one of the categories of interest: object, tool, process or repository. The rate of evaluation is from 0 to 4. Collecting opinions of some specified kind of users and calculating the sample mean as an arithmetical average of expresed opinions, we will have an estimated value for each sub criterion, having any value (mostly non-integer) from interval (0, 4). The results were shown in Fig. 1. The average value in every category has to be defined then because we have different numbers of sub-criteria in categories. In this example the highest position has the Usability with almost equal average values in every of 4 categories.
Another example shows a possibility to visualize the target groups of users against the same quality criteria and a defined number of sub-criteria. This time focus groups have less even opinions on criteria as it was in the Fig.4. over the 10 questions asked on the project web site.

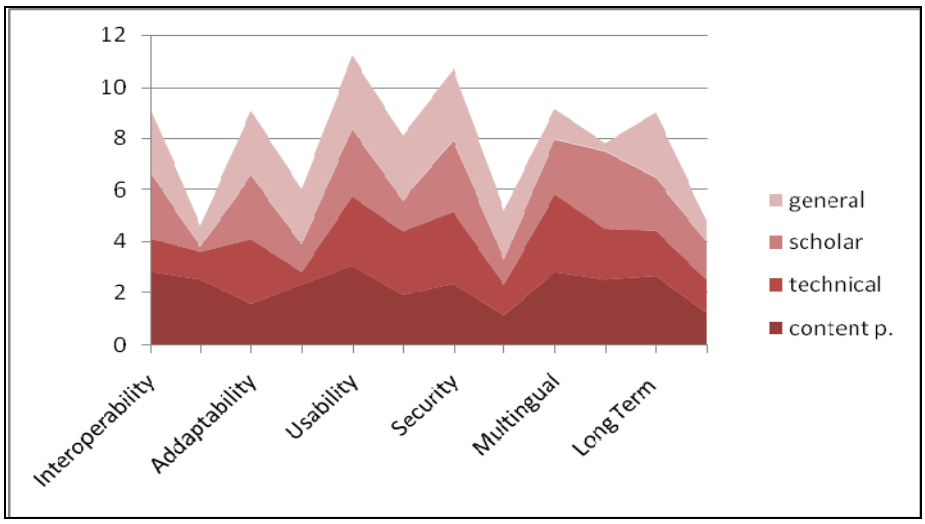


Fig.7. The possible average scores of the main Criteria versus different users groups.

4 Overview of the Quality Evaluation Criteria and Metrics: Categories and Principles Used World Wide

This chapter explores the different approaches applied during the last few years for quality evaluation: *nestor* Catalogue of criteria (The *nestor*, 2006) for trusted digital repository evaluation and certification in Germany, technical criteria applied to Open Access Repositories in New Zealand (OARiNZ), Open Polytechnic NZ technical evaluation of Learning Management System (LMS, 2004), technical evaluation of selected open access repositories in New Zealand (Technical evaluation, 2006).
Open Archival Information System (OAIS) Reference Model, Minerva quality principles for cultural websites (Quality, 2005), ***UsabilityNet*** providing relevant international standards in usability, W3 Consortium initiative – Web Content Accessibility Guidelines. All of them have a lot in common concerning criteria used for evaluation.

4.1. The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification

The ***nestor*** – Network of Expertise for Long-Term Storage and Long-Term Availability of Digital Resources in Germany.

Based on the initial nestor survey and similar to the approach taken by the Certification of Trusted Repositories Task Force by the National Archives and the Research Libraries Group (RLG-NARA , 2006) , the ***nestor*** working group used abstract criteria in the main catalogue instead of asking very detailed and specific questions (e.g. which metadata is used). The nestor catalogue includes best practice values and provides examples and specific literature references for the listed criteria, despite the need to update such examples regularly. The intention is that this criteria catalogue, and its planned revisions, will help customers to share information and expectations. The criteria composed in this catalogue are seen as a sufficient set to demonstrate the trustworthiness of a digital long-term repository. The Catalogue prepared by Susanne Dobratz (Humboldt-University Berlin, University Library), Astrid Schoger **(**Bavarian State Library, Digital Library), Stefan Strathmann (Göttingen State and University Library) was published online (at TEXAS Digital Library, The ***nestor)*** and more complete version (The ***nestor*** - Network, 2006) issued June 2006 as draft for a public comment.

Overview of the Criteria

Within the following table the term "repository" is taken as abbreviation for "long-term digital repository", SIP – Submission Information Package, AIP – Archival Information Package, DIP – Dissemination Information Package.

A	Organizational Framework
1	**The repository has defined its goals.** 1.1 selection criteria, 1.2 responsibility for the long-term preservation of the information represented by the digital objects, 1.3 designated community

