# D3.1.3 FINAL REPORT ON EVALUATION OF ASSETS SERVICES

**Advanced Search Services and Enhanced Technological Solutions for the European Digital Library**

Grant Agreement Number: 250527
Funding schema: **Best Practice Network**

Deliverable D3.1.3 WP3.1

Report
v1.0 – May 29 2012
Document. ref.: ASSETS.D3.1.3.EDLF.WP3.1.V1.0

**Programme Name:** ..... ICT PSP
**Project Number:** ..... 250527
**Project Title**: ..... ASSETS
**Partners**: ..... Coordinator: ENG (IT)
Contractors:
**Document Number**: ..... D3.1.3
**Work-Package**:..... WP3.1
**Deliverable Type:** ..... Report
**Contractual Date of Delivery:** ..... March 31th 2012
**Actual Date of Delivery**: ..... May 29th 2012
**Title of Document**: ..... Final Report on Evaluation of ASSETS Services
**Author(s):** ..... Agnès Saulnier (INA)
..... Preben Hansen(SICS)
..... Cristina Martinez, Inaki Etxaniz (LABEIN)
..... Massimiliano Nigrelli (ENG)
.....Patrick Courounet (INA)

**Approval of this report** ..... APPROVED – Luigi Briguglio (ENG)

**Summary of this report:**..... see Executive Summary

**History**: ..... see Change History

**Keyword List**: ..... ASSETS, Services, Professional Services, Evaluation, Usability, ISO 9241-11, ISO 9126-2

**Availability**..... This report is:
X public

## Change History

| Version | Date | Status | Author (Partner) | Description |
|---|---|---|---|---|
| 0.1 | 08-03-2012 | Draft | Agnès Saulnier (INA) | First draft (plan, introduction, complilation of information from internal report, report of the first MAT evaluation) |
| 0.2 | 16-03-2012 | Draft | Massimiliano Nigrelli (ENG) | Renamed par. 3.4.4 into "Performing the evaluation". Added content to par. 3.4.4 |
| 0.3 | 24-03-2012 | Draft | Agnès Saulnier (INA) | Added MAT evaluation report (4.2) |
| 0.4 | 24-03-2012 | Draft | Agnès Saulnier (INA) | Added Notification evaluation report (4.9) |
| 0.5 | 24-03-2012 | Draft | Agnès Saulnier | Added Notification |

| | | | | |
|---|---|---|---|---|
| | | | (INA) | evaluation API report (5.1) |
| 0.6 | 25-03-2012 | Draft | Agnès Saulnier (INA) | Added progatation report + some minor changements in plan and corrections |
| 0.65 | 26-03-2012 | Draft | Preben Hansen (SICS) | Partial report on UGC tool + some minor additions |
| 0.66 | 26-03-2012 | Draft | Agnès Saulnier (INA) | Document formating |
| 0.7 | 26-03-2012 | Draft | Massimiliano Nigrelli (ENG) | Revised<br>4.9 Digital preservation: Notification<br>5.1 Preservation Notification service |
| 0.8 | 26-03-2012 | Draft | Agnès Saulnier (INA) | Modification of propagation report.<br>Add content by re-use by API access |
| 0.9 | 27-03-2012 | Draft | Agnès Saulnier (INA) | Add evaluation access to log requests<br>+ introduction modifications (methods)<br>+ executive summary (tab) |
| 0.10 | 28-03-2012 | Draft | Cristina Martinez (LABEIN) | Contributions from end-user evaluation<br>+ modification introduction and executive summary |
| 0.11 | 28-03-2012 | Draft | Preben Hansen (SICS) | Part 4.5.2 + appendix for UGC |
| 0.12 | 29-03-2012 | Draft | Massimiliano Nigrelli (ENG) | Amended "Executive Summary" and "Introduction" |
| 0.13 | 29-03-2012 | Draft | Preben Hansen (SICS) | Part 4.5.2 + appendix for UGC |
| 0.14 | 29-03-2012 | Draft | Agnès Saulnier (INA) | Part 4.4 Mat Enrichment + appendix for enrichment<br>Part 4.7 ingestion workflow management (not finished)<br>Part 4.10 (beginning) |
| 0.15 | 29-03-2012 | Draft | Agnès Saulnier (INA) | Part 4.8.1 (input from Patrick)<br>Modification of 2 (input from Cristina)<br>Part 4.10 (input from |

| | | | | |
|---|---|---|---|---|
| | | | | Ghislain)<br>Conclusion (input Preben, Cristina) |
| 0.16 | 05-04-2012 | Draft | Agnès Saulnier (INA)<br>Cristina Martinez (LABEIN) | Several upgrades |
| 0.17 | 06-04-2012 | Draft | Agnès Saulnier (INA) | Reporting of evaluation of the query log service (input from Diego Ceccarelli CNR) |
| 0.18 | 14-05-2012 | Draft | Sergiu Gordea | Review of 4.7 and 4.10 |
| 0.19 | 21-05-2012 | Draft | Patrick Courounet | Part 4.8 |
| 1.0 | 29-05-2012 | Final | Luigi Briguglio | Approval and Release |

# Table of Contents

## Executive Summary

One of the ASSETS objectives is to improve the usability of Europeana. ASSETS services need to be tested on real users to ensure that they are compliant with the project requirements.

User evaluations have been performed to verify the usability of the system. The term "user" includes **end users** that will interact with the ASSETS search engine and also **professional users** that will be involved in the metadata annotation elaboration for the addition of new corpus.

So, both evaluations have been user oriented. Real-end users (General public, Academic learners and Experts) have been involved in the evaluation of end-user services. Evaluation of professional services has been accomplished by both developers and content providers.

For each technical module included in the ASSETS services portfolio, technical evaluations have also been conducted.

This document is complementary to the deliverable D3.1.2. "Report on User Evaluation of Functionalities".

Its goal is to present:

- how the usability issues that were detected during the end user evaluations have been managed,
- how the evaluation of the professional services have been carried forward and its results.

# 1. Introduction

The objective is to evaluate the **quality of use** of both the ASSETS search engine interface (according to end-users experiences) and the ASSETS professional services (based on professional users reports).

This document firstly provides up-to-date recommendations on the usability issues that have been detected during the ASSETS end-user services evaluation and then included into the deliverable D3.1.2.

Secondly, this document deals with the evaluation of the ASSETS professional services and measures the quality in use with tests performed by professional users such as developers, data providers and data managers.

"*Quality in use is the user's view of the quality of a system containing software, and is measured in terms of the result of using the software, rather than properties of the software itself. Quality in use is the combined effect of the software quality characteristics for the user.*» [Nigel Bevan – 1999]

"Quality in use" is the user's view of quality. Evaluation of "quality in use" provides a link between the user approach to the system and the functionalities that it offers. The objective of "quality in use" is a system that can be used by real users in order to achieve selected task(s) in a specific context. This requires appropriate functionalities as well as an interface (GUI, APIs, web services…) to access the system. Due to their various traits, professional services can be evaluated by using their APIs and/or by accessing them through a graphical user interface (GUI).

The following table lists the services ASSETS services that have been evaluated by using a GUI access:

| | **Service** | **Service status** | **Method** |
|---|---|---|---|
| 1 | MAT (*Manual Annotation Tool*) | Early prototype | Cognitive walkthrough |
| 2 | MAT Scenario annotation | Client tool connected to the ASSETS platform | User test |
| 3 | MAT Scenario propagation (*automatic CERTH Propagated annotation with MAT*) | Client tool connected to the ASSETS platform | User test |
| 4 | MAT Scenario enrichment (*manual enrichment propagation*) | Prototype | Cognitive walkthrough |
| 5 | Content creation by re-use (*UGC tool*) | Service integrated into the ASSETS platform | User test |
| 6 | Acces to query log analysis | Prototype tested with a local database | Cognitive walkthrough |
| 7 | Ingestion: Workflow management | Operational phase | User test |
| 8 | Quality of the enriched ingestion (*metadata classification, knowledge extraction*) | Analysis of results obtained from real use conditions | Evaluation of results from quality point of view |
| 9 | Digital preservation: Notification | Finalized service not tested in a real context | User tests aiming to measure some ISO 9126 metrics |

| | | | |
|---|---|---|---|
| 10 | Digital preservation : Risk management & normalization | Finalized service not tested in a real context | User tests |

Other evaluations have been conducted on ASSETS services by using an API access; the chosen methodology consisted of user tests aiming to measure some ISO 9126 metrics:

1. Preservation notification service,
2. Risk management and normalization service,
3. Content by re-use,
4. Query logs service.

# 2. Addressing end-user evaluation usability issues

A user-centered evaluation of the ASSETS Portal has been carried out during the second year of the project. In order to ensure objectivity, this evaluation has been conducted by an external contractor (User Vision) who has tested the usability of ASSETS Portal on real end users (General public, Academic learners and Experts) that broadly match the Europeana 'Personas'. The usability study has looked at two areas in particular, namely the user interface and the user's behaviour when interacting with the services/functions.

After the evaluation, a detailed evaluation report has been produced (*D3.1.2 Report on user evaluation of functionalities*), where all the usability issues and recommendations have been reported along with the severity ratings of each problem. This report has been the feedback for the developers who have analyzed the issues identified in the user tests.

Since, due to limited budget and schedule, not all usability problems found can be afforded, some method for prioritizing them must be defined to provide a solution that offers the best usability per unit of cost.

First of all, we have identified those issues that are not applicable (N/A). They correspond to either accessibility issues -that are not an objective of ASSETS- or issues that have already been addressed by Europeana in the new portal version, e.g. "Missing home link".

Then, we have defined a **priority** rank (High, Medium or Low) for the rest of the issues based on (i) the severity, (ii) the estimated cost to fix and (iii) the relevance of the issue for the ASSETS project.

Next, the developers have addressed the usability issues in **priority order**, starting with the High Priority issues, then as many of the Medium and Low Priority issues as time and resources have allowed. As result, the final version of the ASSETS portal has been released.

The next figure provides a global view about the usability issues detected in the user evaluation and how many of them have been addressed.

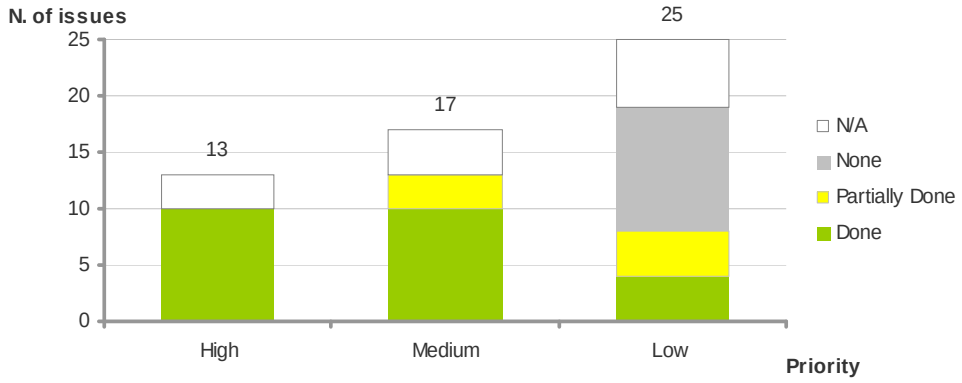



**Figure 1: Usability issues rank and final status**

The main conclusions extracted from the process are*:

- 74% of the issues have been addressed, totally or partially.
- All the High priority issues have been addressed: 100% Done.
- All the Medium priority issues have been addressed: 77% Done and 23% Partially Done.
- 42% of the Low priority issues have been addressed: 21% Done and 21% Partially Done.

(*) Percentages refer to applicable issues.

In the next sections we present a summary of the analysis performed on the usability issues coming out from both, the Heuristic Review and the User Tests. For each issue we show the Group the issue belongs to (e.g. Video Details, Zoom Tool, etc.), the Description, the Severity rank assigned by the usability consultant, whom the issue Concerns to, the Priority rank assigned by the ASSETS team, the performed Action (Done, Partially Done, None, N/A) and some Comments explaining either the changes introduced in the Portal or the reasons for which the issue has not been addressed.

## 2.1 High Priority issues

| N. | Group | Usability Issue | Severity | Concerns | Priority for ASSETS | Action | Comments |
|---|---|---|---|---|---|---|---|
| 16 | Audio details | **"Audio descriptions" is inconsistent with "semantic categories"**: The site uses two different names for the functionality that allows users to find similar content by clicking on descriptors. This inconsistency is likely to prove confusing for users. This is particularly true when it has the same look and feel, as this does. Elsewhere on the web, both "audio descriptors" and "semantic categories" use a more generic term: "tags." | **H** | ASSETS-GUI | **H** | **Done** | The heading "**Tags**" is more related to user annotations and is already used in Europeana.<br><br>The heading '**Audio descriptors**' have been changed to '**Audio categories**'. |
| 3 | General Search | **Poor colour contrast**: Currently the site has various areas with very poor colour contrast. Apart from being an important accessibility issue, it may appear to users as thought the content is disabled. | **H** | ASSETS-GUI | **H** | **Done** | |
| 20 | Relevance Feedback | **RF-Accessibility issue for relaying only on colour identification**: The meaning of a particular element should never be communicated by colour alone as this excludes users with visual colour impairments from understanding and interpreting the results. | **H** | ASSETS-GUI | **H** | **N/A** | Accessibility issue, out of scope Accessibility issue, out of scope.<br>We could add a text (e.g. Relevant/Not Relevant; Like/Dislike), but it would add more noise to the page. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 35 | Relevance Feedback | **Relevance feedback buttons are confusing:** Users were confusing the relevance feedback button with “Zoom in/out”. Additionally, users did not understand the difference between “More like this” and “Similar”. The whole concept of filtering by relevance was very difficult to understand by the users. | **H** | ASSETS -GUI | **H** | **Done** | The design of the GUI for the Relevance Feedback service has been changed:<br>- The RF mode is started by clicking on an action (left panel)<br>- The results page mode changes, a layer will be added over the results showing RF buttons on all of them<br>- The RF search is started by clicking on a button located at the first line of the search results box<br>- In order to exit the RF mode, a Cancel action has been added to the Actions panel |
| 25 | Search box | **Missing spell corrector:** When users misspelled any of the words in the query, the system did not provide any suggestions or spelling corrections. | **H** | Europeana 1.0 | **H** | **N/A** | Addressed in current Europeana |
| 38 | Video details | **Users confused the whole frame with the Similar button:** When asked about finding videos that contained similar images, the users tried to click on the frame itself instead of the icon at the bottom. This was probably caused due to having the same colour when the mouse was hovering. | **H** | ASSETS -GUI | **H** | **Done** | The keyframe border has been changed from blue to yellow.<br>A tooltip has been assigned to the "similar" button: "Search more videos similar to this keyframe" |
| 39 | Video details | **“Similar key frame videos” concept was difficult to understand:** Users did not understand the concept of searching for similar frames. The whole idea was alien and regarded as pointless. Having a functionality that is not user-friendly, adds noise to the site and prevents users to use other useful services. | **H** | ASSETS -GUI | **H** | **Done** | This service represents an Innovative way of searching videos, so there is no reason to remove it.<br>Some GUI changes have been introduced (tooltips, film strip and literals) that help to improve the usability of this service. |
| 6 | Zoom tool | **Icons are not self-explanatory and sometimes missing**: The “Save to my Europeana” and “Find similar” icons are not self-explanatory and will rely on users checking the tooltip to understand its purpose. Furthermore, when the icons are not available, they are not shown as disabled to the users. This creates inconsistency which makes it more difficult for the user to learn how to interact with the system. | **H** | ASSETS -GUI | **H** | **Done** | The "~" button label has been changed by "similar" in Zoom View, Result details and Video Details.<br>The “Save to my Europeana” tooltip has not changed because this literal is inherited from Europeana.<br>The "similar" button is shown as disabled when the item is not in the index. |
| 49 | Notifications | **Lack of instructions:** Users did not understand that they have to choose a term instead of free writing their own search query. Many were left feeling lost when they were typing something that was not on the suggestions list. | **H** | ASSETS -Comm. Services | **H** | **Done** | When the user starts to create a new subscription, now the UI displays a "tool tip" to suggest users to enter some parts of keywords to subscribe. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 43 | Preference query | **The refine button was not visible:** Most users expected the filter to be applied just by clicking the filter options, instead of having to press “Refine”. Additionally when some filter categories are open the button falls under the fold and is not visible. | **H** | ASSETS -Comm. Services | **H** | **Done** | We have followed the recommendation suggested by the usability test team.<br>We have removed the "refine" button. Now changes of filter setting are directly reflected to the search result. (*) |
| 44 | Preference query | **Use of language short code instead of the actual language name:** When filtering by language, the language short code is shown, most users were not familiar with them or were not sure which language they were referring to (i.e. mul, pl, etc.) | **H** | ASSETS -Comm. Services | **H** | **N/A** | This problem is also fixed in the current Europeana. |
| 45 | Preference query | **“Show block by block” was confusing:** Users did not understand what “Block-by-block” meant and assumed it was to show each filter separately. | **H** | ASSETS -Comm. Services | **H** | **Done** | We have followed the recommendation suggested by the usability test team.<br>We have used “Show combination of filters separately” instead of “Show block by block”. |
| 46 | Preference query | **“Next” & “Back” links are not meaningful or visible:** The buttons for moving between “blocks” are not visible at all for users. When probed about them , users are not sure about what they are or what they will get. | **H** | ASSETS -Comm. Services | **H** | **Done** | We have followed the recommendation suggested by the usability test team.<br>We have put links for block navigation into the next line of the breadcrumbs and use "Previous combination / Next combination" for the labels of the navigation links. |

(*) The CSS style and some behaviours of the taxonomy-based notification UI and the preference query UI exactly follow one's of the ASSETS platform that is derived from the code of an "old" Europeana. In the current version of Europeana, this issue is already fixed but such changes are not reflected to the ASSETS codes.

## 2.2 Medium Priority issues

| N. | Group | Usability Issue | Severity | Concerns | Priority for ASSETS | Action | Comments |
|---|---|---|---|---|---|---|---|
| 41 | Audio details | **Audio categories were misunderstood:** Users did not understand what audio categories meant and what was it for. Most of them could not understand the descriptors like “relaxed, happy, blue, tonal” and were confused by them. | **M** | ASSETS-GUI | **M** | **Done** | This is an innovative way of indexing audio content that goes beyond the concept of music style category. Some descriptors are meant for music experts.<br>Tooltips have been added to the audio categories: "search for mood_relaxed"; "search for speed_fast"; "search for mood_happy", etc.<br>Descriptors have been displayed one by one by specifying tuples of the format (descriptor:value) |
| 19 | Draw&Search | **“Reset” option should be located close to the other drawing related calls to action**: All drawing related actions are located to the left of the canvas, meanwhile “Reset” is located under “Search” on the right. This location does not build association. | **M** | ASSETS-GUI | **M** | **Done** | The location of the "Reset" option has been changed and a tooltip has been added to the link.<br>To improve usability we have also changed the literal of the drawing actions (Draw -> Pencil; Erase->Eraser) |
| 55 | Draw&Search | **“Fill” was not understood by the users:** Users did not use the “Fill” option most of the times and therefore the system misunderstood what the user was trying to draw. This meant that models were returned that were not at all relevant to what had been sketched. E.g. a filled triangle returned a pyramid model, but an unfilled triangle did not | **M** | ASSETS-GUI | **M** | **Partially Done** | The algorithm used for the implementation does not allow to provide results that relate to empty shapes as well as filled shapes. Also, filling the shapes automatically is not a trivial issue, since there is no guaranty that the user will close the shape. The best way is to let the user know that s/he must fill the shape.<br>A tooltip has been added for the "Fill" option |
| 1 | General Search | **Recursive link on Pagination:** In the pagination links at the foot of the page, the current page appears as an active link, while the other pages appear to be disabled. This is likely to create a lot of confusion. Additionally, this is an accessibility issue as users using assistive technologies could understand that they are not on the correct page. | **M** | ASSETS-GUI | **M** | **Done** | The navigation style has been changed. The current page is shown with a blue frame around. The other pages have the same link style used in the page. |
| 9 | General Search | **Left hand side menu inconsistency**: On the content details, the left hand side menu changes from the default and is therefore not consistent with the rest of the site. It is important to maintain consistency when using the site to understand where the user is. | **M** | Europeana 1.0 | **M** | **N/A** | The Result details page hs changed on the current Europeana. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 21 | Relevance Feedback | **Missing instructions or context for relevance buttons**: The current relevance buttons do not have any headers or instructions. As this feature is not very common it is likely that this functionality will not be well understood by many of the users. This problem is compounded by the fact that the tooltip is not very explanatory - “Add to positive list” or “Add to negative list” may not have meaning for the users | **M** | ASSETS -GUI | **M** | **Done** | The tooltips have been changed to "More like this" / "Less like this" |
| 8 | Result details | **Actions appear unrelated to the content details (FR)**: Users did not see the Actions list at the left of the file details. Only when probed around were able to notice them but mentioned that they expected them to be below the details, not on the left. | **M** | Europeana 1.0 | **M** | **N/A** | The recommended changes have already been adopted in current Europeana:<br>- Related content has been moved at the bottom of the content page<br>- Actions appear at the right of the content page |
| 23 | Search box | **Search options were not understood**: Many users did not understand what Enter URL or Draw meant on this context. | **M** | ASSETS -GUI | **M** | **Done** | We have added a "?" button close to each of the “more search tools” options. When the mouse is over it a small information box will be displayed. |
| 24 | Search box | **Other search options were very difficult to find:** Users did not expect to find these functions on this page, so they were unsure what the “Other search options” would provide.<br>The new search options should be more visible and explicit (tabs), particularly on the homepage, allowing the users to learn all the different features the website can offer. | **H** | ASSETS -GUI | **M** | **Partially Done** | The Home page should be as simple as possible (simplicity for a user visiting the site for the first time. Expert users will find their way for advanced search options). Also, introducing more Tabs in the search page will result in a crowded design.<br>Besides, the actual solution will allow Europeana to add new search services in the future smoothly.<br>In order to improve usability, we have increased the separation between "More search tools" and Suggestions. Also, the “Draw and search” action will indicate that the results will be 3D objects. The help box on the more search tool options will also help to explain the functionality of the Search options. |
| 26 | Search box | **Suggestions dropdown rows are not fully selectable:** Users often tried to click on the whole row to selecting one of the suggestions and were left confused when it was not selectable. Only clicking on the word itself allowed the row to be selected. | **M** | Europeana 1.0 | **M** | **N/A** | Addressed in current Europeana |
| 12 | Upload&Search | **Upload and search does not relate to search for url**: The method for performing an image search is currently unintuitive. When users are searching by url, they need to click to “Upload and Search” and “Upload a URL”. The | **M** | ASSETS -GUI | **M** | **Done** | The "Enter URL and Search" option has been included at the first level of the "More search options" menu. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | wording used here is not intuitive for this task. | | | | | |
| 52 | Upload&Search | **Confusing option:** Users were not sure which option was selected: Upload File or Enter URL. Using the colour and font weight change was not enough for clarifying which action was selected. | **M** | ASSETS -GUI | **M** | **Done** | This is sequel of last fast changes. Not needed anymore |
| 17 | Video details | **Video keyframes look very much like related videos (ER)**: The incorporation of selected keyframes at the bottom of the video could be confused with related videos. The principle of adding related videos as thumbnails overlaying the video screen is common practice on video sharing sites, such as YouTube.<br><br>**Frames were not completely clear (FR):** Users were unsure of what the keyframes were, on the first glance they thought they were related videos or shorter versions of the same video | **M** | ASSETS -GUI | **M** | **Partially Done** | This service represents an Innovative way of searching videos. If users don't understand the concept "Video Keyframes", is it useful to change the layout? Moving the keyframes out would reinforce the association with related videos (instead of part of the video), so we decide to keep the set together.<br><br>To improve usability we have included the "Video Keyframes" heading, the Tooltip for the "similar" button and the Film strip around the keyframes |
| 36 | Video details | **"Summary" was not being related to having a shorter version:** Some users, particularly in Stockholm, mentioned that "Summary" was not clear and they were surprised of what they saw when clicking on it. | **M** | ASSETS -GUI | **M** | **Done** | |
| 5 | Zoom tool | **Close-up is shown when mouse is over but not on focus**: Having a close-up of the item when the mouse is over is a very useful feature, but it needs to be made accessible by replicating this behaviour when using the keyboard only. | **M** | ASSETS -GUI | **M** | **N/A** | Accessibility issue, out of scope |
| 48 | Notifications | **"Subscriptions" was not related to "Notifications"**: Users did not easily find where to create new alerts or notifications for new content. When looking at "My Europeana" tabs, they thought that Subscriptions was something different | **M** | ASSETS -Comm. Services | **M** | **Done** | We have integrated "Subscription" tab and "Notification" tab into one "Subscription" tab. In the new subscription list, each subscription is coupled with notifications matching the subscription. |
| 42 | Preference query | **Dates were not sorted chronologically:** When users tried to filter by specific dates, they found out that the dates were sorted by the number of results, instead of chronologically. This confused users who are used to search chronologically when referring to dates. | **M** | ASSETS -Comm. Services | **M** | **Done** | Originally, dates were sorted by number of matched items (the behaviour of the ASSETS platform and the previous Europeana). Now dates are chronologically ordered (like the current Europeana) (*) |

(*) The CSS style and some behaviours of the taxonomy-based notification UI and the preference query UI exactly follow one's of the ASSETS platform that is derived from the code of an "old" Europeana. In the current version of Europeana, this issue is

already fixed but such changes are not reflected to the ASSETS codes.