2	**The repository grants its designated community an adequate usage of the information represented by the digital objects.** 2.1 Access for the designated community, 2.2 interpretability of the digital objects by the designated community
3	**Legal and contractual rules are being observed.** 3.1 legal contracts between producers and the repository, 3.2 repository operates on a legal basis regarding archiving, 3.3 repository operates on a legal basis regarding usage
4	**The organizational form is adequate for the digital repository.** 4.1 adequate funding, 4.2 sufficient numbers of qualified staff, 4.3 organizational structure, 4.4 repository engages in long-term planning, 4.5 continuation of preservation tasks even beyond the existence of the repository
5	**Adequate quality management is conducted.** 5.1 definition of processes and responsibilities, 5.2 documentation of elements and processes, 5.3 reaction to substantial changes
B	**Object Management**
6	**The repository ensures integrity of digital objects during all processing stages:** 6.1 ingest, 6.2 archival storage, 6.3 access
7	**The repository ensures authenticity of digital objects during all processing stages:** 6.1 ingest, 6.2 archival storage, 6.3 access
8	**The repository has a strategic plan for its technical preservation measures.**
9	**The repository accepts digital objects from its producers based on defined criteria.** 9.1 specification of SIPs, (Submission Information Package) 9.2 identification of relevant features of the digital objects for the information preservation, 9.3 technical control over its digital objects in order to execute preservation methods
10	**The archival storage of the digital objects is undertaken to defined specifications.** 10.1 definition of AIPs, (Archival Information Package) 10.2 transformation of the SIPs into AIPs, 10.3 storage and readability of the AIPs, 10.4 implementation of preservation strategies for AIPs
11	**The repository permits usage of the digital objects based on defined criteria** 11.1 definition of DIPs, (Dissemination Information Package) 11.2 transformation of AIPs into DIPs
12	**The data management system is capable of providing the necessary digital repository function.** 12.1. persistent identification of objects and their relations, 12.2. metadata for content and formal description and identification of the digital objects, 12.3 metadata for structural description of the digital objects, 12.4 metadata for documenting changes made on the digital objects, 12.5 metadata for the technical description of the digital objects,

	12.6 metadata for the usage rights and terms of the digital objects, 12.7 The assignment of metadata to the digital objects is guaranteed every time
C	**Infrastructure and Security**
13	**The IT infrastructure is adequate** 13.1 The IT infrastructure implements the demands from the object management, 13.2 The IT infrastructure implements the security demands of the object management
14	**The IT infrastructure implements the object management demands.**

4.2. OARiNZ Technical Criteria for New Zealand's Digital Repositories

The CPIT-led Open Access Repositories in New Zealand (OARiNZ) project, funded under the Government's eLearning Collaborative Development Fund (eCDF) aims to design and build the infrastructure necessary to connect all of New Zealand's digital research repositories that meet standards for interoperability and access.

4.2.1. High-level evaluation criteria for institutional repositories

Overall, there is consensus on some high-level criteria that institutional repositories should meet. Based on the work of Genoni (2004), Johnson (2002), and Lynch (2003), Jones, Andrew, and MacColl (2006, pp. 53-4) summarise these, and add defining criteria as follows:

- institutionally defined:
  - no extra-institutional issues to be resolved, as in subject repositories,
  - easily integrated into existing system, in terms of style, semantics and technology;
- scholarly (not necessarily publishable but of value to academics);
- cumulative and perpetual:
  - the collection is expanding
  - items are preserved in perpetuity;
- open and interoperable:
  - access to the collection,
  - its content freely available;
- digitally capture and preserve many aspects of campus life (dependant on decisions made by the institution on scope, discussed on the Implementation Considerations page, and types of material to be collected);
- search without constraints:
  - software employed able to answer queries from human and automated users,
  - effective user and web interface.

4.2.2. Technical Evaluation of Selected Learning Management System by Open Polytechnic of NZ

The goals and objectives of this document developed at Open Polytechnic of New Zealand are

- To define detailed technical criteria for the evaluation of Open Source Learning Management Systems.

- To gain understanding of the design, architecture and implementation details of the short-listed LMSs, with a view towards long-term development and maintenance.
- To evaluate the short-listed LMSs against set criteria.
- To engage members of the Open Source community in the process where relevant.

Methodology proposed consist of two steps:

(1) to develop technical evaluation criteria,
(2) to deploy and to evaluate.

Based on the Initial Evaluation of Open Source Technologies paper provided by Open Polytechnic of New Zealand, the criteria focuses and expands on the technical aspects of the systems. For each criteria rated, a discussion is provided, indicating what is covered, its importance, and a check-list of the aspects observed. Each overall criteria is presented with a brief discussion, with summary tables providing a weighted rating.

Major Evaluation Criteria

1. Overall architecture and implementation (joining 15 different sub-criteria)
2. Interoperability
3. Cost of ownership
4. Strength of the development community
5. Licensing
6. Internationalisation, localisations
7. Accessibility
8. Document transformation.

The first of the criteria - Overall architecture and implementation is expanded over the following detailed investigations: System is scalable, Scale down – minimum requirements, Scale up – high end servers, Scale out – allows efficient caching, Scale out: can use multiple application servers? Scale out: can use multiple database servers? Scalability constraints, System is modular and extensible, Multiple installations on a single platform, Has reasonable performance optimisations, Look and feel is configurable, Security, Modular authentication, System is robust and stable, Installation, dependencies and portability.

4.2.3. Technical Evaluation of Selected Open Access Repositories in New Zealand (Technical Evaluation, 2006)

To pay particular attention to the long-term development and maintenance lifespan of the short-listed Repositories OARiNZ commissioned a report on the Technical Evaluation of Selected Open Source Repository Solutions **(**Technical Evaluation, 2006). Major criteria in this report, developed from the Evaluation of Open Source Technologies paper include:

1. Scalability
2. Ease of working on code-base, extensibility
3. Security
4. Interoperability (ability to integrate with other repositories - OAI-PMH compliance, and ease of integration with systems such as Learning Management Systems)
5. Ease of deployment, ability to support multiple installations on a single platform (required for hosting facility)
6. Ease of system administration (ability to configure for different uses)
7. Internationalisation - multiple language interfaces

8. Open source (type of license)
9. Quality and configurability of workflow tools
10. Strength of community

Developing the Technical Evaluation Criteria, Sub-criteria and Metrics

Major technical evaluation criteria were drafted and reviewed by Steering Committee members and are rather compatible with the Criteria enumerated in the 2.2 section. Moreover each selected criteria was given an importance rating to be used when evaluating the different Repository systems. Major criteria were also broken down into sub-criteria with each sub-criteria also having an importance rating. The importance rating range is 0-4, with 0 being the lowest and 4 being of the highest importance.

Each sub-criteria was then rated using a range of 0-4, these ratings defined as:

0 – Failed or feature does not exist.
1 – Has poor support and/or it can be done but with significant effort.
2 – Fair support but needs modification to reach the desired level of support.
3 – Good support and needs a minimal amount of effort.
4 – Excellent support and meets the criteria out of the box, minimal effort.

Let us consider in detail some of sub-criteria, used in this document.

1. SCALABILITY

1.1 Scale Up – Ability for the Repository to scale higher by adding more resources

1.2 Scale out – The repository supports caching, adding more instances, and other mechanisms to scale higher.

1.3 Architecture - The repository be separated into different local parts and put into different machines. (E.g. separate the database, data directory, components from the repository to distribute to different machines) **<...>**

3. SECURITY

3.1 Data Encryption - Supports encryption of data while transmitting the content, such has using SSL/https.

3.2 Server Security - What does the repository require for installation? Does it follow good security practices e. g. proper file permissions, secure database connection?

3.3 Authentication - The authentication used by the repository to authenticate user

3.4 Authorisation/Access Rights - Support for different roles to properly manage the content and administer the system.

3.5 Ability to restrict access at repository item level (eg view metadata but not content).

Application security: **authentication**

Is the authentication process and authentication token handling secure?

Standard security practices focus on the handling of authentication credentials, and subsequent tokens to prevent replay attacks. HTTP-AUTH is the authentication extension of the HTTP standards. However, HTTP-AUTH is limited, not very secure, and does not offer a good user experience due to bad implementations in popular web browsers. Web applications do well in avoiding it (GARF2001).

Instead, the LMSs use a strategy known as *cookie-based authentication*: the user sends username and password through a form, and is issued an authentication token – stored in a cookie – to use in subsequent requests.

The three systems apply standard industry practices to authentication, and defer token generation and tracking unto PHP's native session management. This session management facility is capable of very good and secure token management. However it is configured by default with settings that make it open to replay attacks. In particular, the hashing algorithm needs a seed (which can be a shared secret in the case of multiple application servers); but

this is not enabled by default. It is strongly recommended PHP session management is configured securely in public deployments of any of the reviewed LMS.

Application security: authorization

After the user has been correctly authenticated, authorization mechanisms decide what the user is allowed to do.

The authorization systems should define available access rights, roles (as a set of access rights) and, where relevant, ownership relationships and groups. They may also define "access control lists" (ACLs), but this is not a requirement. Finally, roles should be configurable, not hard-coded (with the exception of the "super-user" account).

Application security: logging and monitoring

The application framework performs the core of the logging and monitoring for web applications. In this regard, Apache provides excellent logging and monitoring facilities. Certain attacks, however, can only be monitored and logged at the application level. The systems are rated on their handling of:

- Dictionary attack against a single account.
- Attempt to access an object without access rights to it.
- Attempt to edit an object without edit rights to it.

Application security: Validation of input.

Does the system have strong validation of input? Does it effectively prevent users from abusing the system? Software applications, as many other man-made systems, are vulnerable to users providing unexpected input. Systems that can be used anonymously must be hardened to validate all input from users.

In the case of web applications, lack of input validation renders applications vulnerable to HTML injection and SQL injection attacks. HTML injection can lead to an attack against system users known as Cross Site Scripting (XSS). File upload forms also constitute a vulnerability point. Storage – and later retrieval – of files uploaded by users must have strict validation and sanitization of file names/paths to avoid *shell injection* attacks. Successful attacks on file uploads can result on remote users executing commands as the webserver user.

The PHP language implementation has, in past versions, encouraged questionable programming practices by way of a feature known as register_globals. Recent versions of PHP have register_globals disabled, and applications do not require it are safer. Systems are rated on their handling of concrete SQL injection, XSS and shell injection attack attempts. We also consider unsafe practices such as use of register_globals.

The three systems tested have good escaping mechanisms against SQL injection attacks. They are far weaker in the face of HTML injection/XSS attack, in particular if coming from teacher or author account.

5. EASE OF DEPLOYMENT

5.1 Software and hardware requirements - The repository only requires common/basic software and hardware

5.2 Packaging and installation steps

5.3 Separate repository and branding for each institution (Essential)

6. SYSTEM ADMINISTRATION

6.1 Ability to customise look and feel - change the header, theme, footer

6.2 Ease of Publishing - Inexperienced users of the repository can easily publish a content

<…>The developed criteria and sub-criteria set and proposed rating of each criteria were applied to estimate three Learning Management Systems: *Eprints, Dspace* and *Fedora*. The one of such results in evaluation are presented in the Fig.3.1.

4.2.4. Open Archival Information System (OAIS) Reference Model

An archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

The OAIS (2007) Reference model can also be used as a basis for evaluating an open access archive system. Ball (2006) provides arguments why it is not fully appropriate as an evaluation tool, and describes the development of the RLG/NARA Checklist to better fulfil this purpose.

The **OAIS Reference Model** is a useful vocabulary for discussing the preservation of digital objects in a repository context. The issues discussed in OAIS (2007) may be helpful when considering: how to package together the component parts of the extended product model; how to structure a set of documents without altering them; and how to add other forms of metadata to a set of documents without altering them.

Very extensive (148 pages) and well documented (Recomendation, 2002) based on OAIS Reference Model for space data systems, was published and maintained by: CCSDS Secretariat, Program Integration Division (Code M-3), National Aeronautics and Space Administration, Washington, USA.