## 2.3 Low Priority issues

| N. | Group | Usability Issue | Severity | Concerns | Priority for ASSETS | Action | Comments |
|---|---|---|---|---|---|---|---|
| 18 | Draw&Search | **Drawing space is too small for detailed drawings (ER)**: The drawing space is too small to allow the user to draw in any detail, thus reducing the chances of accurate results being returned. There is also no possibility of increasing it.<br><br>**Very small canvas (FR)**: Users complained that the canvas was too small for sketching and that it was difficult to define the details, making the drawing very inaccurate. | **H** | ASSETS-GUI | **L** | **Done** | We have increased the size of the canvas box in the GUI to 150x150 (current size if 100x100).<br><br>The support for resizable control is not a trivial issue and should be avoided. |
| 54 | Draw&Search | **Using the mouse is not a natural way for sketching:** Users mentioned that the mouse was not accurate enough for sketching and it felt unnatural. | **M** | ASSETS-GUI | **L** | **None** | This implementation is just a first approach for testing this type of service, so we are not going to increase the complexity.<br>Choosing shapes or monument types is not an acceptable option, since the objects are not only buildings and monuments.<br>Increasing the size of the canvas will surely improve the User Experience. |
| 2 | General Search | "**Sounds**" is not the best descriptions for content type. The use of "Sounds" is not as broad as "Audio". | **L** | Europeana 1.0 | **L** | **N/A** | This label has been inherited from Europeana. |
| 10 | General Search | **Title includes HTML tags**: Some content titles includes HTML tags as part of the title reducing its readability and creating an accessibility issues for users with screen readers. | **H** | Europeana 1.0 | **L** | **None** | This issue depends on how content providers do the indexing. |
| 4 | General Search | **Inconsistency on breadcrumbs**: Although the use of breadcrumbs is very positive, the wording used needs to be consistent and meaningful for the users. The 'metadata' style wording used here is unlikely to be meaningful to most users. | **M** | Europeana 1.0 | **L** | **None** | Breadcrumbs are used both for tab selection and for refining search. The use of current breadcrumbs is especially useful for the second. |
| 11 | Relevance Feedback | **"relevanceFeedback" is displayed on the search text field when relevance rating is provided**: Once the user has provided the relevance ratings, the search field shows "relevanceFeedback", having this query visible can affect the users trust on the site as they have not introduced it. | **H** | ASSETS-GUI | **L** | **None** | This issue is a constraint of the Europeana architecture. The adopted solution has been defined to integrate complex query types into Europeana with less impact as possible |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 40 | Result details | **"Paid access" created trust issues:** When users were trying to play certain videos or audio files, they noticed the Paid Access icon and thought that they had to pay for accessing the content. This created trust issues with the page as users thought that it was a marketing strategy instead of free content. | **L** | Europeana 1.0 | **L** | **N/A** | |
| 27 | Results page | **Related searches are easily missed:** Users did not look at Related Searches for more information or for other options. Most of the time they didn't see them, confused them with spelling correctors or went to the bottom of the page for something similar.<br>Move the related content and actions to inside the blue box for better association. | **M** | ASSETS-GUI | **L** | **Partially Done** | Things are at the bottom of the page because they are the least used options (e.g. Google advanced search).<br>So, we think that the adequate place for this type of info is below the search box.<br>In order to improve usability, we will change the heading of the suggestions to "Other users searched for" |
| 28 | Results page | **Legend is not clear:** Although many users noticed the legend on the right of the page, they did not understand what it meant for them. This could be caused because it was calling their attention before noticing the images on the results or because it was not related to the results page**.** | **L** | Europeana 1.0 | **L** | **N/A** | Addressed in current Europeana |
| 29 | Results page | **"My Europeana" section is expected to be on top of the page:** Users did not find "My Europeana" link easily or confused it with other information. They mentioned that it was on the wrong location, as the rest of the personal information was on the top of the page line. | **L** | Europeana 1.0 | **L** | **N/A** | Addressed in current Europeana |
| 31 | Results page | **Missing home link:** Some users tried to find a home link and did not recognise the logo as a link | **L** | Europeana 1.0 | **L** | **N/A** | Addressed in current Europeana |
| 32 | Results page | **Tabs need to maintain selection and be automatically selected:** Users found irritating that the tabs were not maintained when searching again or that they were not automatically selected when included as part of the query. Furthermore, some users missed the tabs even when including 3d models or videos as part of the query. | **L** | Europeana 1.0 | **L** | **N/A** | There are no tabs in the current Europeana version |
| 13 | Similar Search | **Meaningless query is shown when searching for similar images**: Finding similar images command should not be visible for users to avoid confusion. | **H** | ASSETS-GUI | **L** | **None** | This issue is a constraint of the Europeana architecture. The adopted solution has been defined to integrate complex query types into Europeana with less impact as possible. |
| 15 | Similar Search | **Results are not similar to the image in the url provided.** In this example, a URL to an image of the Eiffel Tower was provided (below). However, it is not clear why the results provided have been returned for this image. | **H** | ASSETS-GUI | **L** | **None** | Big effort, unaffordable at this stage (*). |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 34 | Similar Search | **Similar button is now very clear:** Users found the new design for similar button very easily and were happily surprised by this functionality. Some users were still unsure about the criteria used for finding similar items.<br>A further explanation of this functionality or being able to select that criteria may be needed for better understanding. | **P** | ASSETS -GUI | **L** | **None** | Big effort, unaffordable at this stage. It would be just for expert users.<br>A help page explaining actual behaviour will be useful when the service is integrated in Europeana |
| 14 | Upload&Search | **Missing "similar to" image (ER)**: When users get the results for the similar images, the actual (source) image is missing. This is inconsistent with other ASSETS services.<br>**Missing similar image or uploaded image (FR):** Users were not sure if the images shown on the similarity by relevance feedback or by uploading were taken into account. Furthermore, they missed the original images for comparison purposes. | **L** | ASSETS -GUI | **L** | **Partially Done** | For the "Upload File" option is not straightforward to implement (needs server-side changes). For Privacy-issues, the file is not stored in Europeana portal. Also Europeana is not currently ready to support content uploaded by users.<br>For the "Enter URL and search" option the source image is shown. |
| 51 | Upload&Search | **Missing "Upload" button:** Some users expected to find an Upload button instead of Search. They expected a two-step action. | **M** | ASSETS -GUI | **L** | **None** | We disagree with splitting up the service in two steps if it's possible to do it in 1 step. Following Google style, the less steps are the better for the user. |
| 53 | Upload&Search | **Missing terms & conditions for the upload:** Some users mentioned that they were not sure about what was going to happen with their uploads and it created trust issues. | **L** | ASSETS -GUI | **L** | **Done** | This notice will be included in the help box: "your content will not be stored in our servers" |
| 37 | Video details | **"Full video" was confused with "Full screen":** Some users confused the "Full video" button with the "Full screen" button. | **L** | ASSETS -GUI | **L** | **Done** | Change "Full video" by "Video"<br>Change ➢ by a ⊙ |
| 30 | Zoom tool | **Expanded image makes it difficult to click some sections:** The lack of delay for popping up the expanded image was making it difficult to click on certain buttons, like the information icon or to move around the different results | **M** | ASSETS -GUI | **L** | **None** | The only clickable thing is the thumbnail itself…we decide that it's better to show the zoom view as soon as possible. |
| 33 | Zoom tool | **Icons for file types are confused with buttons:** Particularly for audio and video files, users tend to click them for playing the content. Users were confused due to not having any action when they were clicking on the icon.<br>Include the icon as the "Open details" link so the users get some action when clicking. | **M** | ASSETS -GUI | **L** | **None** | We have used Europeana type icons (grey, faded) and the cursor doesn't suggest an action.<br>An action associated to open details will imply to remove the "i" button and consistency among types would be lost.<br>Any icon informing about media types that we introduce in the Zoom tool could be misunderstood as a button, i.e. the issue does not depend of the icon design. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 7 | Zoom tool | **Redundant function**: The "i" icon duplicates the behaviour of clicking on the main image. As clicking on the images is an intuitive action in this context, the "i" button is redundant and therefore only generates visual clutter on the page. | **M** | ASSETS -GUI | **L** | **None** | We think that it's good to have both options. Sometimes it's not clear for the user to make a click. |
| 22 | Notifications | **Look and feel inconsistency for Subscriptions:** The style of the tabs and buttons of the subscription service does not follow the style of the rest of the site. This could create trust issues of the services. | **M** | ASSETS -Comm. Services | **L** | **Partially Done** | The original style of the ASSETS' MyEuropeana has a problem in the visibility of buttons. We adopted the button style used in the current Europeana instead of the original ASSETS' MyEuropeana style. (**) |
| 50 | Notifications | **Buttons were not visible enough:** Due to the low colour contrast and not being placed closer to the notification text, the buttons were easily missed by the users. | **L** | ASSETS -Comm. Services | **L** | **Partially Done** | The original style of the ASSETS' MyEuropeana has a problem in the visibility of buttons. We adopted the button style used in the current Europeana instead of the original ASSETS' MyEuropeana style. (**) |
| 47 | Preference query | **Not evident how to sort them:** Users were not sure how to sort the different filters selected. | **L** | ASSETS -Comm. Services | **L** | **Done** | We have followed the recommendation suggested by the usability test team.<br>We use an "up-and-down" arrow cursor icon when a mouse pointer hovers over a sortable item. |

(*) The quality of similarity retrieval is affected by several aspects. A first relevant aspect is related to the content of the searched database. Given that a similarity value can always be measured between to arbitrary images, if the database does not contain images very similar to the query, the results of a query might appear not relevant simply because nothing really similar was found. In addition, it must be pointed out that similarity is a fuzzy and subjective concept. Similarity judgment changes between different people and different applicative scenarios. In this respect, various similarity measures can be used that might give more importance to the colours, shapes, textures, objects, etc. in the images and return completely different results accordingly. The similarity measure used in ASSETS is a combination of MPEG7 descriptors where the colours and the placement of the colours in the picture has the priority. Therefore the result of a query is a set of images having similar placement of colours. Specifically, no object recognition is currently performed.

(**) The CSS style and some behaviours of the taxonomy-based notification UI and the preference query UI exactly follow one's of the ASSETS platform that is derived from the code of an "old" Europeana. In the current version of Europeana, this issue is already fixed but such changes are not reflected to the ASSETS codes.

## 2.4 ASSETS Portal improvements

As result of the process described above, the final version of the ASSETS portal has been released. The next figures provide some insights on the main Portal components where some usability improvements have been introduced:

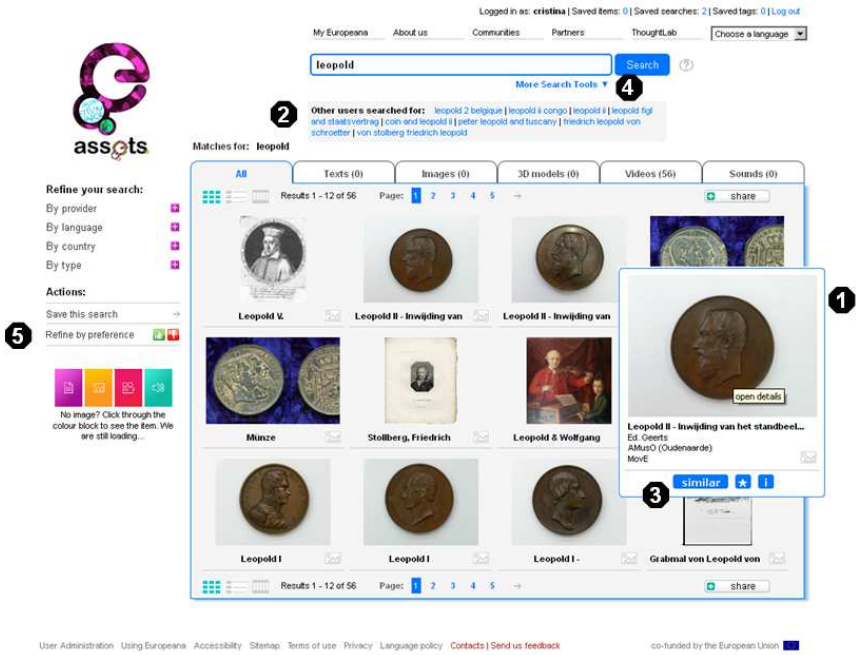



**Figure 2 - ASSETS portal search page**

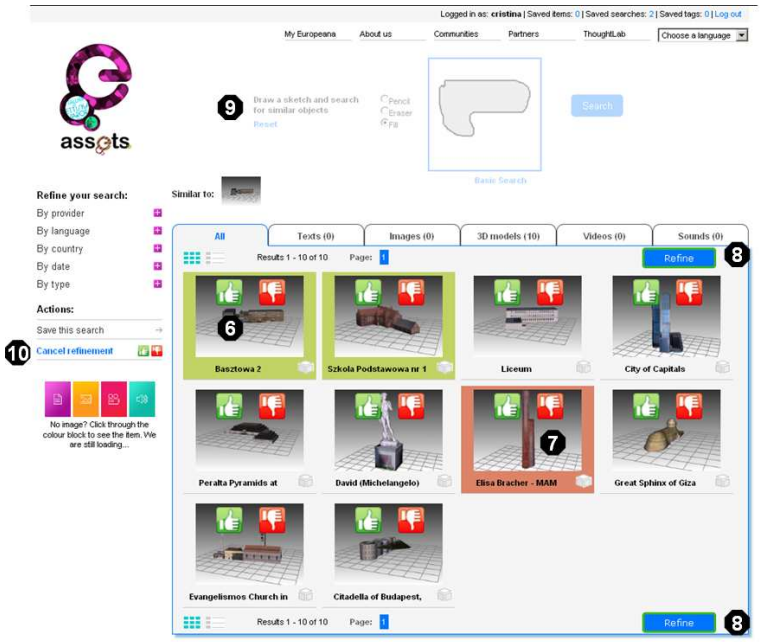



**Figure 3 - Relevance feedback mode and 'Draw and Search 3D'**

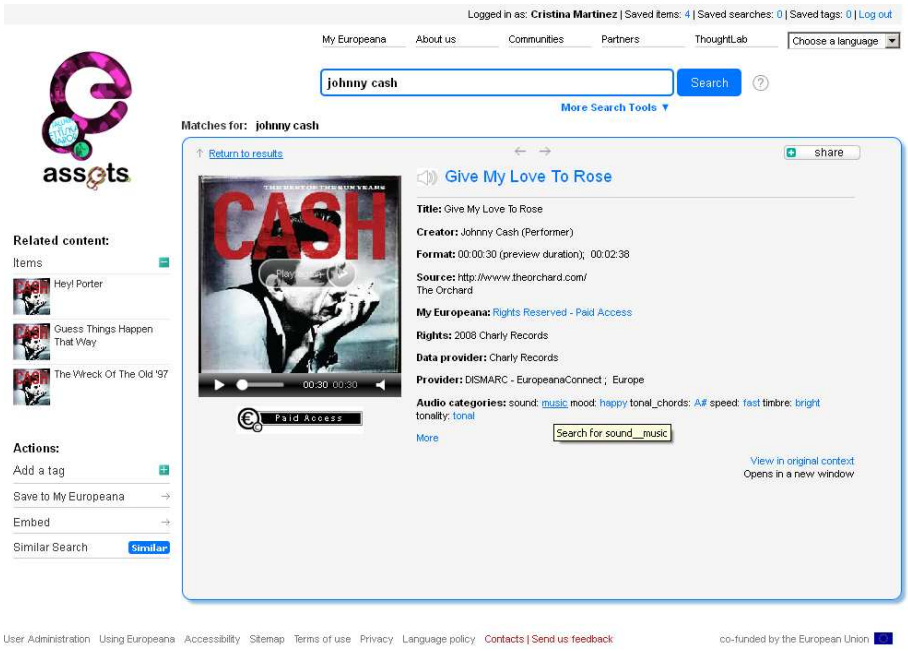



**Figure 4 – Audio details page**


**Figure 5 - Video details page**

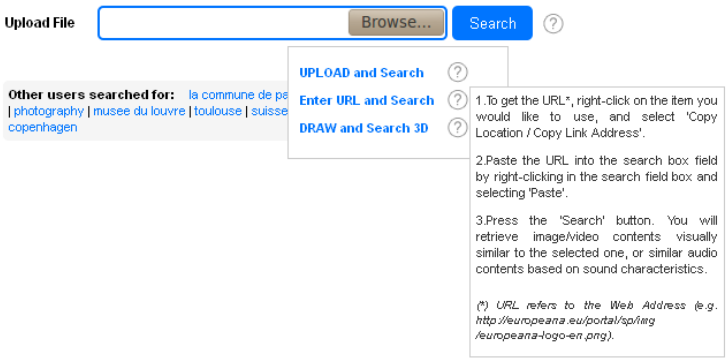



**Figure 6 -Help information for 'More search Tools'**

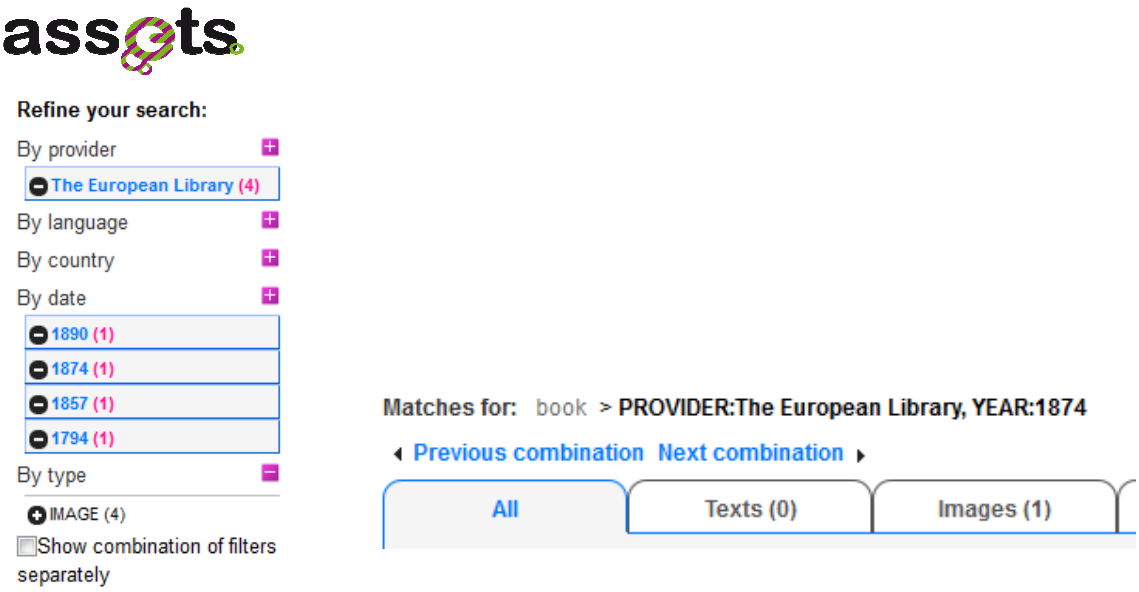



**Figure 7 – Preference query**

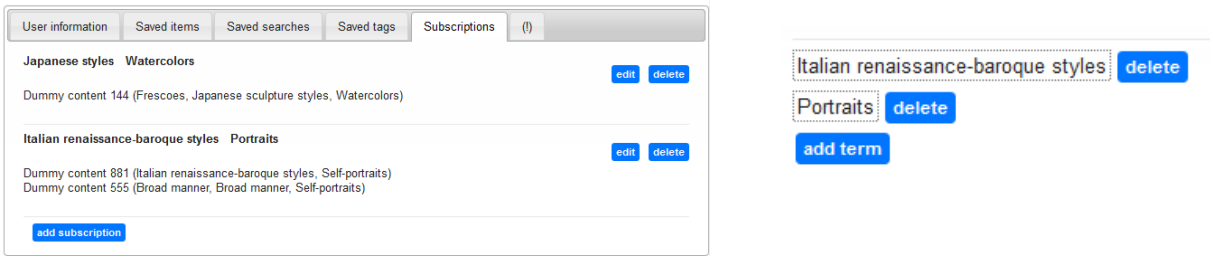



**Figure 8 – Notifications**

# 3. Criteria and methods for the evaluation of professional services

Professional services have to be evaluated with professional users to validate their **quality of use**. Evaluation of "quality of use" of services is in relation to a "broad" definition of usability [Brevan 95]. In fact the major evaluation objectives are to verify that the services enable the intended users to achieve their intended tasks. The methodology relies on measuring criteria relatives to users using the system rather than properties of the system itself.

**Usability** concept has been defined from different aspects. The more narrow aspects are limited to the "ease of use" aspect whereas broader aspects integrate the analysis of system adequacy with the user needs. The purpose of the evaluations presented here is relying on the broader aspect of the usability concept.

The evaluation of "quality of use" can be approached from a user oriented view or from a system oriented view. The first approach is commonly adopted by ergonomists to evaluate system usability from GUI access. The most recognized and most used corresponding standard is ISO 9241-11. The second approach is more often used by developers to evaluate the quality of a system. The most suitable standard is ISO 9126-2 which deals of product quality in a more complete aspect.

## 3.1 ISO 9126-2: Software engineering – Product quality

Technical Reports (ISO/IEC TR 9126-2 *External metrics*, ISO/IEC TR 9126-3 Internal metrics and ISO/IEC 9126-4 Quality in use metrics) provide a suggested set of software quality metrics (external, internal and quality in use metrics) to be used with the ISO/IEC 9126-1 Quality model. Software product quality can be measured internally (typically by static measures of the code), or externally (typically by measuring the behavior of the code when executed).

The objective for the system is to have the required effect in a particular context of use. Achieving quality in use is dependent on meeting criteria for external measures of the relevant quality sub-characteristics, which in turn is dependent on achieving related criteria for the associated internal measures. ISO 9126-4 is focus on the analysis of the user manipulation in "real conditions" whereas 9126-2 is focused on the behavior of the running system in laboratory conditions.

So ISO 9126-2 is the more pertinent part the standard to analyze the professional service from the user point of view before to deploy the complete system in real conditions (Figure 9).