Hosted/Hub Solution

Criteria	Importance Rating	Eprints Evaluation	Eprints Total	Dspace Evaluation	Dspace Total	Fedora Evaluation	Fedora Total
4.1 Scalability							
4.1.1 Scale Up	3	3	9	2	6	3	9
4.1.2 Scale out	4	3	12	3	12	3	12
4.1.3 Architecture	3	3	9	2	6	4	12
Sub Total Scalability			30		24		33
4.2 Ease of working on Code Base							
4.2.1 Add/Change digital object	3	2	6	3	9	4	12
4.2.2 Documentation of code	3	4	12	2	6	4	12
Sub Total Ease of Working on Code Base			18		15		24
4.3 Security							
4.3.1 Data Encryption	3	4	12	4	12	4	12
4.3.2 Server Security	4	2	8	3	12	4	16
4.3.3 Authentication	2	3	6	4	8	4	8
4.3.4 Authorization/Access Rights	2	2	4	4	8	2	4
4.3.5 Ability to restrict access at repository item level	3	2	6	3	9	4	12
Sub Total Security			36		49		52
4.4 Interoperability							
4.4.1 OAI-PMH Compliant (Required)	4	4	16	4	16	4	16
4.4.2 SOAP, UDDI	3	0	0	4	12	4	12
4.4.3 SRU/SRW	3	0	0	4	12	0	0
4.4.4 Bulk Import and Export	4	2	8	4	16	4	16
4.4.5 Institution exit mechanism (Required)	4	3	12	4	16	4	16
4.4.6 Authentication - (e.g. LDAP)	2	3	6	4	8	4	8
4.4.7 Standard metadata - Dublin core, METS.	4	3	12	4	16	4	16
Sub Total Interoperability			54		96		84
4.5 Ease of Deployment							
4.5.1 Software and hardware requirements	2	3	6	2	4	3	6
4.5.2 Packaging and installation steps	3	2	6	2	6	2	6
4.5.3 Separate repository/branding each institution	4	4	16	4	16	4	16
Sub Total Ease of Deployment			28		26		28
4.6 System Administration							
4.6.1 Ability to customise look and feel	2	3	6	2	4	3	6
4.6.2 Ease of Publishing	3	4	12	3	9	3	9
Sub Total System Administration			18		13		15
4.7 Internationalisation							
4.7.1 Localisable UI	4	4	16	4	16	3	12
4.7.2 Unicode Text editing and storage	4	4	16	4	16	4	16
Sub Total			32		32		28
4.8 Open Source							
4.8.1 Open Source License	3	4	12	4	12	4	12
4.8.2 Defined roadmap for the future	3	4	12	2	6	4	12
Sub Total Open Source			24		18		24
4.9 Work Flow Tools							
4.9.1 Workflow integration	3	1	3	3	9	4	12
4.9.2 Support for different workflows	2	1	2	3	6	4	8
Sub Total Work Flow Tools			5		15		20
4.10Community Knowledge Base							
4.10.1 Quality of information on the product's web site	4	4	16	3	12	4	16
4.10.2 Size/Level of activity in the developer community	4	2	8	3	12	2	8
4.10.3 Size of and level of activity in the user community	4	2	8	3	12	2	8
4.10.4 Availability/Range of communication channels	3	3	9	4	12	3	9
4.10.5 Software release history-sustainability,vitality	3	3	9	2	6	3	9
4.10.6 Documentation on how to set up and manage a repository farm (one code base, many independent repositories)	3	0	0	0	0	0	0
Sub Total Community Knowledge Base			50		54		50
Totals for Hub & Hosted Repositories			295		342		358

Fig.8. The results of estimating the LMS: *Eprints*, *Dspace*, and *Fedora*

4.3. Minerva Quality Principles for Cultural Websites and Digital Repositories

A quality cultural website celebrates European cultural diversity by providing access for all to digital cultural content. A good quality cultural website according (Quality, 2005) must:

- be **transparent**, clearly stating the identity and purpose of the website, as well as the organisation responsible for its management
- select, digitise, author, present and validate content to create an **effective** website for users
- implement quality of service policy guidelines to ensure that the website is **maintained** and updated at an appropriate level
- be **accessible** to all users, irrespective of the technology they use or their disabilities, including navigation, content, and interactive elements
- be **user-centred**, taking into account the needs of users, ensuring relevance and ease of use through responding to evaluation and feedback
- be **responsive**, enabling users to contact the site and receive an appropriate reply. Where appropriate, encourage questions, information sharing and discussions with and between users
- be aware of the importance of **multi-linguality** by providing a minimum level of access in more than one language
- be committed to being **interoperable** within cultural networks to enable users to easily locate the content and services that meet their needs
- be **managed** to respect legal issues such as IPR and privacy and clearly state the terms and conditions on which the website and its contents may be used
- adopt strategies and standards to ensure that the website and its content can be **preserved** for the long-term

The ten quality principles proposed in Minerva project are:

Transparent - Effective - Maintained - Accessible - User-centered - Responsive - Multi-lingual - Interoperable - Managed – Preserved.

Several of those criteria have been used also in the sets of criteria, proposed by German experts and OARiNZ activities.

Minerva Quality Principles (Quality, 2005) propose the Checklists to each of 10 proposed Quality Principles. Each Quality Principle has the list of Criteria, the Checklist and Practical tests, but different numbers of items used for investigation of properties in each part of Principle under investigation. The principles for each stage of Website life-cycle (9 stages determined: from planning to maintaining) have the rating from 3 - high priority, 2- mid priority, 1 – low priority. The outcome of every question in the Checklist is: **Yes/ No/ n.a.** Each Criterion has different number of questions. Combined with (10 principles) x (9 life-cycle stages) it make results complicated and hardly perceivable.