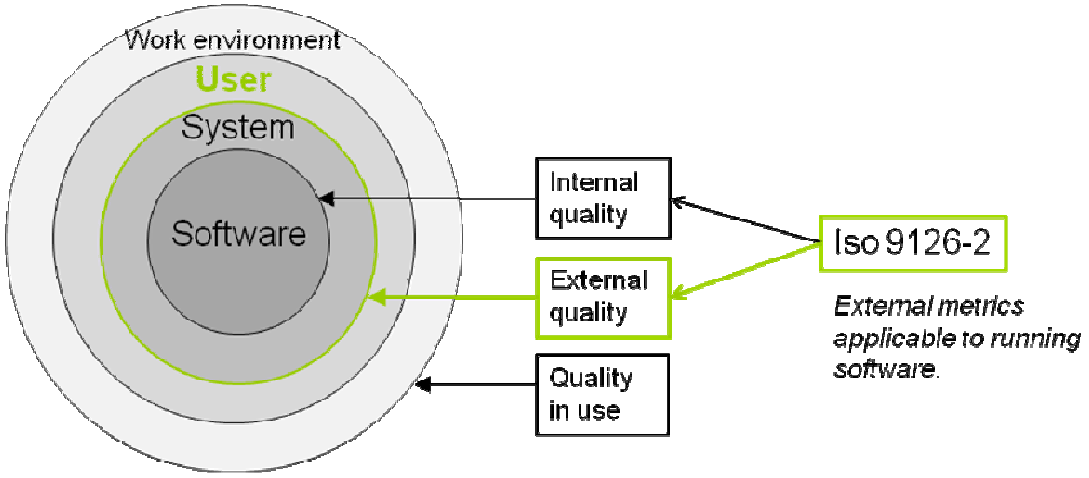



**Figure 9: ISO 9126 - evaluation of software quality**

ISO 9126 provides a general-purpose model which defines six broad categories of software quality that possessing all sub-characterics (Table 1) with measurable attributes:

- **Functionality**: The capability of the software to provide functions which meet stated and implied needs when the software is used under specified conditions;
- **Reliability**: The capability of the software to maintain the level of performance of the system when used under specified conditions;
- **Usability**: The capability of the software to be understood, learned, used and appreciated by the user, when used under specified conditions;
- **Efficiency**: The capability of the software to provide the required performance relative to the amount of resources used, under stated conditions;
- **Maintainability**: The capability of the software to be modified;
- **Portability**: The capability of software to be transferred from one environment to another

| Functionality | Reliability | Usability | Efficiency | Maintainability | Portability |
|---|---|---|---|---|---|
| Suitability | Maturity | Understandability | Time behaviour | Analysability | Adaptability |
| Accuracy | Fault tolerance | Learnability | Ressource utilisation | Changeability | Installability |
| Interoperability | Recoverability | Operability | Efficiency compliance | Stability | Co-existence |
| Security | Reliability compliance | Attractiveness | | Testability | Replaceability |
| Functionality compliance | | Usability compliance | | Maintainability compliance | Portability compliance |

**Table 1: 6 Characteristics and 27 sub-characteristic of ISO 9126**

The actual characteristics and sub-characteristics which are relevant in any particular situation will depend on the purpose of the evaluation. In these professional services evaluations, users are *end-users* as data providers, data managers or even developers who will use a system and have not contributed to develop the system. So, all of the characteristics listed above are not useful but only four of them: "The relationship of quality in use to the other software quality characteristics depends on the type of user, for instance: for the end user quality in use is mainly a result of functionality, reliability, usability and efficiency; » [Nigel Bevan – 1999]

So, for these specific evaluations, we are mainly focused on the following criteria:

- Functionality
- Reliability
- Usability.
- Efficiency.

Just notice that usability criterion is here defined in a "narrow" aspect which we do not restrict ourselves. It defines "the capability of the software product to be understood learned, used and attractive to the user, when used under specified conditions."

- Understandability**:** The capability of the software to enable the user to understand whether the software is suitable, and how it can be used to implement a particular service.
- Learnability**:** The capability of the software to enable the user to learn how to implement this software. Learnability metrics assess how long it takes users to learn to use particular functions, and the effectiveness of help systems and documentation.
- Operability**:** Operability considers the extent to which the user can easily manipulate the software, anticipated usability problems.
- Attractiveness**:** The capability of the software product to be liked by the user.

The sub-characteristics have also to be chosen according to each service to be evaluated and the standard gives all the measure definitions to use for each selected sub-characteristic.

## 3.2 ISO 9241-11: usability

Several definitions and norms have been developed for the usability concept. The definition of Usability given by the **norm ISO 9241-11** is one of the most commonly used ones: "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."

The components that compose the usability of the norm **9241-11** are:

- **Effectiveness**: Accuracy and completeness with which users achieve specified goals. To measure accuracy and completeness it is necessary to produce an operational specification of the criteria for successful goal achievement. This can be expressed in terms of the quality and quantity of output. Accuracy can be measured by the extent to which the quality of the output corresponds to the specified criteria, and completeness can be measured as the proportion of the target quantity which has been achieved.
- **Efficiency**: Resources expended in relation to the accuracy and completeness with which users achieve goals. Efficiency is measured by relating the level of effectiveness achieved to the resources used. For example, temporal efficiency can be defined as the ratio between

the measure of effectiveness in achieving a specified goal, and the time it takes to achieve that goal. This provides an absolute measure of temporal efficiency in a particular context.

- **Satisfaction**: Freedom from discomfort, and positive attitudes towards the use of the product. Satisfaction could be measured with a questionnaire.

There are some predefined measures given in this standard, but as example we can list a set of usual measures [Table 2].

| **Effectiveness** | **Efficiency** | **Satisfaction** |
|---|---|---|
| Percentage of goals achieved;<br>Percentage of users successfully completing task;<br>Average accuracy of completed tasks<br>Number of results errors<br>Number of used functions<br>Number of missing functions | Time to complete a task;<br>Tasks completed per unit time;<br>Number of manipulation to complete a task<br>Number of errors of manipulation | Rating scale for satisfaction;<br>Frequency of complaints |

**Table 2: Examples of usual measures for ISO 9241-11**

Usability analysis could provide recommendations at different levels of a system. Here it is an analysis example that could be done from user test:

- Functional level relative to adequacy between user needs and system functionalities, user performance.
  - ➔ Are all the function presents to achieve the given task?
  - ➔ Do they provide pertinent results in adequacy with user desiderata?
- Interaction level relative to easiness of use.
  - ➔ Are the functions easy to manipulate?
  - ➔ In adequacy with the user knowledge?
- Visual design level (for GUI) relative to easiness of perception and lisibility, graphic compliance with usual standard.
  - ➔ Is the organization and look-and-feel of the visual design relevant?
  - ➔ Easy to understand?

So measures have to be defined for each service to evaluate. The choice of criterion values of measures depends on the objectives of the evaluation and on the main requirements for the system. It may focus on the primary goal of the system or a sub-goal. Some criteria may not be determinant for a system. For example, for a professional service the learnability criteria may not be important as a complex system is not intended to be first completely intuitive but rather to achieve all the tasks accurately. Learning level has to be adapted to the user need.

## 3.3 Discussion between the two standards

The usability defined in ISO 9241-11 presents a broader approach than the only usability characteristic presented in ISO 9126.

In fact usability in ISO 9241-11 focuses on the human issues. In practice usability could be measured by examining the interaction of users with the system. The ease of use is not the only criteria to analyze. In fact effectiveness relies on utility criteria (right functionalities provided for the task?), on reliability criteria (running errors?) and on efficiency criteria (response time). So we could argue that **ISO 9241-11 is in adequacy with a sub-part of ISO 9126-2 2** (limited to functionality, reliability, usability and efficiency criteria).

From this point of view, these two norms share similar criteria but however these two norms will be useful in the evaluation of the professional services because of the different specificities of the services.

Services without GUI are more commonly evaluated with ISO 9126-2 which is more tailored to technical users and will be well adapted for programmers users while services with GUI addressed to content provided will be evaluated with ISO 9241 which is more addressed for ergonomist and non-programmers users. Otherwise, 9126-2 is more adapted to realize some tests on separate functions (fewer close of realistic conditions).

## 3.4 Evaluation methods

Several classifications of methods have been presented ([Nielsen 90], [Jeffries 91], [Bastien 91], [Coutaz 94], [Nielsen & Mack, 1994], [Preece, Rogers and Sharp, 2002] etc.). According to Senach [Senach, 90], mainly two classes of methods have been developed to evaluate or validate human tools: **experimentation** and **inspection**. Both methods may be applied in parallel to get complementary inputs. Experimentation methods are based on user tests and studies while inspection methods involve ergonomic, computer or/and domain experts.

For example, Nielsen and Mack (1994), define usability inspection methods as set of several methods including:

- Heuristic evaluation. Informal method involving usability specialists judging if each interaction element conforms to established usability principles.
- Guideline review. Each interface is checked for conformance with a list of usability guidelines.
- Pluralistic walkthrough. Meetings in which users, developers and human factor people walk through a scenario, discussing usability aspects
- Consistency inspections. Several designers from different projects inspect the interface for consistency.
- Standards inspection. Inspect the interface according a specified standard.
- Cognitive Walkthrough. Explicit and more detailed procedure to simulate a user's problem-solving process at each step in the interactive process.
- Formal usability inspection. A formal usability inspection method that is similar to a code inspection process usually performed by developers.

Furthermore, and with a slightly different conceptualization of the evaluation paradigms, Preece, Rogers and Sharp (2002) describe the following set of evaluation frameworks.

| | Usability testing | Field Studies | Predictive |
|---|---|---|---|
| Role of users | Carry out defined tasks | Natural behaviors | Users not involved |
| Control | Evaluators | Evaluators together with users | Expert evaluators |
| Location | Laboratory | Natural environment | Laboratory oriented, but often on customers premises |
| When used | With a prototype or product | In early design stages to check users needs and assess problems | Expert reviews (often done by consultants) with a prototype. Models are used to assess specific aspects of a design |
| Type of data | Quantitative, sometimes statistically validated. Users opinions collected by questionnaires | Qualitative descriptions often accompanied with sketches, scenarios or other artifacts. | List of problems from expert reviews. Quantitative figures from models e.g. how long it takes to perform a certain task. |
| Feedback into design | Reports on performance measures. Results provide a benchmark for future versions. | Descriptions that sometimes include log statistics. | Reviewers provide a list of problems with suggested solutions. |
| | Applied approaches based on experimentation i.e. usability engineering | Observations (objective) or ethnographic | Practical heuristics. |
| *Techniques* | | | |
| Observing users | Video and Interaction logging | Observation is the central part | -- |
| Asking users | User satisfaction questionnaires + Interviews | Evaluator interview/discuss with participants. | -- |
| Asking experts | -- | -- | Experts use heuristics early in design for prediction |
| User testing | Testing typical users | -- | |
| Modeling users' task performance | - | -- | Models used to predict efficacy of iUI |

**Table 3: Evaluation framework from Preece, Rogers and Sharp**

We have selected highly complementary methods to realize the evaluations of professional services of ASSETS

- **User test**: an experimental method which can provide quantitative and qualitative results.
- **Heuristic evaluation**: an inspection method which can provide qualitative results
- **Variant of cognitive walkthrough**: an inspection method realised with real users which provide qualitative results.
- **ISO 9126 evaluation framework:** user test adapted to collected measures according to ISO
- **Evaluation of results quality from user point of view**: evaluation to complete technical tests about results quality to provide a users' perspective.

Inspection methods are very effective to identify ergonomic problems or conception problem. It could be applied on an "early" prototype. Experimental methods relying on user tests should be conducted in a system which can be used in conditions which are closer to real conditions of use. User tests update problems more precise and specific problems that are directly related to the task of the users and user needs.

### *3.4.1 User Test*

Experimentation evaluations involve the manipulation of the system through users. User tests consist in interpreting the performance of users realizing tasks specific to the evaluation. Those studies might provide feedback on the problems users encountered with a tool. The objective of experimental evaluations is to verify that the system is adapted to the user's perceptive and cognitive characteristics. More often, experimentations gather qualitative or quantitative feedback from users, using objective (duration, performance measures etc.) or subjective (user's opinion) data. User tests may be run with log analysis, observations, interviews, questionnaires, gaze eyes tracking, thinking aloud…. Statistics elaborated from several types of user performances may be computed and compared to given thresholds or to the performances of another group of users, realizing similar tasks on similar data but on a different system. Usability objectives may be absolute ("90% of the users should find at least 3 documents in less than 3 min") or relative ("90% of the users should do better with the new interface than with a reference interface). In the second case the user test is realized to compare to interfaces.

Evaluation variables are dependent on the scenarios, tasks and criteria and should then be adapted to each evaluation.

Examples of measures:

- Effectiveness – Accuracy of achieved results using the system:
  - For a number of typical user tasks compare the accuracy of the result obtained using the system to the results obtained by a classical system
  - Or compare with a predetermined threshold.
- Efficiency – Time, number or complexity of interactions required to complete a task:
  - For a number of typical tasks, compare the time or number of required interactions to complete a task using the system to the time or number of interactions required achieving the same task using a classical system
  - or compare with a predetermined threshold.
- Satisfaction – Ranking regarding aesthetics, user friendliness:
  - For a number of typical tasks users express their satisfaction to execute a task using the system.

- Learnability – Time to learn the system:
  - For a number of tasks compare the evolution of effectiveness and efficiency after completing similar tasks.

**Remote user tests** can also be conducted with the user and the evaluators separated over space. A complete documentation is prepared to guide the user and also to let the user provide complete feedback. When necessary, help could be provided by email, phone or skype.

**User tests in operating phase** are tests conducted in real conditions of use.

### *3.4.2 Expert evaluation*

**Expert evaluation with heuristics**

Inspection may be performed by ergonomic experts or user domain expert, or by a developer and an evaluator. The main difference to the previous method is that the interaction with the system is performed by the evaluator. Heuristic evaluation is the most informal method and involves having usability specialists to judge whether each dialogue element follows established usability principles (the "heuristics"). This method provides qualitative results.

As this method involved no user, the satisfaction can't be analyzed. Some heuristic can give information about effectiveness (goal reached but no the quality of the results) and efficiency (resources needed but no quantitative information as time…). This method is particularly useful to verify a set of ergonomic criteria.

Examples of heuristics:

- Nielsen,
- Bastien and Scapin

**Expert evaluation with user domain expert**

Inspection may also be realized by ergonomic expert and user domain expert. This assessment is based on the knowledge of stakeholders.

### *3.4.3 Variant of cognitive walkthrough*

The cognitive walkthrough is an inspection method. The objective is to simulate a user's problem-solving process at each step through the dialogue, checking if the simulated user's goals and memory content can be assumed to lead to the next correct action.

The origin of the cognitive walkthrough approach is the code walkthrough used in software engineering. The sequence of actions refers to the steps that a user will have to perform to accomplish a task. Often, the main focus of a cognitive walkthrough is to establish how easy a system is to learn.

Some variants of cognitive walkthrough have been proposed [Mahatody, 10]. The objective to analyze the learnability could be widen to integrate the analysis of the quality of the results provided by the functionalities. Learnability could be also replaced by the criterion of ease of interaction. Professional users can also be integrated in the evaluation process to help the evaluator to answer to the questions. They don't have to manipulate the interface but to comment the evaluator manipulation. The interest to include real users in the process is that we can observe how they feel interested and satisfied with the tool. Do they develop spontaneously strategies and/or new tasks? They can also enrich the results giving comparison to their experience.

This method is based on discussions and comments while achieving a scenario of use.

The discussion should answer different types of questions at each step of the scenario:

1. Will the user think that he can do such an action?
2. Will the user find the way to launch the action?
3. Will the user be able to predict the effect of the launched action?
4. Does the action's effect match the user expectation?
5. After the action, is the feedback good enough to encourage the user to continue?

1. Does the user know what to do to achieve the task?
2. Is that easy for users to understand how to achieve the tasks with the interface?
3. Does the user understand the results given by the system?

1. Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?
2. Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?
3. Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?

This adapted cognitive walkthrough (realized with expert user) let obtain qualitative measure for each criteria of the usability:

- Learnability : according to the scenario analysis
- Effectiveness: verification that the predefined tasks could be realized
- Efficiency: comparison made by the user with his previous experience
- Satisfaction: given by the user

A cognitive walkthrough evaluation is conducted in 3 steps:

1. Preparation

   Prerequisites to the walkthrough include:

   - A general description of the user: who is he? What are his competences and relevant knowledge?
   - A specific description of one or more representative tasks to be performed with the system,
   - A list of the correct actions required to complete each of these tasks with the interface being evaluated.

2. Executing the action.

   For each step of the task realisation, the evaluator has to answer several questions with the user.

3. Results interpretation.

   Result analysis should give all the impressions and lead to ergonomic recommendations.

   It is possible to complete this evaluation with user manipulation to let the user discover more precisely the interface, realize some comparison, etc.

### 3.4.4 ISO 9126 evaluation framework

The service evaluation is based on the criteria and measures identified in 3.1

- Suitability.
- Understandability;
- Maturity;
- Functionality compliance;

Evaluation instruments for the users in charge to perform the evaluation are presented in “Appendix 1: Evaluation instruments for ISO 9126"

***a) Suitability***

For details about these metrics, please refer to **Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2.

Among these metrics, two have been chosen for the purpose of the ASSETS professional services evaluation:

- Functional adequacy,
- Functional implementation completeness.

*Functional adequacy*

For the scenario walkthrough, a list of functions featured by the service to evaluate has been addressed. These will represent the list of functions to be evaluated for each service (an example of which is displayed below):

| | |
|---|---|
| FA (1) | <1st function name> |
| …. | …. |
| FA (i) | <i th function name> |
| ….. | ….. |
| FA (N) | <N th function name> |

**Table 4: "Functional adequacy - An example list of the functions to be evaluated for the each service"**

Then a matrix has been built around this list of functions.

While walking through the test scenario (provided by the team responsible for the service development), the matrix is compiled by the evaluator.

Measurement for the metric will be exploited through the formula: **X = 1 – (A / B)**

where:

*A= Number of functions in which problems are detected in evaluation*

*B= Number of functions evaluated*

The interpretation of the measured value is the following: 0<= X <=1; the closer to 1.0 is the better.

*Functional implementation completeness*

For the scenario walkthrough, a list of functions featured by the service to evaluate has been

addressed; these items have been inferred by the functional requirements which the service has to cover.

These will represent the list of functions to be evaluated for each service (an example of which is displayed below):

| | |
|---|---|
| FIC (1) | <1st function name> |
| …. | …. |
| FIC (i) | <i th function name> |
| ….. | ….. |
| FA (N) | <N th function name> |

**Table 5: "Functional implementation completeness - An example list of the functional requirements to be evaluated for the each service"**

Then a matrix has been built around this list of functions.

While walking through the test scenario (provided by the team responsible for the service development), the matrix is compiled by the evaluator.

Measurement for the metric will be exploited through the formula: **X = 1 – (A / B)**

where:

*A = Number of missing functions detected in evaluation*

*B = Number of functions described in requirement specifications*

The interpretation of the measured value is the following: 0<= X <=1; the closer to 1.0 is the better.

***b) Understandability***

Users should be able to select a service which is suitable for their intended use. An external understandability metric should be available, so that new users can understand:

- whether the service is suitable,
- how it can be used for particular task.

For details about these metrics, please refer to **Table 8.3.1 Understandability** of the Norm ISO_IEC 9126-2

Among these metrics, two have been chosen for the purpose of the ASSETS professional services evaluation:

- Function understandability,
- Understandable input and output.

*Function understandability*

For the scenario walkthrough, a list of functions featured by the service to evaluate has been addressed. These will represent the list of functions to be evaluated for each service (an example of which is displayed below):

| | |
|---|---|
| FA (1) | <1st function name> |

| | |
|---|---|
| …. | …. |
| FA (i) | <i th function name> |
| ….. | ….. |
| FA (N) | <N th function name> |

**Table 6: "Functional understandability - An example list of the functions to be evaluated for the each service"**

Then a matrix has been built around this list of functions.

While walking through the test scenario (provided by the team responsible for the service development), the matrix is compiled by the evaluator.

Measurement for the metric will be exploited through the formula **X = A / B** where:

*A= number of functions whose purpose is correctly understood by the user*

*B= number of available functions*

The interpretation of the measured value is the following: 0<= X <=1; the closer to 1.0 is the better.

*Understandable input and output*

For the scenario walkthrough, a list of functions featured by the service to evaluate has been addressed. These will represent the list of functions to be evaluated for each service (an example of which is displayed below):

| | |
|---|---|
| FA (1) | <1st function name> |
| …. | …. |
| FA (i) | <i th function name> |
| ….. | ….. |
| FA (N) | <N th function name> |

**Table 7: "Understandable input and output - An example list of the functions to be evaluated for the each service"**

Then a matrix has been built around this list of functions .

While walking through the test scenario (provided by the team responsible for the service development), the matrix is compiled by the evaluator.

Measurement for the metric will be exploited through the formula **X = A / B** where:

*A= number of input and output data items which user successfully understands*

*B= number of input and output data items available from the system*

The interpretation of the measured value is the following: 0<= X <=1; the closer to 1.0 is the better.

### *c) Maturity*

For details about these metrics, please refer to **Table 8.2.1 Maturity metrics** of the Norm ISO_IEC 9126-2

Among these metrics, only one has been chosen for the purpose of the ASSETS professional services

evaluation:

- Maturity.

*Maturity*

A matrix has been built around the sequence of test scenario steps to be executed while performing the evaluation.

While walking through the test scenario (provided by the team responsible for the service development), the matrix is compiled by the evaluator.

Measurement for the metric will be exploited through the formula **X= A1 / A2** where:

*A1 = number of detected failures*

*A2 = number of performed test cases*

The interpretation of measured value: 0<=X. It depends on stage of testing. At the later stages, smaller is better.[1]

### *d) Functionality compliance*

For details about these metrics, please refer to **Table 8.1.5 Functionality compliance metrics** of the Norm ISO_IEC 9126-2

Among these metrics, only one has been chosen for the purpose of the ASSETS professional services evaluation:

- Functionality compliance.

*Functionality compliance*

Purpose of the metrics: how compliant is the functionality of the product to applicable regulations, standards and conventions (e.g. existing framework, EU projects, EU regulations)?

For the scenario walkthrough, a list of the functionality compliance items (e.g. existing framework, EU projects, EU regulations) has been addressed. The service to be evaluated has to be compliant with these items. This is an example list of the functionality compliance items for each service:

| | | |
|---|---|---|
| FC (1) | <1st model, framework, EU project, EU regulation to be compliant with> | <description of the 1st model, framework, EU project, EU regulation to be compliant with > |
| …. | …. | …. |
| FC (i) | <i th model, framework, EU project, EU regulation to be | <description of the i th model, framework, EU project, EU regulation to be compliant with > |

[1] FOOTNOTES

- The larger is the better, in early stage of testing. On the contrary, the smaller is the better, in later stage of testing or operation. It is recommended to monitor the trend of this measure along with the time.
- This metric depends on adequacy of test cases so highly that they should be designed to include appropriate cases: e.g., normal, exceptional and abnormal cases.

| | | |
|---|---|---|
| | compliant with > | |
| ….. | ….. | ….. |
| FC (N) | <N th model, framework, EU project, EU regulation to be compliant with > | < description of the N th model, framework, EU project, EU regulation to be compliant with > |

**Table 8**: " Functionality compliance – An example list of the functionality compliance items for each service"

Then a matrix has been built around this list of functionality compliance items (see an example of a "Functionality compliance Matrix" in 8.1).

While walking through the test scenario (provided by the team responsible for the service development), the matrix is compiled by the evaluator.

Measurement for the metric will be exploited through the formula **X = 1 – (A / B)** where:

*A= Number of functionality compliance items specified that have not been implemented during testing*

*B= Total number of functionality compliance items specified*

The interpretation of the measured value is the following: 0<= X <=1; the closer to 1.0 is the better.

### *3.4.5 Evaluation of results quality from user point of view*

The problem of results quality is also an underlying condition to provide good effectiveness criteria.

Generally speaking, there are two types of evaluation of data quality:

- Scientific validation: the results are analysed according to a ground truth to provide quantitative measures.
- Study of sources of error: this approach complete the first evaluation giving a user point of view about the results and especially about the level of errors

In this study, we are involved only with the second aspect of results quality study. Quality indicators have to be defined according to each system to evaluate. The idea is to give feedback about the errors, their relative importance and their consequences.

# 4. Evaluation of professional services using a GUI access

## 4.1 MAT Manual Annotation Tool (stand alone prototype)

### *4.1.1 Introduction*

This chapter presents the evaluation of the first prototype of the interface MAT. This prototype already contains many of the features that will be useful for correcting ASSETS manual records. It is not yet connected to the database ASSETS. At this stage, it is useful to conduct an evaluation to confirm the functions and provide recommendations on quality of use.

**Evaluation elaborated and completed by**: A.Saulnier (INA)

**Contact for the service**: P.Courounet (INA)

**3 cognitive walkthrough tests** performed at Ina

**Users**: 3 Ina annotators

**Evaluation place**: Ina (France)

**Date:** June 9th-14th 2011

**Test environment**: Computer DELL PRECISION T1500, screen DELL 2007WFP resolution 1680x1050

**Objectives**

The objectives are derived from the technical use case of D2.0.1. The goal is to verify that MAT allows professional users to perform their tasks with quality and easiness:

- To view a document and its annotation,
- To edit and to correct an annotation,
- To obtain a semantic aid to retrieve named entities

**Method**

Cognitive walkthrough is a usability inspection method that focuses on system's ease of use. For a chosen series of task to be realized by a specific type of user, an expert analyses all the tasks realization according to a set of questions.
We have applied a variant **of cognitive walkthrough** no more focused on how easy a system is to learn but on the quality of results and of interaction which can be reached by a user. This variant also includes users to gather more information as satisfaction ones.

**User profils** are annotators or archivists used to create and modify annotations of document corpus.

The evaluation is conducted with **1 interface expert and 3 annotators** of Ina who have experiences in annotations and are used to search information on Internet for annotation task. They are not involved in ASSETS project.

This method is based on discussions and comments while achieving a scenario of use according to this set of questions:

1. Does the user know what to do to achieve the task?
2. Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?

3. Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?
4. Does the user understand the results given by the system? Does the result quality match the user expectation?
5. Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?

**Scenario**

The scenario contains a specific description of one or more representative tasks to be performed with the system. A list of the correct answers required to complete each of these tasks has been also elaborated with one expert interface, one technical correspondent and one annotator user. It is focused on a video annotation modification. The interface is launched with one video name.

The objective for the user is to find the name of the military chief introduced in the proposed video "US generals at battlefront before Metz" and to complete the annotation of this video.

This scenario had been structured into 6 tasks:

- Task 1: select a document
- Task 3: Read and edit the annotation. Test to make one correction.
- Task 2: View the video and find some entity names missing in the annotation.
- Task 4: Find the entity names and theirs titles in the web
- Task 5: Chose the pertinent field to characterize this person
- Task 6: Import these descriptions in the annotation

**Measures and criteria**

This adapted cognitive walkthrough (realized with professional user) let obtain qualitative measure for each criteria of the usability:

- Effectiveness: verification that the predefined tasks could be realized
- Efficiency: comparison made by the user with his previous experience
- Satisfaction: feedback given by the user

The usability evaluation of interface implies to analyze different levels of the interface:

- Functions: are all the function presents to achieve the given task? Do they provide pertinent results in adequacy with user desiderata?
- Interaction: are the functions easy to manipulate? Are they in adequacy with the user knowledge?
- Visual design: is the organization and look-and-feel of the visual design relevant? Easy to understand?

### *4.1.2 Interface presentation*

This evaluation is realized on a first preliminary prototype of MAT containing only some of the most important features and a design not yet final.

The interface of the Manual Annotation Tool (MAT) (Figure 10) is mainly able to:

- Display a list of documents (panel 1)
- Display a video in a player (panel 3)
- Edit existing annotations through Europeana taxonomy (panel 2)
- Correct or complete annotations. (panel 2 & 4)
- Provide ontology and linked data to give semantic context for semantic tagging of media with concepts and named entities (panel 4)

This prototype is not already connected to ASSETS database at the time of the first evaluation.
This prototype could not allow users to search document or display a list of document. It could neither provide save functionalities. As ontology and thesaurus were not made available to the interface, local taxonomy is not connected to them.

This prototype is connected to DBpedia, Freebase and Geoname to let the user find semantic information.

The interface has been developed with eclipse, so the panel's size and position can be modified by the user.

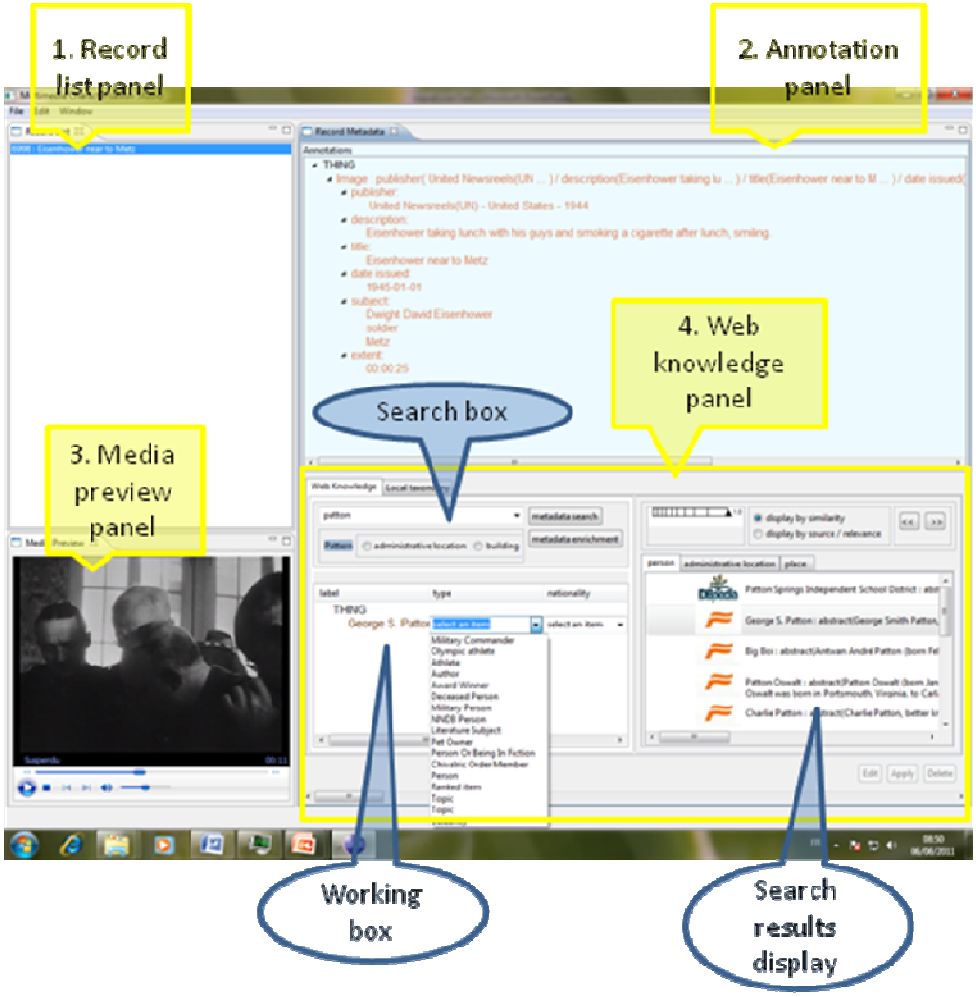



**Figure 10 First prototype of MAT interface**

### *4.1.3 Summary of usability issues found*

The summary of usability issues is extracted from all the results obtained during the evaluation completion. All the results are detailed in the **annexe 2**: **scenario reporting with snapshots, additional user manipulations, questionnaires compilation**.

**P - Priority Legend:**

- **Hi**gh = Critical problem, Task cannot be completed
- **Me**dium = Medium problem, Task completed with significant effort and failed attempts
- **Lo**w = Minor or cosmetic problem, Task completed with minor complications and/or annoyance

Task1: select a document

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 1 | *Record list* panel takes up too much place | Panel not always displayed | 2 | Lo |
| 2 | No (advance) functionalities to search file (BUT *it was not to be test here*) | | 1 | Me |

Task2: read, edit, correct annotation

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 3 | Notice not easy to read: no distinction between field, police size too big, | To modify the display style | 2 | Me |
| 4 | First line of the annotation not useful | To verify with ASSETS users | 1 | Lo |
| 5 | Difficult to find the edition mode | Edition mode in the same window than display (with management of edition right) | 3 | Me |
| 6 | Difficult to add, suppress, correct a term – Too much interactions (ex: 3 panels for one action) and no homogeneity between field edition management. | To modify the interaction mode, see traditional annotation interface. | 3 | Hi |
| 7 | Two edition modes | Only one edition mode | 2 | Lo |
| 8 | In the edition mode, the proposed fields couldn't be displayed by alphabetical order. | See 6 | 1 | Lo |
| 9 | No confirmation asked in suppression mode | Add one | 1 | Me |

Task4: Search in Web knowledge

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 10 | The text entered in the box is automatically changed and the information given by the user is lost | Bug to resolve | 2 | Me |
| 11 | Sometimes the user can't enter a new word in the search box | Bug to resolve | 2 | Hi |
| 12 | Pattern not easy to read and select. Not all pattern useful (only *place* and no: *administrative location, building, geographic location*) | List display? | 2 | Me |
| 13 | Entity/keyword selection hidden if the screen is not enough large | To change the disposition of the menu | 3 | Hi |
| 14 | No suggestion mode when entering a term in the search box | Functionality to add | 2 | Me |
| 15 | Arrow not well visible in the results list | Design modification | 3 | Lo |
| 16 | No possibility to stop the search | | 1 | Lo |

Task4: View results list

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 17 | Arrow not well visible in the results list | Design modification | 3 | Lo |
| 18 | Lines too long (always need browsing) | | 2 | Lo |
| 19 | No automatic (or easy) display of corresponding web pages when passing the mousse upon line of result. Too much interactions: necessity to put the line in the working box | To add web page display as verification mode. | 3 | Me |
| 20 | No indication of time wait to obtain the results | | 1 | Lo |
| 21 | Results tabs not always corresponding with selected pattern | | 3 | Lo |
| 22 | More pertinent results are not always at the top of the list (*it depends of the exemples*) | | 1 | Me |

Task4: Import result in working box

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 23 | For the first time, difficult to know that the result line need to be dragged and | But could be learned | 3 | Lo |

| | | | | |
|---|---|---|---|---|
| | dropped in the working box | | | |
| 24 | Some place left in the working box | | 1 | Lo |

Task4: Verify content in the media panel

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 25 | Difficult for the first time to know how to visualize a content | But could be learned | 3 | Lo |
| 26 | Very difficult to drag and drop the line in the media panel window | | 3 | Hi |
| 27 | No functionalities of classical web browser (back, forward..) | Prefer to have a separate but complete web browser | 3 | Hi |
| 28 | No highlighted searched term in web pages | | 3 | Lo |
| 29 | Function of searching a term in the web page only accessible in shortcuts (ctrl F) | Menu | 2 | Lo |
| 30 | Video lost when web pages are displayed | Two tabs: one for video display and one for web page display | 3 | Me |
| 31 | Automatic web page display (from action in the working box) not always pertinent (ex: military person). | To change or suppress the automatic load of web pages | 3 | Me |

Task5: Modify fields in the working box

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 32 | It is not always possible to suppress a field | Bug to resolve | 3 | Me |
| 33 | When a field has been displaced, the remove action suppresses the previous field which was positioned in the place in spite of the new one. | Bug to resolve | 2 | Me |
| 34 | No possible to select two fields for one entity | | 1 | Me |

Task6: Import these description in the annotation

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 35 | Not easy to create different lines containing the information. | | 2 | Lo |
| 36 | Bug: once, the imported line has been placed at the top of the notice up to Image label | Bug to resolve | 1 | Me |

General remarks

| # | Usability issues | Suggestions for solution | Nb of users | P |
|---|---|---|---|---|
| 37 | One non useful slider at the bottom of the interface | Suppress it | 1 | Lo |
| 38 | Cursor always displayed as active | | 2 | Me |
| 39 | Names are missing to recognize different panels especially when they are iconified | | 2 | Me |
| 40 | Design and organization of the general could be improved (left space and too much sliders) | | 3 | Me |

### *4.1.4 Analysis*

The analysis is realized from the cognitive walkthrough completion, from the questionnaires and also from discussions with the users and from additional tests done by users.

Three level of the interface is discussed.

**1. Functions level**

The MAT functions allow the archivists represented in this evaluation to achieve the given task. The user is able to find a named entity in semantic database and to import it in the annotation.

The MAT interface provides ontology and linked data to give semantic context for semantic tagging of media with concepts and named entities. This is a new functionality for archivists which could provide a real advantage. Users don't need to find specific information in the web pages. In fact, the automatic structuring of the information proposed to the user helps the understanding and the labelling of the document: it may be an important time saving function. The web page reading is only required as verification step. Document annotation contains two main steps: enter semantic categories and references to position the document in the archive, and to summarize the content. This evaluation shows that semantic data provides pertinent results in adequacy with user desiderata for the first step. This function seems very suitable, and in fact, users were enthusiastic. They would have liked to test it more deeply and they have made a lot of suggestions.

The more important recommendations that could be analysed are:

- To have classical web navigation functions in the web display (back, forward, highlighting searched term in the web display…)
- To have the functionality of suggestion when entering search terms

On the other hand, the modification of the annotation is not efficient and should be improved:

- too much information is given while clicking on a word (list of Europeana fields, 50 fields not by alphabetical order….)
- interface is too complicated and not homogeneous: for example, modifying a date requires 4 operations : click, select the field, click on "date property" "+", that lunch a popup windows that contains the "enter date field"

- field manipulation functions : functions to modify, move and separate fields are not available, delete function : exists but is difficult to use and does not ask for confirmation !
- intermediate save for work in progress.
- final Save/Export in the database

The working area should also be improved:

- Selection of Web knowledge: users ask for multi-selections in the label fields. For one label it was not possible to select two values (i.e. for the label “military personnel involved” we can choose only one name)

In the search box

- There are too many patterns proposed and they are not homogeneous (*geographic location* and *administrative location* are included in *place*…).

The selection of documents has to be improved

- Selection of a set of annotations : Annotations should be searchable through a classic search engine that returns a list

The video display incorporates only basic functions but they are adequate for the given task as the user have not to create new extracts but only to correct the annotation of the current video.

**2. Interaction level**

Functions are not always easy to manipulate.

The interface look-and-feel and basic environment interaction are based on eclipse environment. Users were not accustomed to this technology but they can exploit the interface without using all the interaction proposed by Eclipse such as modifying windows size and position. The only problem providing by Eclipse was the drag and drop interaction to load data in the working area and to load a web page. These interactions are not in adequacy with the habits of the users.

There are often too many interactions to realize one action and there are also some bad interactions. A lot of recommendations have been suggested by the users:

- To improve the editing functions of the working area because interaction is not easy (difficult to suppress a property)
- To improve the importing function from the working area to the annotation (creating several lines…)
- To load the web page quickly and easily when checking a result row without having to go through the working box.
- Better management of automatic pages switching as the web page displayed is always changing depending on the working box. Often irrelevant pages are displayed as generic page (creator, ... ..)

**3. Visual design level**

The visual design of the different windows is easy to understand but not completely relevant because the user needs often to browse otherwise there are some space left. The interface includes all the displays inside even the web browser but to the detriment of navigation options of the web browser.

A lot of recommendations have been expressed.

- To name the windows
- To display video and web page in two separate windows
- To diminish the selection window and enlarge the search window
- To modify the display of the annotation (different size or color for the different fields…)
- To correct the displays when there are too much lines to browse
- To improve the pattern list which is not easy readable

### *4.1.5 Conclusions*

This evaluation was designed to study the usability of the interface MAT in the context of a scenario of modification of video annotation. This study demonstrated that the results of the evaluation of all the functional parts of the interface were generally very positive but that the ergonomics were more negative. The tests also identified several very specific problems that were identified by multiple users "observer".

Usability is characterized by the following qualitative measures;

1. **Effectiveness** is positive for the search task but more negative for the modification of annotation task. All the predefined tasks could be realized except the suppression of term in the notice. The search module proposes new functionalities very useful and effective to rapidly obtain preformatted named entity. On the contrary, the annotation edition module is less competitive than traditional system. Some functions are also missing for the annotation edition as well as for the web page display.

2. **Efficiency** is also positive for the search task but more negative for the annotation modification task. The search of term in semantic database could diminish the number of interaction to import entity names in notices. It was easy to realize the search task, except to load the web pages. The comparison made by the users with their previous experiences is positive except for annotation edition.

3. **Satisfaction** has a right level but could be ameliorate. Feedbacks given by the users are positive especially for the exploration of semantic database. A lot of remarks have been expressed about missing functionalities they used to have especially for the edition module and about the design.

There is a strong expectation from the users on this type of functions that provide them a real help to collect very precise information already formatted.

## 4.2 MAT Manual Annotation Tool (prototype integrated to ASSETS database)

### *4.2.1 Introduction*

This chapter presents the evaluation of the interface MAT (Manual Annotation Tool) connected to ASSETS DB. A first evaluation has already been performed on an early stand-alone prototype. This evaluation validates the quality of use of the annotation service proposed by the MAT interface, checks the improvement according to the recommendations collected during the first evaluation and verifies the connection access to the database.

***Evaluation elaborated by***: A.Saulnier ( Ina)

***Technical contact for the service***: P.Courounet (Ina)

***2 User tests performed at Ina by***: two users of Ina

***8 Remote users test performed***: two users from FLM, one user from DW, three users from HASC, two users from CVCE

***Date***: from February 22sd to March 16th 2012

**Objectives**

The objectives of the evaluation are:

1. To analyze the usability of the functionalities of manual annotation
2. To test the integration of MAT in ASSETS platform
3. To verify the improvement of easiness of access (amelioration of MAT interface since June)

The scenario is derived from the technical use case of D2.0.1. The goal is to verify that MAT allows professional users to perform their tasks with quality, easiness and satisfaction:

- To search documents from the database
- To display document list, document and annotation in viewers,
- To edit, to correct and to back-up copy of annotations,
- To obtain a semantic aid to retrieve named entities

Limitation of this evaluation:

We are not evaluating the annotations for propagation.

We are not evaluating the access to the taxonomy in this evaluation.
It will be done in a separate evaluation.

**Methods**

Several tests are performed:

- **"User tests"** are conducted at Ina with a supervisory control allowing collecting observation information.
- **"Remote user tests"** are conducted in content provider offices.

All the user tests are planned with similar protocol:

- Presentation of the interface (verbal presentation or written presentation) with a walkthrough example to do.
- Annotation of the same set of notices according to precise instructions.

- Collect of results and data linked to the task performance (by the observer or by the user).
- Additional free tests when user wants to do it.
- Questionnaires to fill.

**Users**

Users involved in the tests are content providers from FLM, DW, HASC, CVCE and INA.

Their profiles are:

- 9 annotators
- 1 research project manager in charge of development of archivist tool

They are all involved in the domain of document annotation.

Three of the users are involved in annotation activity since less than one year, the other are experimented.

None of the users is accustomed to work with semantic database.

### *4.2.2 Interface presentation*

A complete walkthrough scenario can be found in D2.5.3 Multimedia characterisation, Appendix 1.

The interface of the Manual Annotation Tool (MAT) (Figure 11) is mainly able to:

- Search and save document annotations.
- Display a document (panel 1)
- Edit existing annotations through Europeana taxonomy (panel 2)
- Correct or complete annotations. (panel 2 & 3)
- Provide ontology and linked data to give semantic context for semantic tagging of media with concepts and named entities (panel 3)
- Verify named entities in web documentation (panel 4)

This prototype is connected to DBpedia, Freebase and Geoname to let the user find semantic information.

The interface has been developed with eclipse, so the panel's size and position can be modified by the user.

**Figure 11: MAT interface (second version)**

### *4.2.3 Scenario description*

**Scenario: Manual completion/correction of ESE records with "web knowledge help"**

Context of the scenario: manual correction or completion of annotations (ESE metadata file) belonging to document which has been already sent to Europeana ingestion.

The scenario contains a specific description of several representative tasks of correction of

annotations to be performed with the system.

Data for the test are a set of 5 images belonging to the database of Google 3D model (same images for all the users).

The objective for the user is to find one or several missing entity names to complete an image notice. In order to assure a comparison of use of the tool between all the users, the missing entity to search are specify in the instructions given to the user.

This scenario had been structured into 5 tasks:

- **Task 1**: controlled task where the users have to follow precise directives letting him testing all the functionalities of the service while searching the location and architects of “Peoples Heroes Monument” to complete the annotation.
    1. Search of document
        - Search the document “Peoples Heroes Monument” and open the record.
        - View the media and metadata.
    2. Search on the web knowledge panel
        - Search on the web knowledge “people heroes monument” as “building” type
        - Chose the document about the monument of China in the “building” tab
    3. Load document in the working box
        - Drag and drop the chosen document in the working box
        - View the corresponding page in the Web documentation panel
    4. Direct modification in the record metadata panel
        - Add a field “alternative title” in the annotation
        - Enter with the keyboard this sentence “Monument to the People's Heroes, Beijing” in the alternative title of the annotation
        - Add a value field in the “descriptive” field of the annotation
        - Complete the description from the web page with a copy and paste
    5. Modification from the web knowledge panel.
        - Configure the categories in order to add in the notice two fields (illustration provided):
            - “Contributor” field with the name of the two architects
            - “Spatial coverage” field with Beijing, Tiananmen Square names
        - Integrate these semantic data in the annotation with the “metadata enrichment” button
        - Verify the modification done in the annotation
    6. Save

- **Task 2**: Users have to complete the annotation of the “London-eye” contained in the 3D Google database to add location information. They can also add some other information and clear some fields in the annotation if they wish.

- **Task 3**: Users have to complete the annotation of the “Colosseum, Rome” contained in the 3D Google database to add building functionality information.

- **Task 4**: Users have to complete the annotation of the “Brandenburg Tor” contained in the 3D Google database to add location and architect information.

- **Task 5**: Users have to correct and complete the annotation of the “Birds nest” contained in the 3D Google database.

### *4.2.4 Report of the results*

Five tasks of annotation correction have been defined to test the use of the service. As similar tasks have to been proposed to the user, the annotations have been chosen from Google 3D database belonging to none of the user’s company. The first task (directive task) is very well detailed to guide the users to perform this task unlike the four other tasks (free tasks) which only give general instructions to the users. They have to find named entities to complete some annotations. They can use the web knowledge functionalities to find missing information and use the basic functionalities of edition if needed.

Users had more or less difficulties, but they all understood the different ways of working offered by this interface.

To summarize the final results obtained by the users (Figure 12):

- The London-eye task: everyone has found the location to complete the record. For one user, the field has been dispatched into *subject* field instead of *spatial coverage* field but he was able to manually correct this error. Another user has just made a change of layout in the instructions to shorten a line. Regarding the search for web pages, one user didn’t notice that he had mistakenly selected the item "any type", so he had to make several attempts before he realized why he couldn’t find the correct results. This problem could occur when users do not know the interface and do not read the manual. Another user has had trouble dispatching the result. In fact, it is very interesting to notice that a user has use the system in a different way that it was supposed to be. In spite of searching “London-eye” in the web knowledge to find the exact place, he searched London. Then the term London was automatically dispatched as “subject”.

- The Colosseum of Rome task: A user does not find the output because it has not found the right web page containing the information and therefore he couldn’t perform properly the following operations. Another user has trouble because he has dispatched « Colosseum » as subject.

- The Branderburg Tor task : Everybody has found *Berlin, Brandenburg Gate, Carl Gotthard Langhans*. One dispatching was not perfectly done. One user propagates the “architect” as “subject” and has to correct it.

- The BIRD’s nest task: A user does not find the result because it has encountered many technical problems. Other users have found it difficult to quickly identify the right page web. This task was more difficult because users had to find that Bird's Nest was the name of Beiijing national stadium. Suggestion was very effective in this task because it propose “Beiijing national stadium” while entering “birds” in the search box.