4.3.1. The Life Cycle Approach

The structure of this document reflects a 'life cycle' approach to the digitisation process, and (with some modifications) parallels the structure of the *Good Practice Handbook* developed within Work Package 6 of the Minerva project. The document is divided into the following main sections, each reflecting a stage in that life cycle. In practice, there are relationships and

dependencies between activities within these different stages and indeed some of the stages may not be strictly sequential.

1. Preparation for digitisation
2. Handling of originals
3. The digitisation process
4. Storage and preservation of the digital master material
5. Metadata capture
6. Publication Disclosure
7. Reuse and re-purposing
8. Intellectual property and copyright

4.3.2. Requirement Levels

The approaches taken to conformance to standards and guidelines vary between programmes, along a spectrum from encouraging the adoption of good practice to mandating conformance to standards as a condition of grant award. Typically the standards and guidelines adopted by programmes encompass different levels of requirement, and it is possible to distinguish between requirement levels. Standards, that are widely accepted and already in current use. Projects **must** implement standards that are identified as requirements. **Guidance** that represents good practice but for which there may be reasons not to treat it as an absolute requirement, for example, because those standards are still in development. Projects **should** maintain and demonstrate awareness of these standards.

4.4. Basics of Usability. How is Usability Defined

Usability means making products and systems easier to use, and matching them more closely to user needs and requirements.

The international standard, ISO 9241-11, provides guidance on usability and defines it as:

The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

Usability is about:

- Effectiveness - can users complete tasks, achieve goals with the product, i.e. do what they want to do?
- Efficiency - how much effort do users require to do this? (Often measured in time)
- Satisfaction – what do users think about the products ease of use?

…. which are affected by:

- The users - who is using the product? E.g. are they highly trained and experienced users, or novices?
- Their goals - what are the users trying to do with the product - does it support what they want to do with it?
- The usage situation (or 'context of use') - where and how is the product being used?

Usability should not be confused with 'functionality', however, as this is purely concerned with the functions and features of the product and has no bearing on whether users are able to use them or not. Increased functionality does not mean improved usability!

There are a series of international standards for usability and user centred design.

4.4.1. Relevant International Standards in Usability

Usability and user-centred design standards can be divided up into three main categories:

1. The product development process
2. The use of a product (how well users perform when using it and how satisfied they are when using it)
3. The design of the user interface and interaction

1. The main standard affecting the product development process is **ISO 13407: Human-centred design processes for interactive systems**. This standard provides guidance on human-centred design activities throughout the life cycle of interactive computer-based systems. It is a tool for those managing design processes and provides guidance on the human-centred approach.

The standard outlines four essential user-centred design activities:

- understand and specify the context of use
- specify the user and organisational requirements
- produce designs and prototypes
- carry out user-based assessment

The activities are carried out in an iterative fashion, with the cycle being repeated until the particular usability objectives have been attained. The recommended process (*UsabilityNet.org*) is shown below:

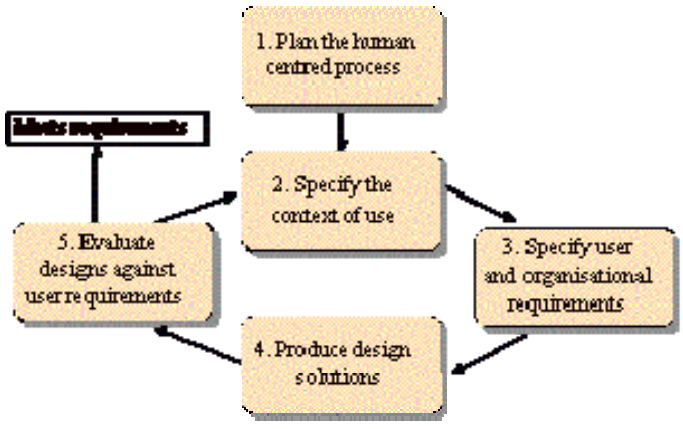


Fig.9. The process of achieving usability objectives.

UsabilityNet recommends the usability methods that should be used during this process.

2. The main standard dealing with product usage characteristics is ISO 9241**: Ergonomic requirements for office work with visual display terminals**, and part 11, which provides guidance on usability and the definition of usability:

Usability: the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

ISO 9241-11 explains how to identify the information that is necessary to take into account when specifying or evaluating usability in terms of measures of user performance and satisfaction. It includes an explanation of how the usability of a product can be specified and evaluated as part of a quality system, for example, one that conforms to ISO 9001. 3.

3. The main standard for design of the user interface and interaction is ISO 9241. Parts 12-17 provide detailed guidance on the design of user interfaces.

4.4.2. Categorisation of Standards Related to Usability

Standards related to usability can be categorised as primarily concerned with:

- the use of the product (effectiveness, efficiency and satisfaction in a particular context of use)
- the user interface and interaction
- the process used to develop the product
- the capability of an organisation to apply user centred design

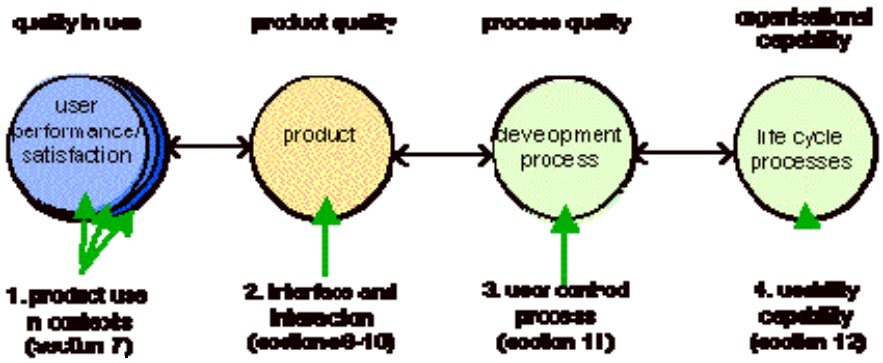


Fig. 10. Categorization of Standards related to usability.