Some meaningful percentages can be calculated from the performance results:

- 90% of users successfully completing their tasks: 9/10 users have achieved all their tasks with success.
- 95% of goals achieved: 38/40 goals of annotation completion have been achieved (with dispatching from working box and/or manual edition functionalities).
- 87,5% of the propagation of the selected named entities has been correctly automatically dispatched in the metadata from the data defined by the users in the working box.
- Very good average accuracy of completed tasks: all the annotations which have been completed by the users are accurate because if the users faced some problems with the automatic propagation they could manually correct the errors imported in the annotations. There is no wrong information entered in metadata.
- Number of searched done in the web knowledge in order to find the asked information:
    - 20 tasks have been achieved in 1 search step
    - 10 tasks have been achieved in 2 search steps
    - 5 tasks have been achieved in 3 search steps
    - 5 tasks have been achieved in more than 3 search steps

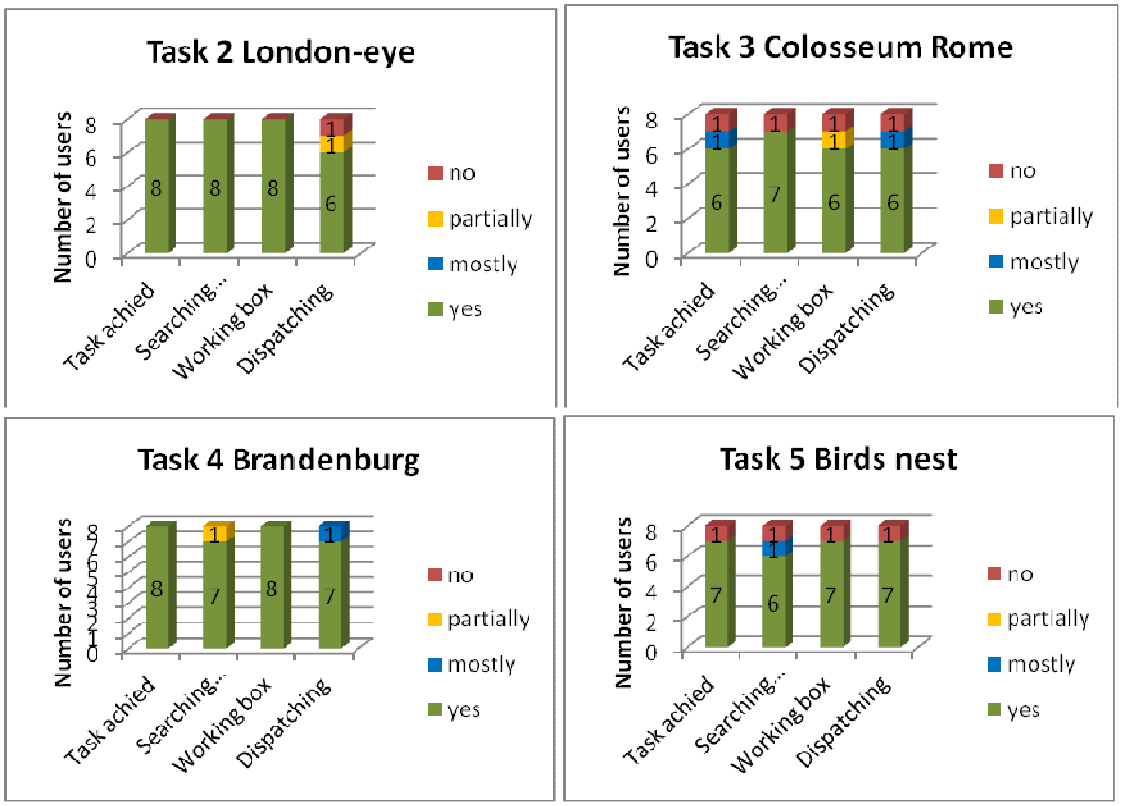



**Figure 12 : Summary of the results obtained for each task.**

Regarding the realizations of the tests, several problems have been reported.

Firstly, several technical problems have been noticed. They could be due to the connection to the web knowledge database or to the computer power. Sometimes system answers were very slow to

give suggestions or to give back results of web pages. Sometimes users have to restart the system after an annotation task. These connection problems made some manipulations difficult to realize and probably in some ways have a little biased the result sets of evaluation in terms of usability. Due to the slowness of executing the result of drag and drop, a user has tried to repeat the operation a couple of times, thus causing the lock of the program. He had to close the annotation and reload in order to proceed in the test.

Secondly, some problems of access to ASSETS test database have been noticed. Few times, users have found difficulties to search a new annotation metadata on the database. They couldn't find the results of a request with two terms or they received answers about the last request. In these cases, users have do try several times to search the annotation metadata in the database before to open it.

Thirdly, some problems appeared when users would like to search an entity name with an apostrophe in the menu of the web knowledge panel. Moreover, the suggestions proposed names with apostrophe which couldn't provide any results. So, in these cases the users have to enter another request formulation without apostrophe. This problem has probably bothered a lot the user who was unable to perform all his tasks.

Few other problems could be reported. There is no "undo" functionality to apply when you have done a wrong action in the edition mode of the record metadata panel.

Questionnaires let us collect the user's perception of aspects such as usefulness, easiness of manipulation, effectiveness and learnability.

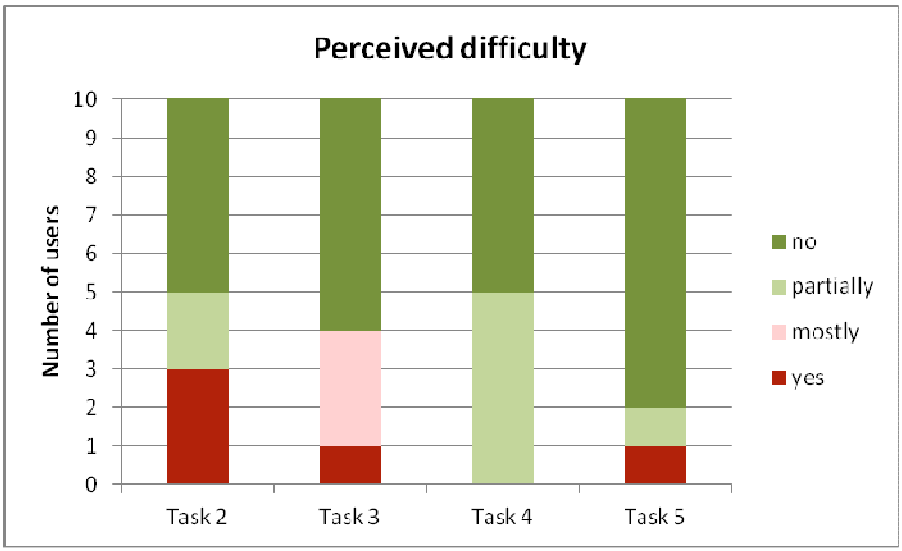



**Figure 13 : Perceived difficulty to realize each task**

Perceived difficulty is less important at the end of the evaluation than at start. For the first task, two users had difficulties but achieved their tasks with the manual edition functionalities to correct the automatic dispatching. One user faced difficulties for task 2, 3 and 5 and didn't achieve his task. The perceived difficulty is correlated to the quality of results obtained by the users.

The user's perception of the service (Figure 15) is globaly good because all the marks given by the users are above the mean. *Useful* is related to the interest of the user. This service is very useful to realize annotation task. *Esay to use* is related to the manipulation difficulties encoutered to achieve the task. *Easy to use* criteria of the service is the lowest measure and several recommandations have been expressed to improve the ease of use. This measure contains information gathered from interaction level (i.e formitting line size in record metadata), as well as from visual design level (i.e.

size of working box panel too little) or conception of the design (i.e. open file in Annotation/New Annotation tab) but it is also influenced by underlying technical problem. *Effective* caracteristic is related to the performance reached by the user. The lowest performnance is for the web knowlege panel which contains the more innovative functionalities.

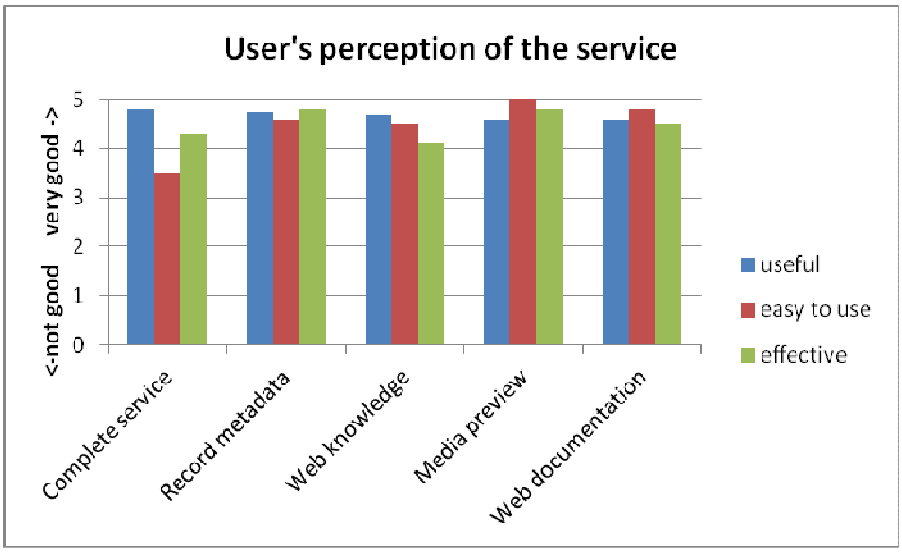



**Figure 14 : Users' perception of MAT functionalities**

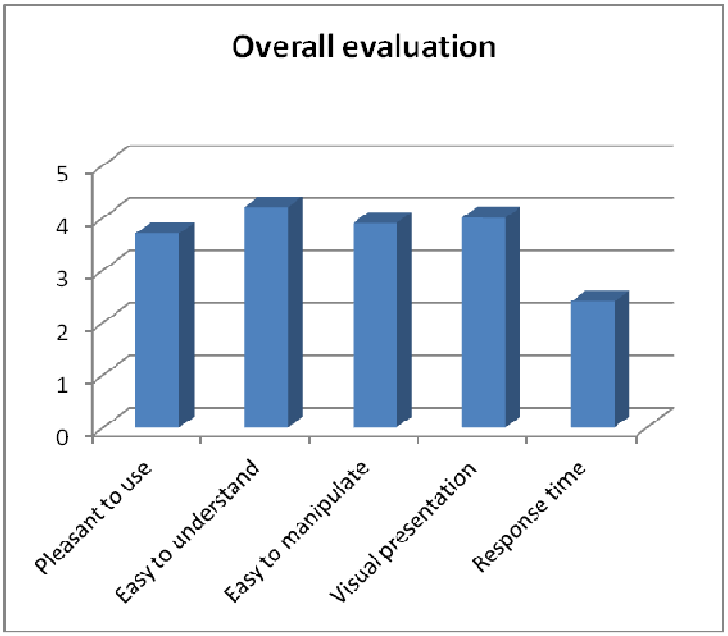



**Figure 15: Users' perception of MAT overall evaluation**

About the overall evaluation, response time is the only criterion which appears as critical. It illustrates partly the technical problems encountered during the evaluation.

**Improvements since the first version**

During the evaluation of the first prototype of MAT, 40 recommendations have been reported. This new evaluations have been performed with two of the users who have done the first evaluation.

The more important recommendations have been realized. Classical web navigation has been incorporated in the interface. The edition mode has been simplified. Intermediate save is possible. The working area has been improved. Suggestions have also been inserted in the search box, it may return highly relevant results but this functionality is often working too slowly.

Only 4/40 recommendations have not been solved:

- number 9: No confirmation asked in suppression mode (priority medium)
- number 15: No possibility to stop the search (priority low)
- number 16: One sign "+" not well visible in the results list (priority low)
- number 19: No indication of time wait to obtain the results (priority low)

### *4.2.5 Analysis of the usability*

**Effectiveness**

Measures of effectiveness relate the accuracy and completeness with which the users can achieve their goals as well the compliance with the user need. Effectiveness relies also on technical criteria related to the system performance as reliability of the service.

In this evaluation, **effectiveness** is measured by how many annotations have been correctly completed. Figure 12 indicates corresponding results. 90% of users successfully completed their tasks and 95% of goals were achieved. The average **accuracy** of completed tasks was perfect. These are positive results, especially taking into account that the user who has not met its objectives reached many technical problems. All the functions proposed in this service were in adequacy with the user need in term of **usefulness**. Several functionalities are proposed to the user to complete an annotation and their complementarity is a real advantage of this service. The edition functions of the record metadata panel are in adequacy with traditional service of annotation while the web knowledge panel provides a more advanced functionality helping the user to find named entity. If the user meets problem to organize the data in the working box, he can always consult the web page in the web documentation panel and make a copy-and-paste for example. We can observe different ways to use this service: 1) classical edition mode of metadata annotation; 2) automatic dispatching of metadata annotation with named entities found on web knowledge; 3) automatic dispatching with manual correction; 4) copy and paste from web page into the annotation. All these functionalities have been used by the users and this complementarity ensures the users to be able in nearly all cases to correct metadata.

So all the **functionalities** proposed in this service are in **adequacy with the user's need**. Nevertheless, one missing function has been notified. There is no undo functionality to let the user come back when he has made an error. Then the **suitability** is good except for this point.

These tests also give rise to several technical problems relative to the service **maturity** belonging to reliability aspect. A first problem of connection to ASSETS test database has immediately been resolved after the evaluations. The API of access to the ASSETS database has been replaced. The second problem of the apostrophe for the search in the web knowledge is now solved. The response time of the service is also not in adequacy with the user need, but this point is relative to a lot of parameter as computer performance, connection ( (to the web, to the server).

**Efficiency**

Efficiency relates the level of effectiveness achieved to the expenditure of resources. Here expenditure of resources is analyzed from mental effort. Time could not be a criteria of analysis because remote tests are not conducted in the same material condition (computer performance, connection) so we can't compare the users performance. Time is also dependent of the step of learning and experience of each user. So, for this test, efficiency is measured by 1) number of searches in web knowledge; 2) understandability; 3) number and type of errors; 4) the perceived difficulties; 5) the adaptability to user experience.

Time efficiency cannot be calculated from the user performance but several users have expressed the fact that the answer of the system was too slow because of problem of connections to web knowledge. The problem of the **response time of the system** was the worst aspect of the system from the user point of view (Figure 15). This technical problem prevented to achieve a good efficiency for the user even if he has no problem of manipulation of the system.

The number of searches in the web knowledge varies from 1 to 6. 6 searches have been realized when the user faced technical problems. 35% of the searches reached the results in one launch and for the fifth task 50% of the searches are done in one search. We can notice that the learnability progress have been observed during these tests. This is confirmed by some user commentaries which expressed some difficulties to understand the management of the labels in the working box at the beginning of the test and rapidly controlled much better the choice of labels. So this interface is not completely intuitive but **learnability** is rapidly improved with manipulations. Moreover, when the user succeed to organize the labels in the working box and enriched properly the annotation, this operation needs less information to read than for classical system where the user needs to read a web page to find information and then do a copy and paste which remains anyway possible.

**Understandability** is related to the learnability and understandability measurements. It can be indicators of the learnability potential of the system. Figure 15 shows that users have globally a good perception of the understandability of the system, except for one user. For the input actions, users need a period of time to understand the management of the working box. One misunderstanding has been noticed. Users do not understand that the first line of the working box was supposed to be ventilated in subject, which proves that this assumption was false. It could be perhaps more adapted to propose the way to dispatch the information in the metadata annotation than to do it automatically. For the output actions, according to the examples, users may have some difficulties to understand which result he has to select to find the appropriate web page. So for output action, understandability may be more dependent of the context of the task. Sometimes, the system gives not enough feedback to the user because there is no enough indication of wait during the actions of the system.

Some other manipulation errors illustrate the need of **learning** before to use the system with efficacy. About the search step in web knowledge, two users have forgotten to fix the type as "building" and then have not search the results in the right tab. But when users have discovered the problem, they haven't done this interaction mistake a second time. A problem of guide design has been noticed. The opening of annotation was difficult to find. Also, visual design requires the users to browse to often in the web panel box. Observations have been collected in the recommendations (see below).

The users have expressed the difficulties that they perceived for each task (Figure 13). These perceived difficulties reveal the **cognitive effort** developed by users. If we expect the user who met a

lot of technical problem, no user have perceived strong difficulties for task 3, 4, 5. So after the familiarization step, no difficulty has been reporting about manipulation problem. Figure 14 reporting the general user's perception of the system is not so good because all these resulted are influenced by the technical problems (connection to ASSETS database, connection to web knowledge). We can precise that there is a huge dispersion in the collected measures collected of the user's perception about "easy to use" criteria. So this perception was almost bad for one user and very good for two users.

This service also proposed several advanced option such as ("search as" from the annotation metadata) to fasten the use for advanced users. These options have been tested only by one user and provide a good adaptability of the system according to the profile and user experience. They let the user work faster.

**Satisfaction**

Satisfaction describes the comfort of use. The measures characterize the user satisfaction to execute a task. These measures are extracted from the questionnaires. The users find the service very useful, almost effective and in a lesser extent easy to use.

This system is very useful firstly because the system caters for immediate Europeana compatible metadata without the user having to know the Europeana standard. It is very useful for leading through the process of dispatching. Metadata dispatching is a powerful tool. Users liked very much the interaction between record metadata window and working box. One user remarks that this service is much more opened than the system they used "at home" to complete.

The possibility to obtain suggested terms through the web knowledge panel is very effective and can greatly facilitate the annotation task. The major problem is the slowness of search observed sometimes (for obtaining results of web pages and for drag and drop of one result line in the working page). Connections to ASSETS db was not always properly working. Sometimes also searches were not reliable (because there is no pertinent information in the web knowledge corresponding for the search). In fact, there is a technical constraint imposed by semantic database. Response time is longer for searches made inside structured and typed database than for searches made with Google on the web. A long-term solution might be to get the databases locally.

Users like the overall layout mixing several functionalities, but they found the box for the web documentation too small. Some information are hidden. Anyway they like the results with image, the easy and effective drag-and-drop function. The perception of ease of you of the working box was not the same for all users. All users found that the functionality of labels is very useful, but some said that it is not very user-friendly. For other, as this functionality is innovative, it just requests a short learning period. The system is not complicated but it does require a phase of learning for operators.

### *4.2.6 Recommendations*

High level

- To Improve the slowness of performing some operations (too long to search on web knowledge, program not reactive enough to update metadata, …)
- To Improve the relevancy of searches
- To resolve the problem of connections with ASSTS DB (already resolved – new API given by AIT)

- To resolve the bug due to apostrophe (already done since the evaluation performances )

Medium level

- To propose to the user to control the dispatching of the information collected in the working box into the fields of the annotation (automatic proposition to be validated or corrected by the user)
- To add *undo* functionality
- To improve the windows management (independence of annotation panel and web knowledge panel in order to spread the full width box working)

Low level

- To display indication of waiting when the system works (at least a small animation)
- To modify the denomination of “new annotation” for the opening of metadata annotation, same for saving.
- To enlarge the images of the web knowledge panel
- To automatically format the length lines in the annotation panel
- To let more intuitive to find in the edition mode how to add a field in metadata panel

### *4.2.7 Conclusions*

This service has been evaluating as a powerful system for the user but it still needs some additional modifications to improve technical performances before to be use in a real environment. The major quality of this service is that it is very useful for leading through the process of dispatching; the automatic extraction of knowledge from web pages is really an effective idea, which can greatly facilitate the annotation task. Moreover, the user doesn’t need to know the Europeana format to complete/correct a metadata annotation. The program is not complicated but it does require a phase of learning for operators, including those who have a certain expertise in metadata annotation.

## 4.3 MAT propagation (automatic CERTH Propagated annotation with MAT)

### *4.3.1 Introduction*

This chapter presents the evaluation of the usability of the visual propagation service through the MAT interface. This evaluation is focussed on the usability of the system offered to the professional users who will have to prepare a corpus of images for the propagation system. MAT is used to enhance the attribution of "semantic tags" on a selected corpus by indexing manually some representative instances of this corpus and propagate this indexing to the remaining of the corpus. MAT behaves essentially as a GUI to facilitate access and tuning of the propagation service.

This evaluation will complete the evaluation of MAT done in chapter 4.2: evaluation of MAT in chapter 4.2 is focussed on the correction/completion of a document whereas the evaluation of MAT in this chapter is focus on the manual enrichment of a corpus.

This evaluation deals with the analysis of the service offered to the professional user to achieve his task but will not deals with the quality of the enrichment.

***Evaluation elaborated by***: A.Saulnier ( Ina)

***Technical contact for the service***: M.Lazaridis (CERTH-ITI), P.Courounet (Ina)

***2 User tests performed at Ina by***: two users of Ina

***2 Remote users test performed***: two users from CVCE

***Date***: from February 22th to March26th 2012

**Objectives**

1. Test the usability in terms of
   - Accuracy and completeness with which users achieved the task of propagation preparation
   - Easiness of achievement
   - Satisfaction of the interface

   of the functionalities relative to
   - The corpus selection
   - The manual attribution of semantic tags (with the use of a taxonomy)
   - The consultation of results
2. Test of the complete chain of propagation (from corpus selection to propagation in ASSETS database)

Limitation of this evaluation:

- This evaluation is not focused on the quality of the ASSETS taxonomy.
- This evaluation is not focused on the quality of the image similarity service. For this aspect, this evaluation has also been subject to "technical evaluation". Quality evaluation of the results has to be done for a large corpus which can't be treated as part of these tests. Result quality belongs to corpus size and quality of the training set (according to visual criteria analyzable by the system).

**Methods**

Several tests are performed:

- **"User tests"** are conducted at Ina with a supervisory control allowing collecting observation information.
- **"Remote user tests"** are conducted in content provider offices.

All the user tests are planned with similar protocol:

- Presentation of the interface (verbal presentation or written presentation) with a walkthrough example to do.
- Test a limited corpus
- Questionnaires to fill

**Users**

Users involved in the tests are content providers from CVCE and INA.

Their profiles are 4 annotators.

They are all involved in the domain of document annotation.

### *4.3.2 Interface presentation*

Visual propagation with MAT

A complete walkthrough scenario can be found in D2.5.3 Multimedia characterization, appendix 2

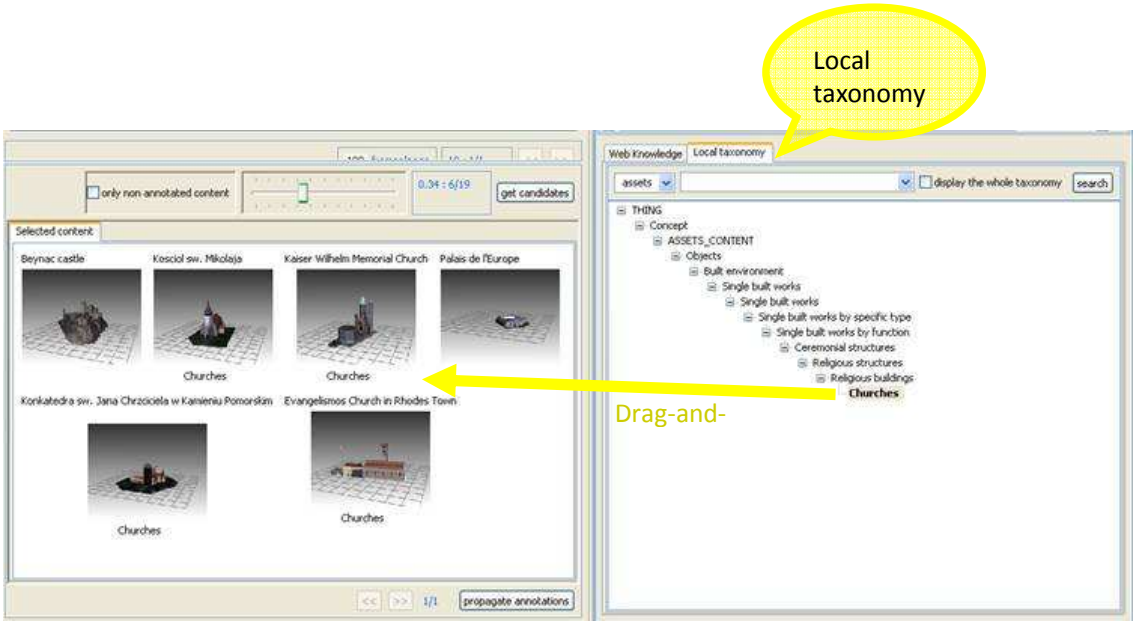



**Figure 16: MAT interface for propagation service**

Through this service the professional user is able to:

- define a corpus to work with
- manually attach "semantic tags" to a sub-set of objects of this corpus
- initiate the propagation of the attached annotations to the entire corpus
- corrects some results if necessary

This interface has the same look-and-fill than the one evaluated previously. All the tested functionalities in 4.2 can be accessible.

### *4.3.3 Scenario description*

The evaluation relies on a scenario of visual propagation on a corpus of 3D images extracted from Google 3D.

The scenario is structured into 3 tasks:

1. Selection of a corpus (between 30 and 40 images) images extracted from Google 3D database. Call content selection service to determine the candidates for the training set.
2. Attribution of semantic tag on the representatives instances of the corpus given by the service :
   - Select the size of the training set of images to annotate (a size between 30% and 40% is recommended but in the real case it depends on the homogeneity of the corpus)
   - Indexe by drag&drop of semantic tags directly on thumbnails from the taxonomy. User can give more than one concept for document.
   - Propagate
3. Visualization of the results
   - Take a look at the results
   - Correct some of the results (just to validate the easiness manipulation)
   - Save the propagation

### *4.3.4 Report of the results*

The proposed tasks correspond to a complete scenario to propagate visual semantic field on a corpus. All the users have already been involved in the evaluation of MAT connected to ASSTES db, so they had some prior knowledge on the graphical interface.

Users have completed all the tasks (Figure 17) and obtained expected results. No running error happened. One user finds the operation to erase 15 images too long to execute. No difficulty was reported (Figure 18). On the whole, users have no problems and find these tasks easy to realize. They succeeded to configure the interface and customized their work area. However, the user appreciates to position the window where he wants.

For task 1, users select images to define a sub-corpus of Google 3D. All useful functionalities were available. They were all easy to find, except the web documentation for which a user does not find it immediately. Users appreciated to remove a set of images instead of having to delete them one to one.

For task 2, all users found adapted semantic tag in the ASSETS taxonomy. Chosen tags were all about architecture “buildings, castles, tower buildings, skyscrapers, churches,...”. The terms were easy to find in the taxonomy and well adapted to the corpus. The two displays have been used. All users used the suggestion for query search in the taxonomy. For example while typing “tow” you have 5 propositions, not only beginning by these letters “tower buildings, radio towers, towels, wall towers, tower clocks”. Users also appreciated the pop-up display upon taxonomy term to provide additional explanation about the term. Then users gave tags to the images with drag-and-drop.

For task 3, all the users have succeed to check the results. They have no problem to see the results with their tags and to make some correction if necessary. There were no running or interaction errors. Users expressed a common desire to have feedback on images that have not been propagated by the system.

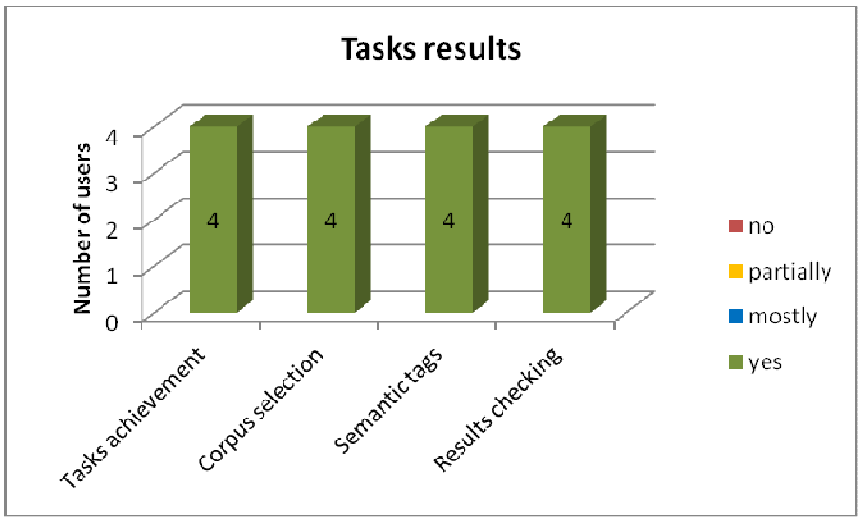



**Figure 17: Achievement of tasks for all users**

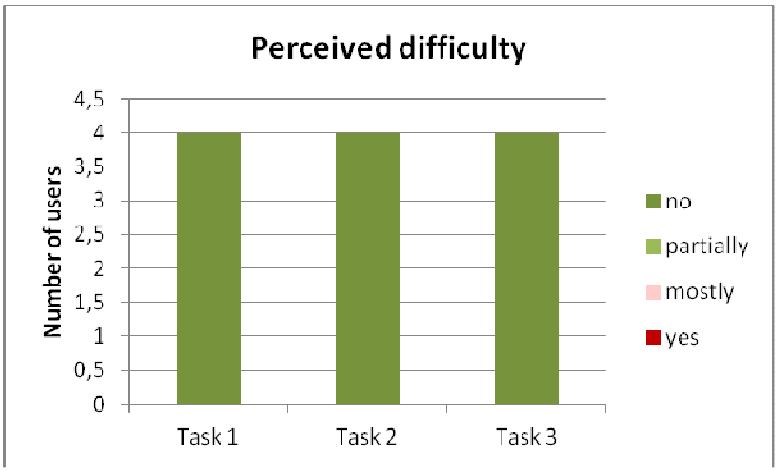



**Figure 18: Perceived difficulties for propagation service**

Some measures have been computed from exit questionnaires (see Appendix 4: MAT propagation).

The users' perception of the service (Figure 19)is globally very good because all marks are upper 4. All functionalities are useful for the user. "Easy to use"" critera has not the maximum mark of 5 because sometime users didn't find immediately a function (for example web documentation or remove tag). As well, efficacity has not the maximum mark of 5 because users think that the manipulation of the taxonomy could always be improved and for checking results users waste time to find which images had not be propageted. Some recommendations have been expressed to improve the system.

The overall evaluation (Figure 20) expressed by the users is also good. We can notice that for this use of MAT there is less problem of response time than for the evaluation of MAT connected to ASSETS DB . Interactions that must be repeated many (e.g. assign tags) are very fast.

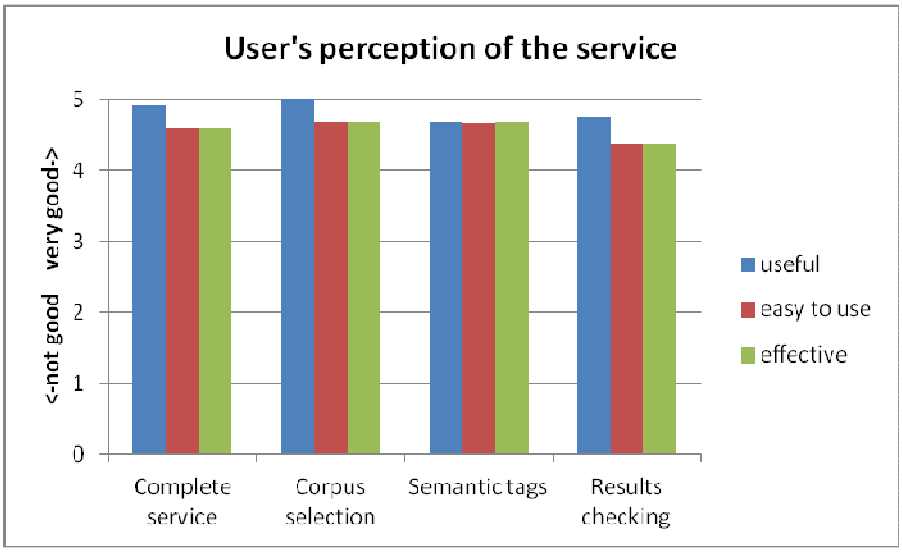



**Figure 19: User's perception of the propagation service**

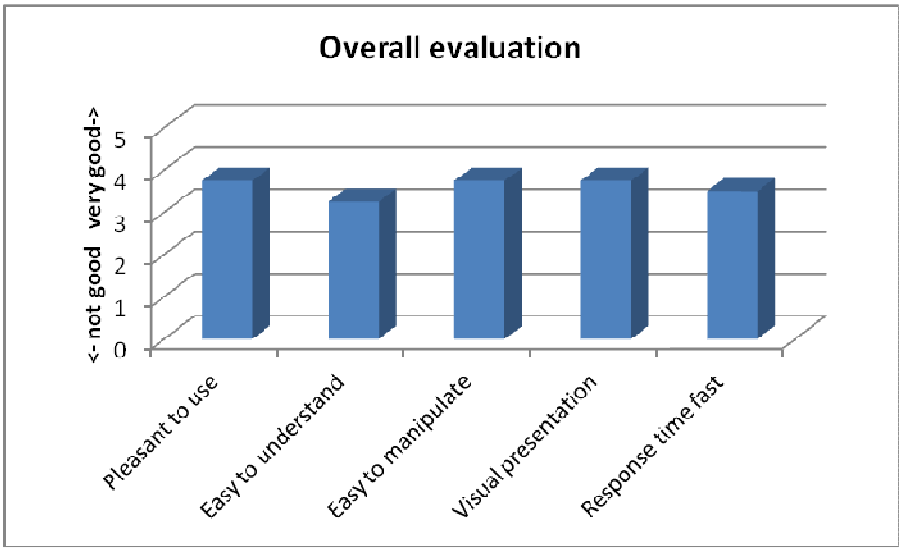



**Figure 20: Overall evaluation of the propagation service**

### *4.3.5 Analysis of the usability*

**Effectiveness**

Effectiveness relates the accuracy and completeness with which the users can achieve their goals as well the compliance with the user need. It is measured by the tasks results obtained by the users. Here 100% of the tasks have been achieved and 100% of the users succeeded to achieve their task. So the **effectiveness** is absolutely positive. It should be noted that these results are obtained from a reduced-scale test, but they are considered auspicious for a job in real conditions. Similarly, adaptation of the taxonomy is crucial for the success of this task.

All the **functionalities** proposed in this service are in **adequacy with the user's need**. Nevertheless, one missing functionality has been notified. Users would like to obtain information about the image which have not been propagated. The system should provide the list of documents that failed to be propagated so that the user can analyze the system's capabilities and complete by hand certain result if he wants. So the **suitability** is good but could be improved.

Moreover, the performance of the system was good. Drag and drop interaction time was compatible for users' actions but some operations are a little too long (as erase a lot of images). No running problem has been observed, so the **maturity** of the service has been also very well evaluated.

**Efficiency**

Efficiency relies on expenditure of resources which is here analyzed from mental effort criteria. It is measured by 1) understandability; 2) number and type of errors; 3) the perceived difficulties; 4) the adaptability to user experience.

Understandability is quite good but need a learning period with manipulation. During the tasks realization, all users understood the functionalities objectives as well as all the input and output provided by the interface.

Some output are missing to indicate when the system is working (i.e.. erase of images, propagation, …) or when an action has no result.

No error of any type was observed and response time was adapted to interaction operations. This contributes to a good performance of the system use.

There was no perceived difficulty. Selection of corpus functionalities is intuitive. Search into the taxonomy is very effective and easy to use. The two displays are useful. Suggestions are fast and very pertinent. All useful terms are returned. Then drag and drop is very easy and fast to attribute tag to the images. Moreover all needed functionalities are available by clicking the right mouse button, which is very convenient. The check of results task is also intuitive to achieve. So the functionalities are adapted to the users' experience.

**Satisfaction**

The system has been evaluated as relatively "pleasant to use" by the users (Figure 20). In fact users are not accustomed to this kind of interface and they need a period of adaptation. Anyway, the drag-and-drop functionality has been very well enjoyed by all the users.

### *4.3.6 Recommendations*

Medium level:

- Add marks to have feedback about images which have not been propagated inside the corpus defined by the user.
- Confirmation need when clicking on "Initialize corpus" button.

Low level:

- A little bug display: when asking two times the annotation showing of one image, the annotation appears and disappears.
- Put a visual marker to indicate when the system works
- When the system return no result , this has to be indicated by a short message
- “get candidates” on the button is not a very clear label

### *4.3.7 Conclusions*

The usability evaluation of the propagation service with MAT interface gives positive results about the work than professional user will have to do to manage the visual propagation process. However, these results demonstrate that the efficiency is very good on a small corpus. This should be confirmed on a larger corpus.

The verification step gives rise to a discussion as it is an enrichment process automatically. Users, as librarians responsible for data quality, all wish to access the results to verify them. This step seems to them much less restrictive and laborious than having to manually process a large corpus, especially if the results are sorted by levels confidence in the results.

As the taxonomy quality was not evaluated (and was supposed to be totally in adequacy with the visual image to annotate), it remains an issue to be addressed: what happens if a user finds no suitable descriptor in the taxonomy? It would require additional testing in relation with the scientific evaluation to assess the impact of the quality of the taxonomy on the visual propagation in a real case.

## 4.4 MAT enrichment (manual enrichment propagation)

### *4.4.1 Introduction*

This chapter presents the evaluation of the propagation of manual enrichment with MAT interface. This scenario has been developed at the end of the project and was tested on a prototype. MAT is used in a process allowing to enrich independently of any context some recurrent named entities and concepts. This feature is based on the assumption that for a given provider a same term keeps the same meaning in the corpus.

**Evaluation elaborated and completed by**: A.Saulnier (INA)

**Contact for the service**: P.Courounet (INA)

**1 cognitive walkthrough** performed at INA

**Users**: two Ina annotators

**Evaluation place**: Ina (France)

**Date:** March 22sd 2012

**Objectives**

The objectives are to verify that MAT allows professional users to perform their tasks with quality and easiness:

- To select a corpus
- To manually enrich named entities and concepts
- To propagate this annotation

**Method**

We have applied the same variant **of cognitive walkthrough** than for the first evaluation of MAT (see 4.1).

**User profils** are annotators or archivists used to create and modify annotations of document corpus.

The evaluation is conducted with **1 interface expert and 2 annotators** of Ina who have experiences in annotations and are used to search information on Internet for annotation task. They are not involved in ASSETS project.

This method is based on discussions and comments while achieving a scenario of use according to this set of questions:

1. Does the user know what to do to achieve the task?
2. Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?
3. Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?
4. Does the user understand the results given by the system? Does the result quality match the user expectation?
5. Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?

**Scenario**

As this module is quite recent and still more a proof of concept a quick overview of the scenario is given in annex 3 of D2.5.3 Multimedia characterisation.

**Measures and criteria**

This adapted cognitive walkthrough (realized with professional user) let obtain qualitative measure for each criteria of the usability:

- Effectiveness: verification that the predefined tasks could be realized
- Efficiency: comparison made by the user with his previous experience
- Satisfaction: feedback given by the user

### *4.4.2 Interface presentation*

This scenario of manual enrichment involves several functionalities of MAT interface (Figure 21):

- Selection of corpus
- Display of corpus
- Display of entities extracted from the annotation of the corpus
- Display of one metadata annotation
- Web knowledge display to enrich one entity
- Web page display

The manual enrichment may provide (Figure 22):

- some URL of readable pages
- a set of interoperable URIs (in various languages)
- a type compatible with high level EDM ontology and sometimes alternative labels
- geo-coordinates for location

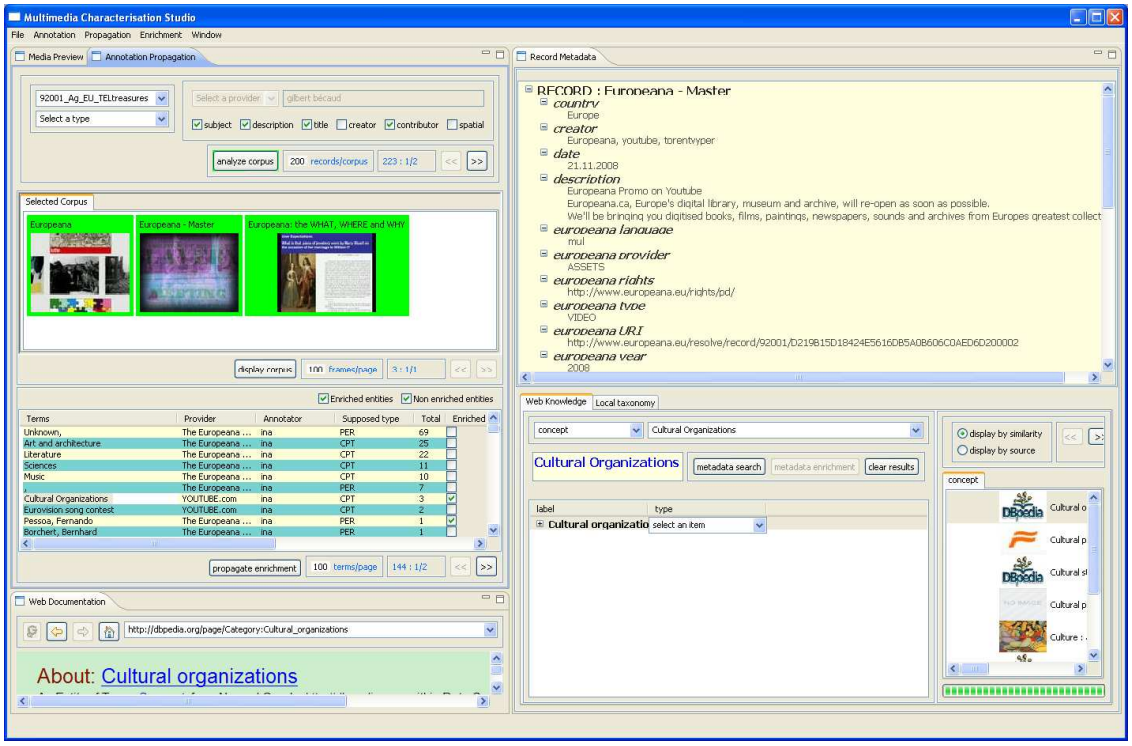



**Figure 21: MAT interface for manual enrichment scenario**

| Terms | F | S... | Total | E... | P.. | type | link | uri | long | lat | newLabel |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Palais Brongniart-Bourse de Paris | G. | i. | CPT | 1 | | | | | | | |
| Obserwatorium Wysokogorskie na Sni... | G. | i. | CPT | 1 | | | | | | | |
| Newgrange Passage Tomb | G. | i. | CPT | 1 | | | | | | | |
| Beijing National Stadium | G. | i. | CPT | 1 | ✓ | OBJ | http://en.wikipedia.org/... | http://rdf.freebase.com/ns/e... | 116.3906 | 39.9917 | |
| The birds nest | G. | i. | CPT | 1 | | | | | | | |
| Prefecture of Rhodes | G. | i. | CPT | 1 | | | | | | | |
| Great Wall of China Mutianyu South | G. | i. | CPT | 1 | | | | | | | |
| Mykonos Windmills | G. | i. | CPT | 1 | | | | | | | |
| Mury Miejskie czesc II | G. | i. | CPT | 1 | | | | | | | |
| Great Wall of China Mutianyu North | G. | i. | CPT | 1 | | | | | | | |
| Blok przy ul. Basztowej 11 | G. | i. | CPT | 1 | | | | | | | |
| Beynac castle | G. | i. | CPT | 1 | | | | | | | |
| 1010 Washington St | G. | i. | CPT | 1 | | | | | | | |
| European Parliament Office in Spain | G. | i. | CPT | 1 | | | | | | | |
| Statue Place Bellecour Lyon | G. | i. | CPT | 1 | | | | | | | |
| Bergedorfer Schloss Hamburg | G. | i. | CPT | 1 | | | | | | | |
| Award-Winning Work | G. | i. | CPT | 0 | | CPT | | http://rdf.freebase.com/ns/a... | | | |
| Olympic venue | G. | i. | CPT | 0 | | CPT | | http://rdf.freebase.com/ns/o... | | | |
| Ai Weiwei | G. | i. | PER | 0 | | PER | http://en.wikipedia.org/... | http://rdf.freebase.com/ns/e... | | | |
| Jacques Herzog | G. | i. | PER | 0 | | PER | | http://rdf.freebase.com/ns/e... | | | |
| Li Xinggang | G. | i. | PER | 0 | | PER | | http://rdf.freebase.com/ns/e... | | | |
| Pierre de Meuron | G. | i. | PER | 0 | | PER | | http://rdf.freebase.com/ns/e... | | | |
| Beijing | G. | i. | LOC | 0 | | LOC | http://en.wikipedia.org/... | http://rdf.freebase.com/ns/e... | 116.4 | 39.91 | |
| China | G. | i. | LOC | 0 | | LOC | http://en.wikipedia.org/... | http://rdf.freebase.com/ns/e... | 110.46... | 32.90429 | |

**Figure 22: Enrichment of named entities**

### *4.4.3 Scenario*

Task 1: Selection of corpus about Brassens (177 records).

Task 2: Get the terms

Task3: Manual enrichment

Task 4: propagate the enrichment

### 4.4.4 Results

Details results, screen shots and answers given to predefined set of questions are given in “162”.

### 4.4.5 Usability analysis and conclusions

The MAT interface is very well suited to achieve this scenario. The interface is still based on the same features than the ones evaluated previously in the other MAT scenarios: web knowledge panel, record annotation panel, web documentation panel, selected corpus panel (similar than for propagation). A new panel provides the display of terms extracted from the annotations corpus. The visual design and the principles of interaction always keep a consistency between all panels.

The effectiveness is good because this interface let the user perform the scenario without any error. The interface brings together all useful information to manually add annotation in the most efficient way. However, these tests are not sufficient to indicate that there are no missing features. This service is dedicated to professional users to achieve very specific tasks. A learning step is needed to use this service which is very new for annotators used to classical system. There will be no difficulty to understand the functionality with documentation. Also, documentation is necessary to discover some interactions. Drag-and-drop is very easy to do but all the functionalities are not easy to find for the first time. Several functionalities are hidden, for example drag-and-drop a term in the selected corpus panel to display all the documents having this term in their metadata. It is also not intuitive to know that you have to drag-and-drop the label from the working box to the upper box in order to enter it in the table of terms. When you have discovered this action, it is then very efficient. Visual design is well adapted to the scenario but could be improved again (see recommendations). Users have expressed some satisfaction with this scenario. They appreciate especially the efficiency of the management of the formatted data in the web knowledge. They also appreciate the general efficiency of interaction with drag-and-drop and short-cut (as the box on the top of the working box).

So the conclusions of the cognitive walkthrough are positive, no conception or major ergonomic problem has been found. Effectiveness, efficiency and satisfaction are very good according to this context of use given by the scenario. Additional user tests need to be done to give deeper conclusions about easiness of use and about effectiveness provided with a longer test scenario. Moreover, tests have to be done with distant database and the propagation time has to be improved.

This evaluation showed the advantage of using this tool to perform/propagate this type of manual enrichment, however this assessment does not include the interest to make such kind manual enrichment as it depends on the way ASSETS / Europeana will draw benefits of these enrichment for search and inter-linking process (server and portal side). It was explained by the developping team this prototype provides a proof of concepts in the following sense:

- The storage format adopted is still indicative, and should be validated.
- At now many actions which should be performed on the server are taken in charge at the client level. This client / server dialog may be seen in various way to optimize considerably the performances and release the client once propagation has started. Technical choice remains to be done at this level.
- The operated enrichment remains a proposition, it may be modifed.
- Portal side modification should be handled to show the enrichment done and allow assessment of the interst /quality of query results in the web portal.

Lack of time prevents complete the technical and user validation cycle and results exploitation on server and portal side, for this scenario which were not forecast at the start of the project.

So this evaluation aimed only at presenting functions and a (very slow) way to perform them. If the idea is considered worth to be carried on, additional technical validation should be done after additional technical development on various side (client, server, and eventually portal) and new user validation should be done to check resulting performances. But this new validation should not give any bad surprises if the present prototype is considered functionally valuable.

### *4.4.6 Recommendations*

High level:

- Improve propagation time
- Connection with distant database

Low level:

- Drag-and-drop possibility from working box to the table of terms
- Suppress the left place in the working box (independency of panels)
- Improve the display of thumbnails when there are long titles
- Put a visual marker to indicate when the system works

## 4.5 Content creation by re-use (UGC tool)

### *4.5.1 Introduction*

In this section, we present the evaluation of the Content Creation by re-use service through the interaction with its functionalities through the user interface. This evaluation is the first evaluation involving the user interface and functionalities besides the more technical evaluations that have been performed by the developing team of this service.