The standards described here are divided into these categories, and are listed in the table below.

	Principles and recommendations	Specifications
Use in context	ISO/IEC 9126-1: Software Engineering - Product quality - Part 1: Quality model	ISO 20282: Usability of everyday products
	ISO/IEC TR 9126-4: Software Engineering - Product quality - Part 4: Quality in use metrics	
	ISO 9241-11: Guidance on Usability	
Interface and interaction	ISO/IEC TR 9126-2: Software Engineering	ISO 9241: Ergonomic requirements for

	- Product quality - Part 2 External metrics	office work with visual display terminals. Parts 3-9
	ISO/IEC TR 9126-3: Software Engineering - Product quality - Part 3 Internal metrics	ISO/IEC 10741-1: Dialogue interaction - Cursor control for text editing
	ISO 9241: Ergonomic requirements for office work with visual display terminals. Parts 10-17	ISO/IEC 11581: Icon symbols and functions
	ISO 11064: Ergonomic design of control centres	ISO 13406: Ergonomic requirements for work with visual displays based on flat panels
	ISO 14915: Software ergonomics for multimedia user interfaces	ISO/IEC 14754: Pen-based interfaces - Common Gestures for text editing with pen-based systems
	IEC TR 61997: Guidelines for the user interfaces in multimedia equipment for general purpose use	ISO/IEC 18021: Information Technology - User interface for mobile tools
		ISO 18789: Ergonomic requirements and measurement techniques for electronic visual displays
Documentation	ISO/IEC 18019: Guidelines for the design and preparation of software user documentation	ISO/IEC 15910: Software user documentation process
Development process	ISO 13407: Human-centred design processes for interactive systems	ISO/IEC 14598: Information Technology - Evaluation of Software Products
	ISO TR 16982: Usability methods supporting human centred design	
Capability	ISO TR 18529: Ergonomics of human-system interaction - Human-centred lifecycle process descriptions	

ISO/IEC 9126: Software product evaluation - Quality characteristics and guidelines for their use (1991)

In the software engineering community the term usability has been more narrowly associated with user interface design. ISO/IEC 9126, developed separately as software engineering standard, defined usability as one relatively independent contribution to software quality associated with the design and evaluation of the user interface and interaction:

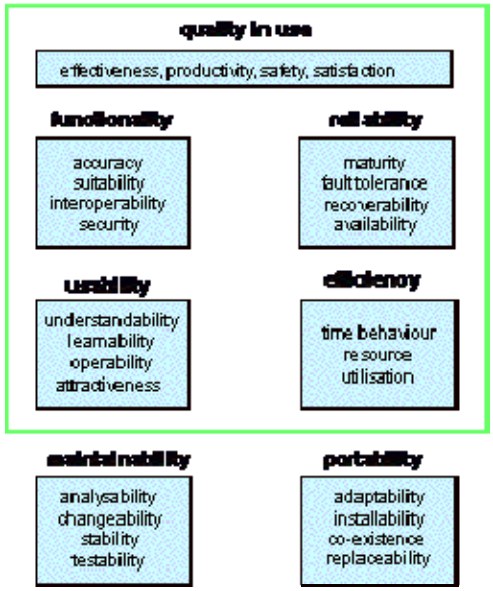
Usability: a set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.

4.4.3. Quality in Use – Product Use in Context

ISO/IEC FDIS 9126-1: Software Engineering - Product quality - Part 1: Quality model (2000)

ISO/IEC 9126 (1991) has recently been replaced by a new four part standard that has reconciled the two approaches to usability. ISO/IEC 9126-1 describes the same six categories of software quality that are relevant during product development: functionality, reliability, usability, efficiency, maintainability and portability.

The definition of usability is similar:



Usability: the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions.

The phrase "when used under specified conditions" (equivalent to "context of use" in ISO 9241-11) was added to make it clear that a product has no intrinsic usability, only a capability to be used in a particular context.

The standard now recognizes that usability plays two roles (Bevan 1999): a detailed software design activity (implied by the definition of usability), and an overall goal that the software meets user needs (similar to the ISO 9241-11 concept of usability). ISO/IEC 9126-1 uses the term "**quality in use**" for this broad objective:

Quality in use: the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use.

Quality in use is the combined effect of the six categories of software quality when the product is in use. The overall objective is to achieve quality in use, both for the end user and the support user. Functionality, reliability, efficiency and usability determine quality in use for an end user in a particular context.

The support user is concerned with the quality in use of maintenance and portability tasks.

Other parts of ISO/IEC 9126 define metrics for usability and quality in use.

ISO/IEC 9126: Software Engineering - Product Quality

ISO/IEC 9126-1 defines usability in terms of understandability, learnability, operability and attractiveness. Parts 2 and 3 include examples of metrics for these characteristics. These can be used to specify and evaluate detailed usability criteria.

ISO/IEC 9126-2 Part 2: External metrics (DTR: 2001)

This technical report describes metrics that can be used to specify or evaluate the behaviour of the software when operated by the user. For example: how long does it take to learn to use a function, can users undo functions, do users respond appropriately to error messages?

ISO/IEC 9126-**3** Part 3: Internal metrics (DTR: 2001)

This technical report describes metrics that can be used to create requirements that that describe static properties of the interface that can be evaluated by inspection without operating the software. For example: what proportion of the functions is documented, what proportion of functions can be undone, what proportion or error messages are self explanatory?