This evaluation validates the quality of use of the content creation by re-use service and point out important issues for improvements.

***Evaluation elaborated by***: P. Hansen (SICS)

***Technical contact for the service***: N. Aloia (CNR)

***11 Remote users test performed***: eleven users, SICS (3), FLM (3), DW (1), HASC (3), Ina (1)

***Date***: March 2$^{sd}$ – March27$^{th}$ 2012

**Objectives**

The objective for the evaluation of the Content creation by re-use service is:

1. To analyze the overall usability of the functionalities of the content creation by re-use service tool
2. To test the ability to perform a set of interconnected tasks and the success of completing three tasks of different complexity levels.
3. To test the usability in terms of learnability (e.g. learning the functions during the different task complexity levels, efficiency (e.g. the ability to use and complete the tasks) and satisfaction (how satisfied the users were as regard to their expectations.

The scenario is based on the technical use case "Mona Lisa Seminar Scenario" in which a user of ASSETS decides to create a UGC object describing a seminar on Mona Lisa and to submit the newly created object to the EUROPEANA digital library. The overall goal is to verify that professional users may be able to perform expected tasks using the Content creation by re-use tool with easiness, quality and satisfaction, such as:

- To upload objects to the content creation tool
- To assign metadata to the objects,
- To search Europeana, and
- To link different objects to each other and enhance them with metadata

Limitation of this evaluation:

- The evaluation is not focused on pure esthetics of this tool since it will be a tool not manipulated by end-users.
- The evaluation is not involved in technical aspects of this tool/service, and we would like to point to the technical validation and technical quality assessment done by the developing team of this tool.
- The evaluation only involved a limited set of objects and metadata connected to each

other. The current evaluation only dealt with a pre-defined training set.

**Methods**

For this service, the main method used for collecting data is

- Remote user testing.

The remote user testing involves that the users have been equipped with a pre-defined evaluation package to be used. The assigned user then unpacks the evaluation package and performs the test on site at their own offices.

This test, as well as the other tests, are planned and executed with a similar set of evaluation tools and protocol (since the tools evaluated can be very different, there may be differences in the evaluation set-up, protocol and measures used.

The evaluation package contains:

- The presentation of the tool:
  - Description of the evaluation procedure
  - Presentation of the user interface,
  - Presentation of the functionalities, and
  - Description of the evaluation procedure
  - The user were also provided with an image to be used (upload task)
  - A set of 5 different questionnaires to be filled out
    - Pre-questionnaire
    - Task 1 questionnaire
    - Task 2 questionnaire
    - Task 3 questionnaire
    - Post-questionnaire

**Users**

In order to perform the test we needed to involve professional users from within the consortium. For this specific service, we used 13 professional users from CVCE, HASC, FLM, SICS, Ina, DW.

### *4.5.2 Presentation of user interface of Content creation by re-use*

A complete set of images can be found in “Appendix 6: Content by re-use”.

The interfaces developed for the Content creation by re-use have basically the following panes:

- Accessing the Assets application
- Upload a file from own desktop or from an external source.
- Search Europeana
- Add a property (feature) to a Europeana Object
- Create a compound object by linking 2 or more objects

The interface below shows the general interface for the work-area. To the right there is an open space for working with an object (adding metadata and linking objects). The left there are four tabs that allow the user to upload an object, search the Europeana repository, handle own objects and

add metadata properties.

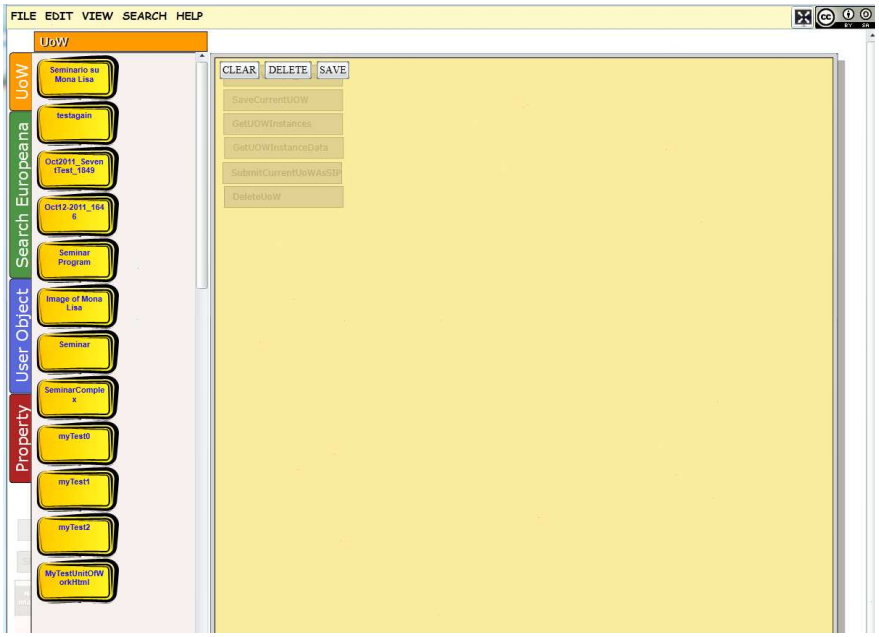



**Figure 23: Work-area for the content creation by re-use tool.**

### *4.5.3 Scenario description*

The objectives of this service are aimed at facilitating the information exchange between the users and the digital library: users find in ASSETS and Europeana the sources of information needed for their work and, in turn, they enrich the library with the results of their own work. In particular, the services allow ASSETS users to contribute to the contents of the digital library in several different ways, such as uploading simple media objects along with their descriptions, annotating existing objects, enriching existing descriptions or creating new complex content by extracting and recombining in various forms existing or new digital objects or parts thereof. Rather than focusing on a specific set of UGC applications, ASSETS has developed a general purpose, back end component that aims at supporting any UGC service Europeana will want to offer to its users. To this end, the ASSETS back end component implements an Application Programming Interface (API) for creating, storing and manipulating UGC instances, and for submitting these to Europeana, in the form of Europeana Submission Information Packages (SIPs).

In order to show this, we designed three different levels of scenarios (tasks) with an increasing complexity level. The objectives for the services is described in the following scenario which has been transformed into 3 different tasks:

- *Low complexity*:
  - ASSETS server initiation
  - Uploading of a media file from the user's workstation
  - Defining and assigning some metadata values.
- *Medium Complexity*:
  - Search Europeana for an object using a query input to the search box
  - Adding some metadata values to the object found
- *High Complexity*: Creating a complex object,
  - Combining an Europeana object and a user uploaded object using UGC functionalities

- Providing appropriate metadata values to that new set of objects

### *4.5.4 Report of the results*

The evaluation concerned 3different but connected tasks using the Content creation by re-use service. These three tasks contained partially three different parts of various complexity as well as parts common to all the three tasks.

Each user had to finalize 5 questionnaires. The pre-questionnaire contained questions about demographics and the experience of searching and using metadata. The post –questionnaire contains question about the overall experience and satisfaction of using the UGC tool.

Questionnaire 1-3 each corresponds to the three tasks of different complexity. Each of these tasks contains a unique part and a 2 common parts across the three tasks. The unique part is as follows:

- Questionnaire1: uploading an object
- Questionnaire 2: searching Europeana
- Questionnaire 3: linking objects

The two common parts across all three tasks are:

- Adding metadata value to an object
- General comments

*Pre-questionnaire: The users and their experiences:*

All 13 users came from different countries: France, Germany, Greece, Italy, Luxembourg, and Sweden. 9 male and 4 female users. The background for these 13 professional users was as follows: 9 content providers, 3 developers, 1 other (not specified). 92% (12 users) were connection to the Assets project.

Search experience: 100% used Google; 38% used Wikipedia; 8% used regularly Europeana; 27% used other search engines in order to find information.

Metadata experience:

a) 18% (2) did not have any experience of adding metadata to an object and 82 % had some or regular experience of adding information to an existing object.
b) Regarding experience of using metadata, 46% had less than a year of experience, while 23% had more than 10 years of experience.

Those participants that used Europeana, answered that they used it for:

- Developing professional services for it. (3)
- To check company items (2)
- To look at overall content from other content providers' (2)
- For collect information about my collections (2)

In the following section, we will describe the task specific functions separately and then follow up with the tasks common across all tasks – adding metadata.

***Questionnaire 1-3: The three UGC tasks***

The three questionnaire (1-3) asked task specific questions on:

- Task 1: uploading an object

- Task 2: searching Europeana
- Task 3: linking objects

*Task 1: uploading an object*

13 users performed the task and the user had to upload an image from their own desktop of their choice or use the image provided by the evaluation package. In general the users were satisfied with the functionality and said that the principles behind the upload function was clear and easy to understand. They were also glad that the tool did not time out during the performance.

Regarding the implementation one user said that it was reliable, while 5 users pointed out problematic issues related to platforms: a) difficulties: Notebook with Win XP and not Chrome compliant, b) no problems with Mac OsX 10.7.3 with Safari 5.1.3, Workstation (Win 7). Another issue that was raised was that 3 users were alerted that there was low memory through a pop-up message.

Closely related to the upload functionality, the following aspects were pointed out by the users

- Semantics: The users (2) pointed out that the meaning of UOW tab was not clear and what it was about.
- Repetition: a user experienced problem when uploading an object. The reason was that the URL was then demanded and the user didn't remember it.

As can be seen in the figure (X) below, the users did not have any problems with electing an object, uploading the object and naming the object. Some locating the object presented some difficulties, but overall it went without problems and all users managed to perform this task.

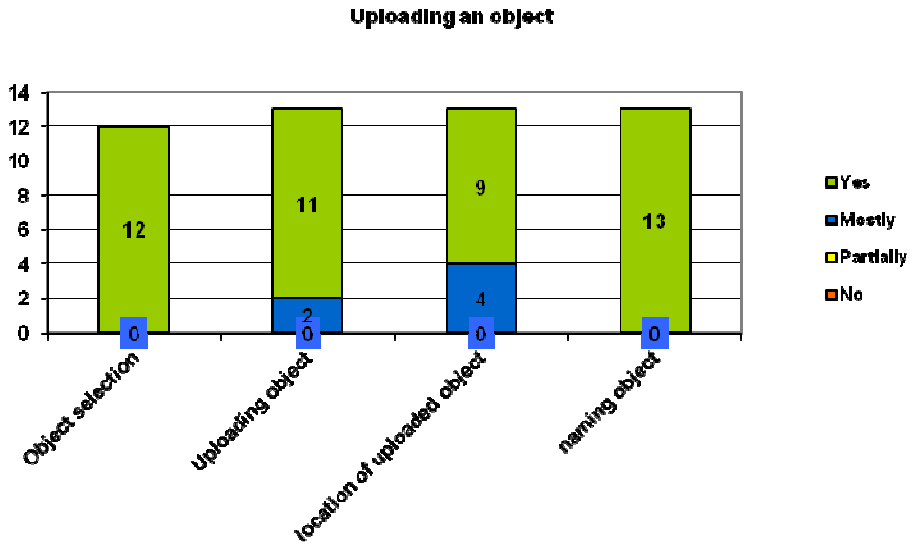



**Figure 24: Uploading an object**

*Task 2: Searching Europeana*

The second task specific functionality was to search the Europeana repository in order to upload an external image. In general, the users managed this task properly as well. The following issues was raised during the performance:

- The drop-and drop function was sometimes slow and difficult to handle.

- 2 users pointed out that they got messages about short 'memory'

According to the table below, the users successfully used the search functionality for searching Europeana repository. Furthermore, 11 out 13 users understood the drag and drop functionality even though 2 users had some difficulties. Most problematic aspect of this task was the presentation of the objects in the result list. At least 4 had more or less problems with understanding the presentation.

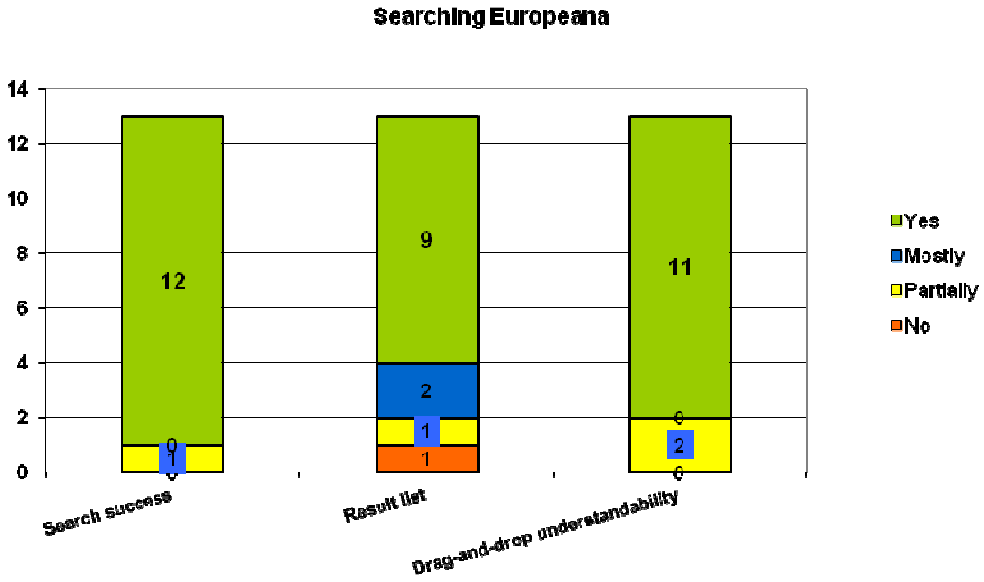



**Figure 25: Searching Europeana**

*Task 3: linking objects*

The third task concerned is with linking two objects. This is done by using the uploaded object from the user's desktop and linking it with the object uploaded from the Europeana search. The table below shows that the user's had no problems navigating between the different tabs on the left side and no problems of using the drag-and-drop function. Some problems occurred regarding the amount (number of times) of action used in order to be able to create the link between 2 objects. Although everyone did manage to do that, I user was not satisfied.

Issues mentioned in the questionnaires related to linking objects were:

- A user experienced that adding links properties to an object didn't always work.
- Another user thought that it was cumbersome to change relationships. Instead of simply replacing the object of the relationship, you need to remove the relationship property, then add it again and then assign the new object as part of the relationship.
- The Drag-drop functionality were experience both positively but could also be perceived as problematic: Some users had to repeat several times the drag and drop over the blue "+" in order to link the two objects.

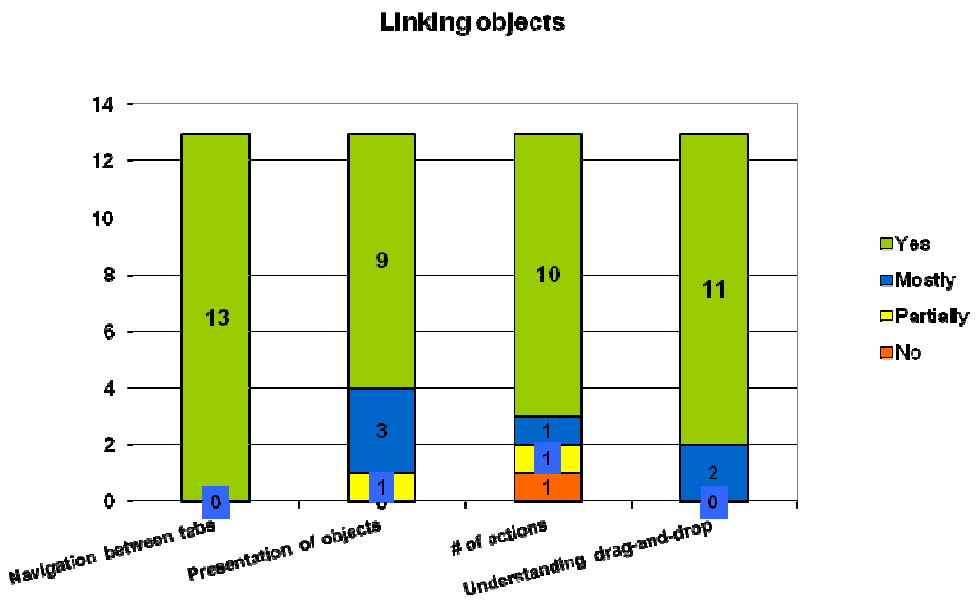



**Figure 26: Linking object**

Since this task was the more complex tasks of all three performed by the users, there were additional comments made about the overall functionality.

- 3 users pointed out that it would have good to see the published result at the end. When the last step of saving the work as a UOW with an ID, was done, the object did not appear in the list of UOW's in the left-hand pane. Furthermore, the users also wanted to have a message that the work was actually saved.
- Correction issues. Some raised the problem of making correction in a smoother way. For example: adding properties to the Europeana object two times, it was problematic to add new properties. The user then needed to cancel the first property added and re-start the action.

*The metadata assignment task across three tasks*

Next, we will present the result of the one common task across the three separate tasks complexity levels: Adding metadata properties to an object. Adding metadata to an object is part of all three tasks in different complexity degrees. Adding metadata to uploaded object from your own desktop; adding metadata to an object searched and retrieved from Europeana and finally, adding metadata to a combined set of objects. One thing we can monitor is the learning curve in two ways: first since the activity of adding metadata is done repeatable three tomes and also at an increasing complexity level.

First we will present the comments made performing the metadata assignment tasks in task 1,2 and 3.

During the metadata assignment tasks the users mentioned that

- The question marks are helpful for the properties, but the information disappears too fast from the screen.
- Previous knowledge/ Experience of Europeana:
    - Another user pointed out that one need to be familiar with the Europeana terminology to be able to select properties in an efficient way.

  - You need to know Europeana constraints and definition to edit with confidence the various properties.
- Knowledge level:
  - It requires some familiarity with the properties, e.g. deleting an added property, getting details on a property, referring to a related property, being aware of differences between closely/ seemingly related or overlapping properties.
- Lack of information:
  - A user mentioned the problem using the retrieved Europeana object. User needs more detailed information about the object.
  - 2 users mentioned that the magnification glass button does not work. This makes it impossible to get any information on the information object. It makes metadata assignment difficult. Is this an efficient way to add metadata to information objects that have already been added to Europeana since no context is available for figuring out exactly what object you are manipulating.
- One user mentioned that not all metadata assignment resulted in a blue '+' sign.

*Learnability*

In the table below we can see that there is some positive learning curves related to the three sub-tasks monitored.

From task1 to task3 there was a positive learning curve regarding the issue of the easiness to expand an object with metadata even though the complexity increased (6-10-12 users found it easy to perform this task).

Regarding the usage of the '+' sign, one can see an increased understanding of how to use this function between task 1 and 2. Again, at task 3, the number of users feeling comfortable with the function decreased to 10 users. This may be due to the complexity level of the task in task 3 when the users needed to add metadata to a combined object.

Regarding the final observation, understanding what 'property' meant, we see an increasing learning curve from task 1 to 3 (8,12 resp. 12).

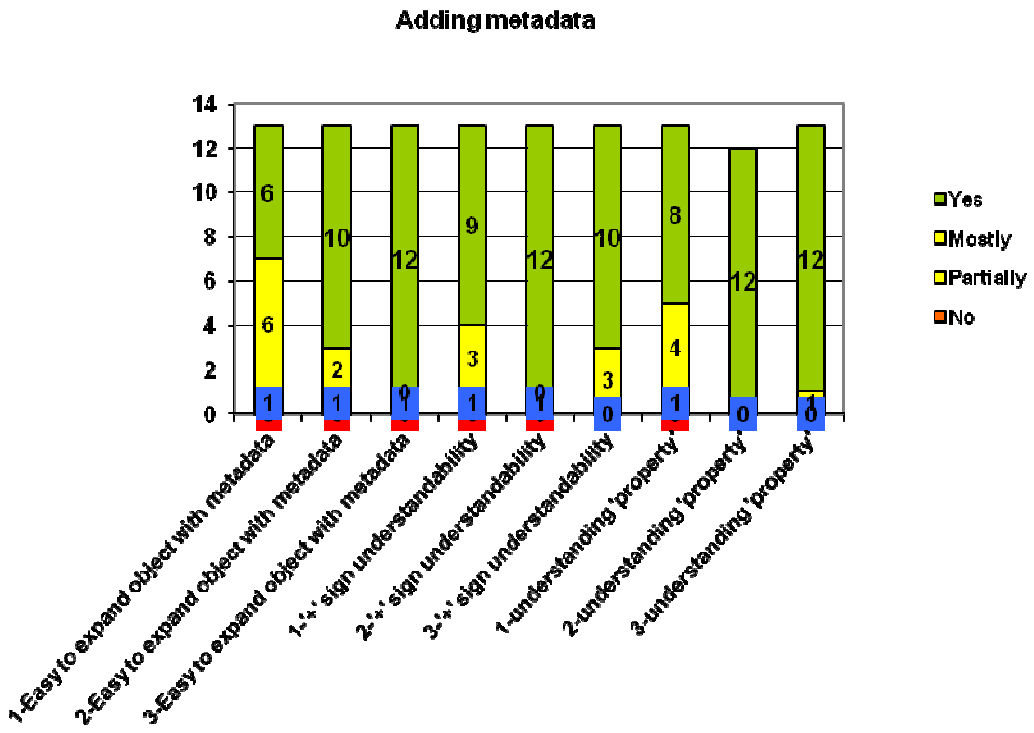



**Figure 27: Adding metadata**

This notion of learning was also commented by the users during task performance.

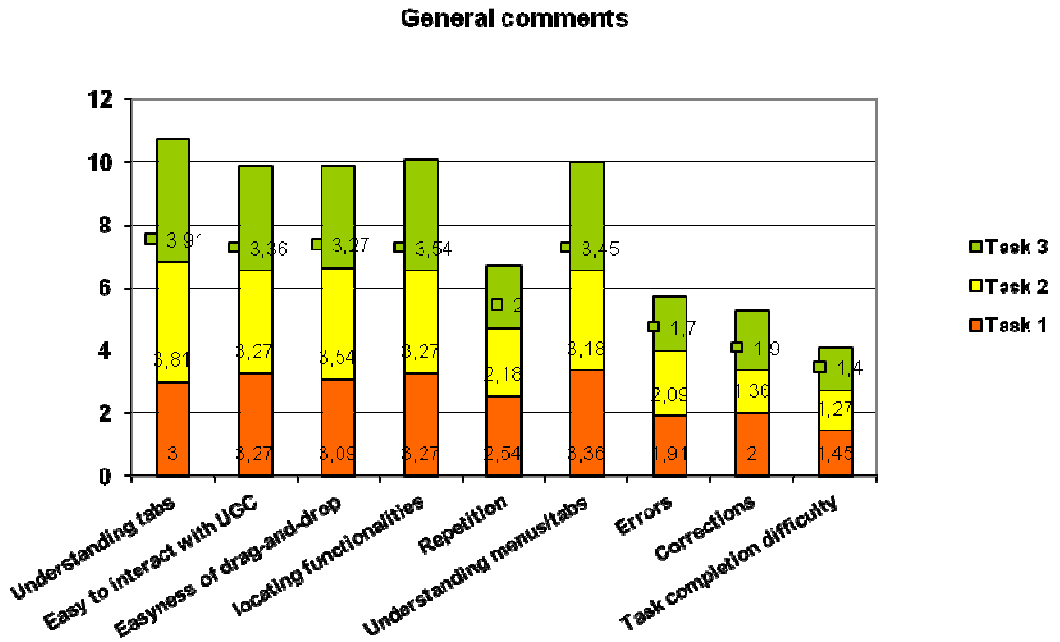



**Figure 28: General comments**

### *4.5.5 Analysis of the usability*

**Efficiency**

Efficiency relates the level of effectiveness achieved based on the available resources. Here expenditure of resources is analyzed from mental effort such as learnability and understandability. Efficiency as regard to time was not an issue here because remote tests are not conducted with the same equipment and locality. For this test, efficiency is measured by 1) adding metadata; 2) understandability; 3) number and type of errors; 4) the perceived difficulties

Even though time effort could not be calculated, there were users pointing out that the drag-and-drop functionality had a long response time. Connections or other more technical issues were not considered as problematic.