4.4.4. Measuring System Usability

ISO/IEC DTR 9126-4: Software Engineering - Product quality - Part 4: Quality in use metrics (2001)

This technical report contains examples of metrics for effectiveness, productivity, safety and satisfaction. Specifying usability requirements and verifying that they have been achieved in a usability test is an important component of user centred design (ISO 13407). ISO/IEC 9126-4 suggests metrics for effectiveness, productivity, satisfaction and safety that can be used for this purpose. The results can be documented using the Common Industry Format for usability test reports, which is included as an example in an Annex to ISO/IEC 9126-4. System Usability Scale (SUS) proposed in SUS (Brooke, 1986).

This is a mature questionnaire, developed by John Brooke in 1986 and not published until years later. It is very robust and has been extensively used and adapted. It is public domain and nobody has published any standardization data about it. Of all the public domain questionnaires, this is the most strongly recommended.

Using *SUS*
The SU scale is generally used after the respondent has had an opportunity to use the system being evaluated, but before any debriefing or discussion takes place. Respondents should be asked to record their immediate response to each item, rather than thinking about items for a long time.
All items should be checked. If a respondent feels that they cannot respond to a particular item, they should mark the centre point of the scale.

Scoring *SUS*
SUS yields a single number representing a composite measure of the overall usability of the system being studied. Note that scores for individual items are not meaningful on their own.

To calculate the SUS score, first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. For items 1, 3, 5, 7, and 9 the score contribution is

the scale position minus 1. For items 2, 4, 6, 8, and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

SUS scores have a range of 0 to 100.

4.5. W3 Consortium – Web Content Accessibility Guidelines 2.0

Web Content Accessibility Guidelines 2.0 (WCAG 2.0) covers a wide range of recommendations for making Web content more accessible. Following these guidelines will make content accessible to a wider range of people with disabilities, including blindness and low vision, deafness and hearing loss, learning disabilities, cognitive limitations, limited movement, speech difficulties, photosensitivity and combinations of these. Following these guidelines will also make your Web content more accessible to the vast majority of users, including some older users. These guidelines however are not able to address the needs of all people with disabilities.

WCAG 2.0 success criteria are written as testable statements that are not technology-specific. Guidance about satisfying the success criteria in specific technologies as well as general information about interpreting the success criteria are provided in separate documents. The main points in accessibility are:

Perceivable

- Provide text alternatives for any non-text content so that it can be changed into other forms people need such as large print, Braille, speech, symbols or simpler language
- Provide synchronized alternatives for multimedia
- Create content that can be presented in different ways (for example spoken aloud, simpler layout, etc.) without losing information or structure
- Make it easier for people with disabilities to see and hear content including separating foreground from background

Operable

- Make all functionality available from a keyboard
- Provide users with disabilities enough time to read and use content
- Do not create content that is known to cause seizures
- Provide ways to help users with disabilities navigate, find content and determine where they are

Understandable

- Make text content readable and understandable
- Make Web pages appear and operate in predictable ways
- Help users avoid and correct mistakes that do occur

Robust

Maximize compatibility with current and future user agents, including assistive technologies

4.6. Other Approaches to Evaluation

Xia and Sun (2007) suggest that ‘depositorship’ (whether the item is deposited by the author(s) or not) and ‘availability of full-text’ are key criteria in assessing repositories. An issue raised by Probets: Jenkins: 2006 that should be included in a set of evaluative criteria is

the quality of documentation related to an individual repository. They argue that a key criterion in any successful repository is that the "purpose and aims of the IR (should be) clearly defined and that the IR documentation itself should be concise and easy to understand, with the rights and responsibilities of stakeholders clearly presented."

Goh (2006) developed an alternative set of criteria for evaluating Open source software for repositories from a survey of the literature, and tested them on four software packages. Their criteria are:

- content management,
- content acquisition,
- metadata,
- search,
- access control and security,
- report and inquiry,
- preservation, interoperability,
- user interface, standards compliance,
- automatic tools and support.

They provide a full checklist of criteria assessing each of these categories, with weightings dependent on the critical nature of each.

DRAMBORA (Digital Repository Audit Method Based on Risk Assessment). It is the latest document evolved from the ***nestor*** project and RLG/NARA Task Force activities.

The Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE) 2007 announced the release of the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit. This toolkit is intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognize their strengths.

It is mostly based on identifying risks, assess and managing risks. For evaluating risk probability score and risk impact score the scale from 0 to 6 is proposed. Zero impacts are quantified as zero loss of digital object authenticity and understandability. Considerable impact, quantified as 5, results in widespread loss, including unrecoverable loss of digital object authenticity and understandability. The highest score rated by 6 is cataclysmic impact resulting in total and unrecoverable loss of digital object authenticity and understandability. It is very extensive document covering over 220 pages.

Digital repositories are still in their infancy and this model is designed to be responsive to the rapidly developing landscape. The development of the toolkit follows a concentrated period of repository pilot audits undertaken by the DCC, conducted at a diverse range of

organizations including national libraries, scientific data centers and cultural and heritage data archives. At the moment it seems to be not easy applicable toolkit.

5 References

Ball, A. (2006). Briefing Paper: the OAIS Reference Model. http://www.ukoln.ac.uk/projects/grand-challenge/papers/oaisBriefing.pdf, accessed on 8 January 2008.

BASELINE. Data for User Validation in Information Engineering (available supporting documents on User Validation Assistant, Usability Validation Plan format, Usability Validation Report format , Context of Use analysis)[accessed on 21 January 2008] http://www.ucc.ie/hfrg/baseline/index.html

Bevan N. (1999) Quality in Use: Meeting User Needs for Quality, *Journal of Systems and Software.*

Brooke, John (1986) SUS - A quick and dirty usability scale, [accessed on 6 January 2008] http://www.usabilitynet.org/trump/documents/Suschapt.doc

Digital Curation Center http://www.dcc.ac.uk/ accessed on 8 January 2008.

DPE – Digital Preservation Europe, http://www.digitalpreservationeurope.eu/ accessed on 8 January 2008.

DRAMBORA (Digital Repository Audit Method Based on Risk Assessment), (2007) http://www.repositoryaudit.eu/ accessed on 15 February 2008.

GARF(2003) Garfinkel, Spafford, Schwartz, Practical Unix & Internet Security, 3rd Edition,

Genoni, P. (2004). Content in Institutional Repositories: A Collection Management Issue. *Library Management. 28*(6/7), 300–306.

Goh et al. (2006). A Checklist for Evaluating Open Source Digital Library Software. *Online Information Review. 30* (4). 360–379.

Johnson, R. (2002). Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication. D-Lib Magazine. 8(11). http://www.dlib.org/dlib/november02/johnson/11johnson.html"2007095026′

Jones, R. (2004). D Space vs. ETD-db: Choosing software to manage electronic theses and dissertations. *Ariadne, 38*.

Jones, Andrew & MacColl (2006). *The Institutional Repository.* Oxford, Chandos Publishing.

Kaczmarek et al. (2006). Using the Audit Checklist for the Certification of a Trusted Digital Repository as a Framework for Evaluating Repository Software Applications. A Progress Report. *D-Lib Magazine. 12*(12).

Kirakowski J**.,** B Cierlik (1999) Context of Use: Introductory Notes, HFRG, Cork.

Lynch, C. (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *Portal: Libraries and the Academy. 3*(2). 327–336

LMS Evaluation (2004) - Open Polytechnic NZ Technical Evaluation of Learning Management System.

Manuscriptorium Digital Library http://www.manuscriptorium.eu accessed on 8 May 2008

MINERVA Technical Guidelines document, [accessed on 8 January 2008] http://www.minervaeurope.org/publications/technicalguidelines.htm

Moore, R. & Smith, M. (2006) Assessment of RLG Trusted Digital Repository Requirements. 11–15 June Joint Conference on Digital Libraries. Chapel Hill, North Carolina. http://www.ils.unc.edu/tibbo/JCDL2006/Moore_Smith-JCDLWorkshop2006.pdf, accessed on 8 January 2008.

OAIS - Open Archival Information System Reference Model http://public.ccsds.org/publications/archive/650x0b1.pdf, accessed on 8 January 2008.

OARiNZ - Open Access Repositories in New Zealand http://www.oarinz.ac.nz/index.php accessed on 8 January 2008

Organ, M. (2006). Download Statistics: What do they tell Us? The Example of Research Online, the Open Access Institutional Repository at the University of Wollongong, Australia

Quality principles for Cultural Websites: a Handbook, 2005 Minerva Project**,** http://www.minervaeurope.org/userneeds/qualityprinciples.htm accessed on 8 January 2008

Recomendation for Space Data System Standards Reference Model for an Open Archival Information System (OAIS) CCSDS 650.0-B-1, 2002.

RLG-NARA (2006) "Audit Checklist for Certifying Digital Repositories" Edited by the RLG-NARA Task Force on Digital Repository Certification, http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf accessed on 8 January 2008.

Shearer, M. K. (2003). Institutional Repositories: Towards the Identification of Critical Success Factors. *Canadian Journal of Information and Library Science. 27*(3). 89–108.

Shipsey, F. (2005). Versions in the lifecycle of academic papers - user requirements and guidelines for digital repositories. Delivered at CERN workshop on Innovations in Scholarly Communication (OAI4). Geneva Switzerland. http://eprints.rclis.org/archive/00006365/, accessed on 8 January 2008.

Technical Evaluation of Selected Open Access Repositories in New Zealand, version 1.3, Report funded by the Tertiary Education Commission of New Zealand, 2006, 41 p.

The ***nestor*** Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification http://journals.tdl.org/jodi/article/viewArticle/199/180#10, TEXAS Digital Library, accessed on 8 January 2008.

The ***nestor***– Network of Expertise in long-term STORage / nestor Working Group on Trusted Repositories Certification: **Catalogue of Criteria for Trusted Digital Repositories,** Version 1, June 2006, c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2006060710

Thibodeau, K. (2006). What Constitutes Success in a Digital Repository? 11–15 June Joint Conference on Digital Libraries. Chapel Hill, North Carolina. http://www.ils.unc.edu/tibbo/JCDL2006/Thibodeau-JCDLWorkshop2006.pdf

UsabilityNet: Usability Resources for practitioners and managers http://www.usabilitynet.org/home.htm; http://www.usabilitynet.org/tools/methods.htm accessed on 8 January 2008.

Web Content Accessibility Guidelines 2.0 W3C Working Draft 17 May 2007 http://www.w3.org/TR/2007/WD-WCAG20-20070517/ accessed on 8 January 2008.

Webopedia - The online dictionary and search engine, http://webopedia.com/ accessed on 8 Jan 2008.

Westell, M. (2006). Institutional Repositories: Proposed Indicators of Success. *Library Hi Tech. 24*(2). 211–226.

Xia, J. & Sun, L. (2007). Assessment of Self-Archiving in Institutional Repositories: Depositorship and Full-Text Availability. *Serials Review. 33*(1), p.p. 14–21.

Zuccala et al. (2006). Digital Repository Management Practices, User Needs and Potential Users: An Integrated Analysis, accessed on 26 May 2007. http://cybermetrics.wlv.ac.uk/DigitalRepositories/FinalReport.pdf