The user usually made 1 attempt for each metadata task that gives us 39 attempts to add metadata to an object. In 8% (3 out of 39) the user needed to redo the metadata addition. In close relation to this, the users that re-added metadata only did that once in during the first task. Thus, the learnability curve was high and the efficiency of adding metadata was enhanced. Some users also made this explicit. The level of understandability also showed a high score (3,92 out of 4,00) at the end questionnaire (figure X). Regarding number and type of errors, they usually belonged to the third task and to the task of linking objects and adding metadata properties to the combined object. Thus could be due to the level of complexity in the third task.

The users did not describe major difficulties with the UGC tool, rather expressed that, with few exceptions, that the tool was easy to use and easy to understand. The cognitive problems had its reason in less knowledge on metadata issues.

**Effectiveness**

Measures of effectiveness relate the accuracy and completeness with which the users can achieve their goals as well the compliance with the user need. Effectiveness relies also on technical criteria related to the system performance as reliability of the service.

In this evaluation, effectiveness is measured by how many task out of the 5 original main tasks that have been correctly completed. The main tasks were uploading an object from own desktop, add metadata to that object, search Europeana repository for an object, link 2 objects and provide metadata to the new object. 100% of the professional users completed all their tasks successfully without leaving any task unfinished. Thus, the user's goals were achieved and in this sense the tool was effective.

Although all 13 users achieved their goals successfully, 2 users pointed out that they were unsure if the work had been saved at the end, and that can be calculated as 85% in terms of accuracy.

All functionalities provided by the service were in line with what the users expected, in terms of usefulness and most users used the functionalities in a very similar way across the different task complexity levels. One user wanted to add more metadata properties to the objects and another user wanted to search and upload more than 1 Europeana object. The users with high metadata knowledge, the functionalities seemed rather straightforward.

Another measure for the criteria of effectiveness is if there was any function missing. The result shows that no major function was missing in order to complete the task. However, the users pointed out that it would be good to have a function for deleting or correcting a mistake. They also asked for textual guidelines for certain task such as for the metadata taxonomy and the metadata construction of a Europeana object.

As regard to the maturity of the service, there were some problems with using the UGC tool on some computer platforms. Some tests and documentation on these tests might be useful for the users. Also, when repeating some actions did not work and the user was not sure what was happening and in order to get on track again, the restarted the task.

**Satisfaction**

Satisfaction describes how comfortable the user is with the use of the tool. Usually, the measures is characterized the user satisfaction to execute or perform the task. The data for these measures are extracted from the post-questionnaire and measures the level of usefulness, effectiveness and easy to use at the end of the evaluation.

The satisfaction criteria is divided into several sections that concern the overall functionality, uploading and naming an object, adding metadata, searching Europeana object and linking 2 object

Regarding the satisfaction with the overall functionalities (Figure 29), users were with the overall usefulness, easy to use and effectiveness of the UGC tool. Even higher score of satisfaction did the number of functions designed for the tool get with an average of 3,8 of a 4,0 scale.

The work-area was also found useful and effective. Some lower score on easy to use and this may be due to that some user pointed out that it was problematic to manipulate the objects within the work-are.

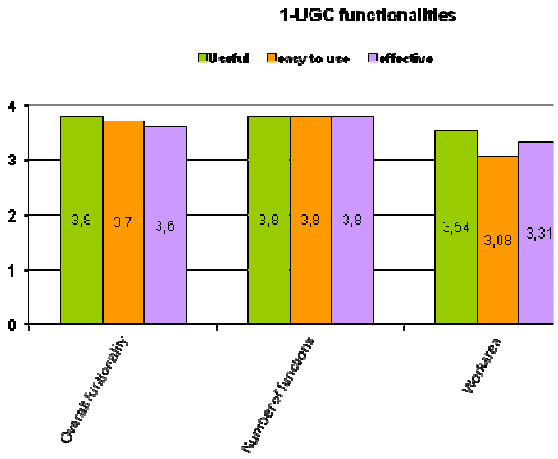



**Figure 29: UGC overall functionalities**

Regarding the actions of uploading, naming the object and moving the object (Figure 30), they were all considered very useful (3,85, 3,77 and 3,92 respectively) by the users and without any major problem. However, uploading and moving around the object were considered somewhat problematic and uneasy. The uploading issue may have its reasons in the problem with the URL earlier mentioned.

Adding metadata showed some different scores. Adding metadata in general was considered very useful, as did the display of metadata properties. However, the metadata properties display and adding metadata scored comparatively low regarding the effectiveness. Reasons for this is that there were users that were not skilled in metadata assignment tasks and had difficulties of knowing what they meant. Display of metadata, metadata taxonomy and adding metadata all scored around 3.

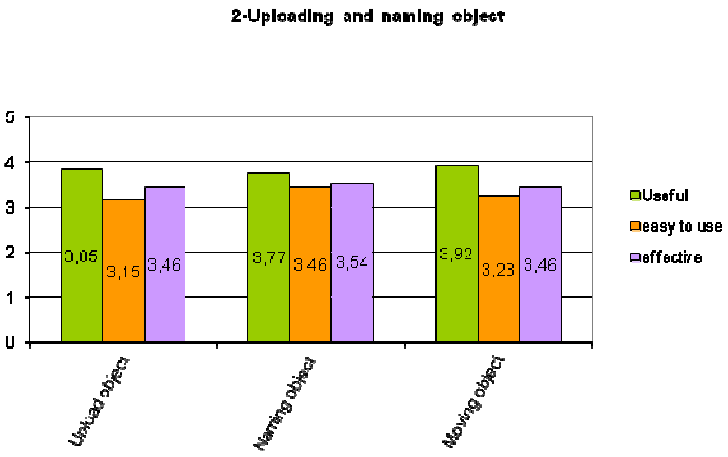



**Figure 30: Upload and naming object**

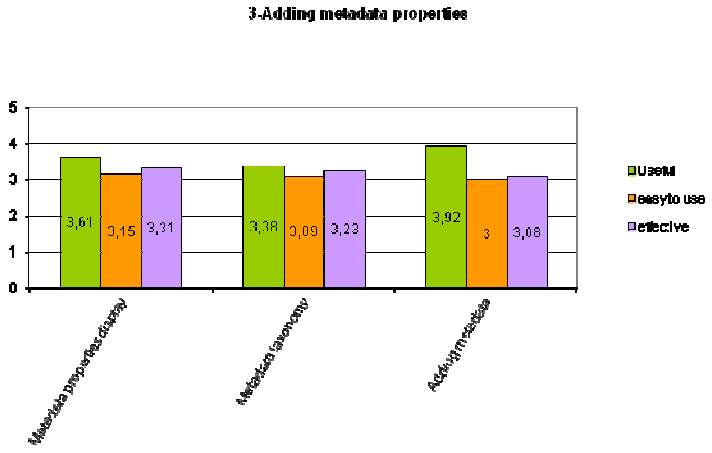



**Figure 31: Adding metadata properties**

Searching Europeana was considered useful, effective and easy to perform. The task of describing an object was considered very useful, easy to use as well as effective. However, the function of result presentation showed low scores (3,08, 3,00 and 3,08). The users pointed out reasons such as that the thumbnails should be expandable and that the objects contained very little information.

The users found the linking function very useful, but not effective. One comment made by 2 users was the issue of scalability. One object could be linked to many objects and it would then be cumbersome to do all the linking. The functionality of creating a new object by linking several objects was considered both useful, effective as well as easy to use by the users. Finally, the users perceived that it was easy to understand the overall linking procedure (3,38).

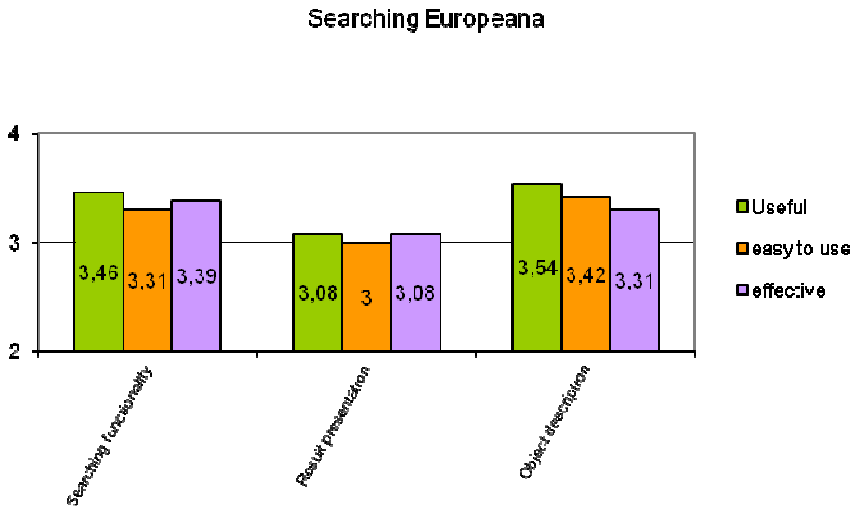



**Figure 32: Searching Europeana**

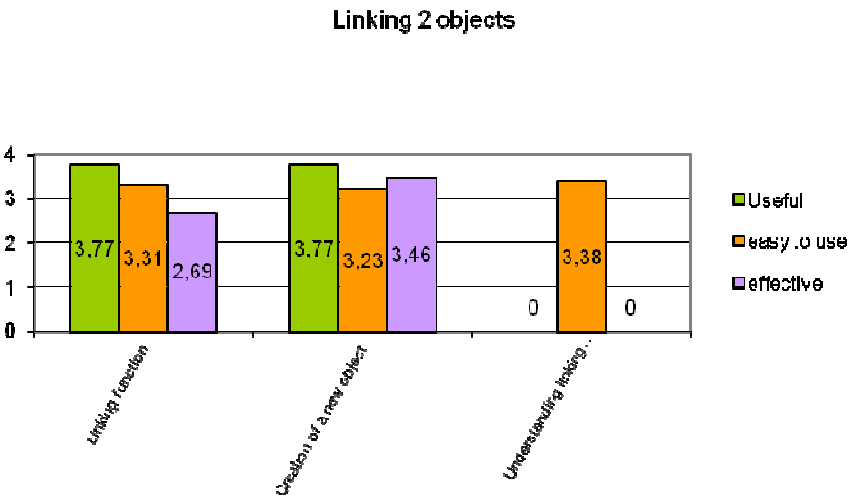



**Figure 33: Linking 2 or more objects**

Finally, we measure the satisfaction regarding overall performance. The users agreed that it was satisfactory to use the UGC tool and that it was easy to understand the UGC tool as well as understanding the purpose of the service. Less agreeable by the users were the visual presentation of the interface, the response time and the cognitive clarity of what was going to be done during task performance.

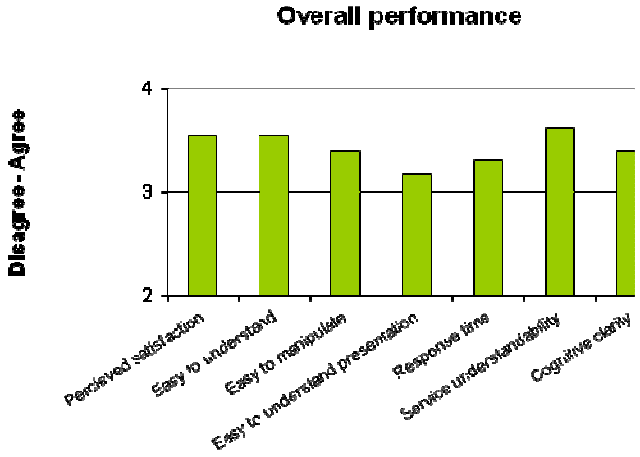



**Figure 34: Overall performance**

### *4.5.6 Recommendations*

There are three levels of recommendations back to the developers of the UGC tool: High, medium and low.

High level

- To provide a delete or undo functionality. It should be easier to make corrections for the user.
- To provide functionality that remembers the URL when using the reloading functionality. This could avoid cumbersome repetition of reloading tasks.
- Scalability. In order to scale the linking functionality, the tool probably needs to be tuned for a smoother linking functionality. Example: if a user needs to link 8 different objects together in order to create a 'family' of connected object, this can be a cumbersome task.
- To provide a message to the user when the linking of objects was finished and the creation of an object ended, in order to know if the work saved

Medium level

- To provide a better explanation and guidelines on what the results list shows and a better explanation of the objects in the list.
- To make making the task of changing the link between 2 objects easier and more user friendly.
- To provide guidelines and instructions on metadata procedures and how Europeana objects are designed for non-experienced users. If there is a situation in which inexperienced or less experienced person works with this tool, a there might be good idea to be able to have guidelines or a general descriptions on metadata properties in general and more specifically how Europeana objects have been tagged. This would also solve the problem with uncertainty as regard to knowing what object you really are manipulating (the retrieved Europeana object or a copy of that object retrieved and placed at your work-area).

Low level

- To explain to the user why the pop-up message turns up telling the user that there is low memory
- To provide a list of known problems with certain computer environments.

## 4.6 Access to query log analysis

### *4.6.1 Introduction*

This chapter presents the evaluation of the professional service of access to query logs. The Graphical User Interface has been evaluated with a cognitive walkthrough protocol because it was ready to be tested only at the end of the project. It has been tested with log requests belonging to Europeana data from august 1st to February 31st. The system was tested with a local database.

***Evaluation elaborated by***: A.Saulnier (INA),

***Technical contact for the service***: GUI:Tsuyoshi Sugibuchi (LRI), logs: Claudio Lucchese and Diego Ceccarelli (CNR)

***1 cognitive walkthrough*** with four Ina users, one evaluator and the GUI provider.

***Date***: March 27th 2012.

**Objectives**

The objectives are to evaluate the usability of the GUI providing access to query logs.
Will the user be able to analyze a corpus of log request with this interface?
Will the user have useful functionalities to achieve his task?
Will the user be able to use them?

Limit of this evaluation: we are not evaluating the quality of the log requests.

**Measures and criteria**

This adapted cognitive walkthrough (realized with professional user) let obtain qualitative measure for each criteria of the usability:

- Effectiveness: verification that the predefined scenario could be realized
- Efficiency: comparison made by the users with their previous experience
- Satisfaction: feedback given by the user

**Cognitive walkthrough method**

Cognitive walkthrough can provide feedback to provide some feedbacks about effectiveness and efficiency. Since users do not handle the system by themselves, satisfaction can only be inferred from perceptions that users have had of the system.

The service provider demonstrates his interface in front of the users and evaluator. This method is based on discussions and comments while achieving a scenario of use according to this set of questions:

1. Does the user know what to do to achieve the task?
2. Does the user understand the proposed functionalities?
3. Are the proposed functionalities easy to find, easy to combine?
4. Does the user understand input and output of the system?
   Does the result match the user expectation?
   Is there a positive guide and feedback for the user? Will he know how to continue?

5. Is the visual design of the interface relevant?
   Are the different elements of information easily readable and distinguishable?

A questionnaire has also been filled by the users.

**User profils** are three expert users (developers) of GUI, one user involved in field studies.

### 4.6.2 *Interface presentation*

Interactive GUI for exploratory log analysis.

The user chooses parameters to define a visualization that he can then filter or modify.

The interface is divided in two parts:

- Cube navigator: displays “dimensions” which can be selected and manages filters
- Chart table: displays column x rows x cells (with char template)

Three main operations are available according to OLAP operations done in the database:

- Pivoting: change selection of dimensions to display
- Slicing & dicing: restrict values of a dimension
- Drill down/up: change level of dimension

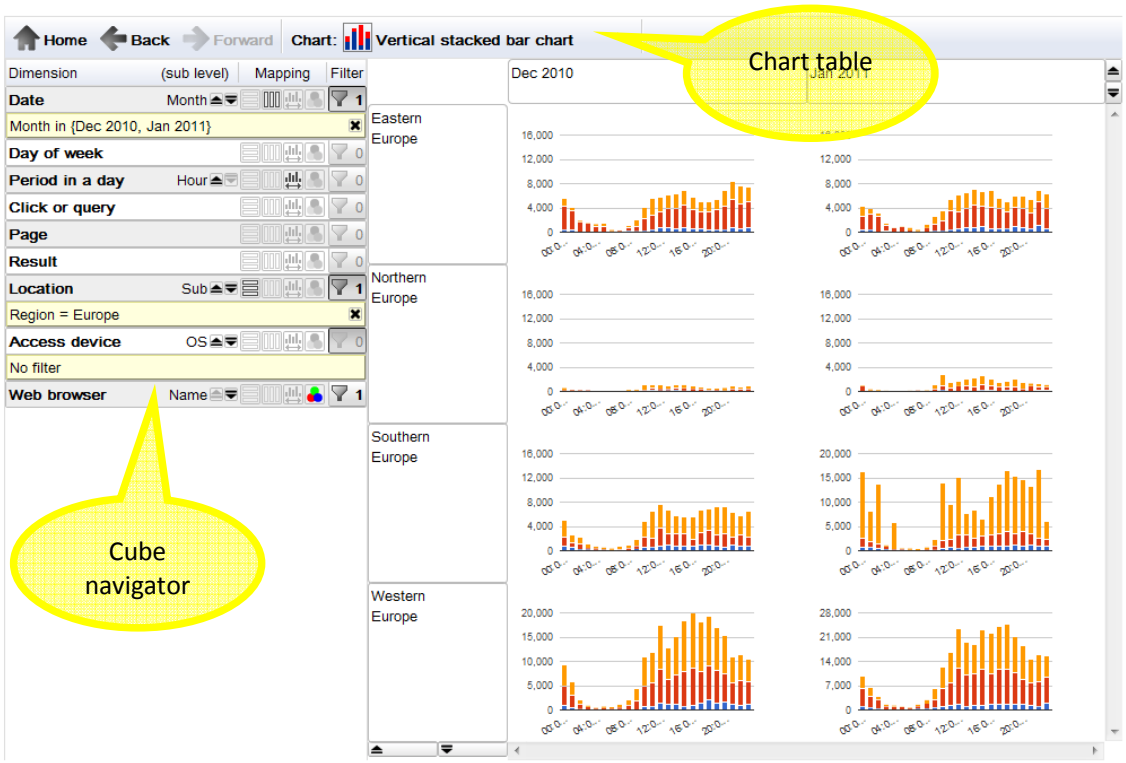
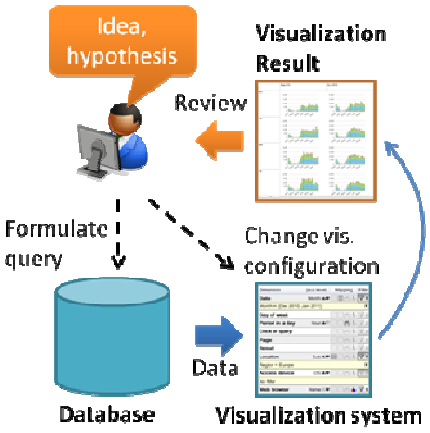



**Figure 35: GUI of access to log requests**

### 4.6.3 Scenario

"Analysis of the log requests in order to understand the popularity of the web browser in Europe."

The scenario relies on the exploratory visualization of a database log request extracted from Europena (from august 1st to February 31st)).




The user has a hypothesis about the logs content. He wants to analyze the logs database and confirm his analysis while manipulating. He could also affine his idea or find another interesting results relative to his topic of search.

**Figure 36: Exploratory visualization**

The proposed scenario contained four representative tasks to be performed with the system:

- Task 1: Define an initial visualization Quarter / Region
- Task 2: Focus on sub-region in Europe
- Task 3: Observe the number of web access during 4 periods (of 6 hours) in a day.
- Task 4: analysis of the different web browsers.

### 4.6.4 Results

Complete results (scenario, screen shots and answers to questions of the cognitive walkthrough and questionnaires) are reported in the appendix "Appendix 7: Access to Query log analysis".

All answers to the predefined questions were positive according to the proposed scenario.

Questionnaires, about users 'perceptions of the system, were also very positive.

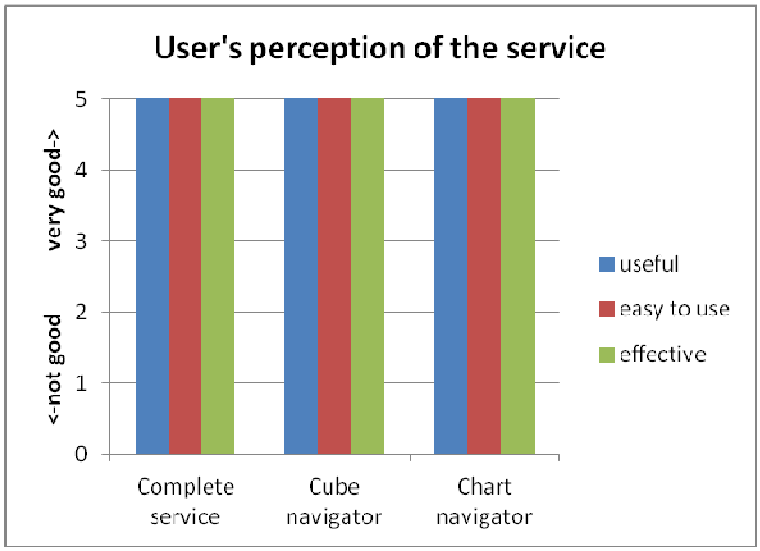



**Figure 37: User's perception of access of log requests**

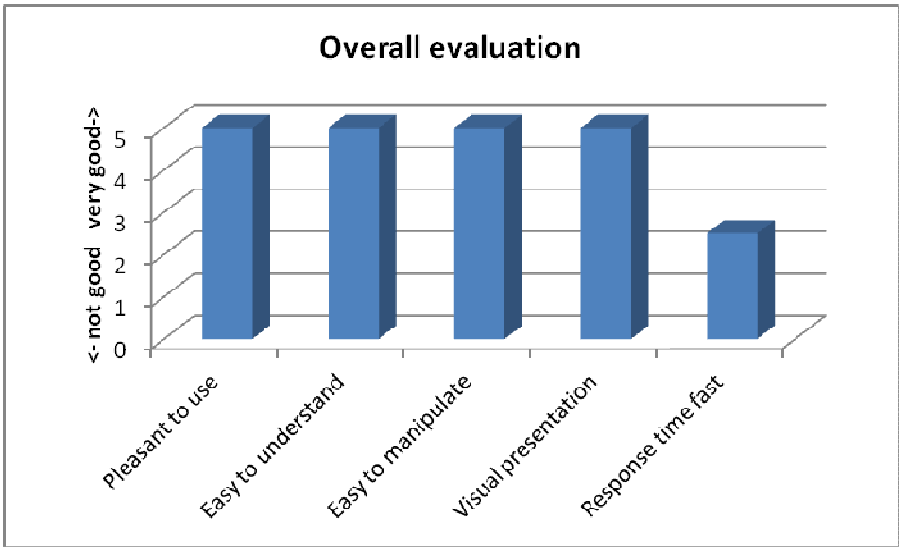



**Figure 38: Overall evaluation of access of log requests**

### *4.6.5 Usability analysis and conclusions*

The feedbacks were very positive. The conclusion of the walkthrough is that all the functionalities were present to realize the proposed scenario with effectiveness and efficiency. All the useful functionalities were present and let the user configures visualization in order to be able to make pertinent comparisons between data. All the functionalities are very intuitive and seem easy to manipulate. The user can understand very rapidly the interactions between cube navigator and chart table. Moreover, users accustomed to chart, this make the system easy to learn. The visual design is also relevant. All the functionalities are easy to identify and are very well put together in a same menu. Presentation is very well structured. In cube navigators, menus are visually separated into blocks on the horizontal and vertical axes. In chart table, visual characteristics are very well adapted to map the data. User satisfaction was good according to the scenario realization.

So the conclusions of the cognitive walkthrough are very positive, no conception or ergonomic problem has been found. Effectiveness, efficiency and satisfaction are very good according to this context of use given by the scenario. However, the tests are not sufficient to indicate that there are no missing features. Indeed, we would have to treat a much wider scenario (which more data) for this analysis. Similarly, feedbacks were collected just from observations and not from interactions done by users, so they cannot include “real” difficulties.

When envisaging a more complete analysis scenario, an additional evaluation should be done to analyze one problematic issue: how the user can choose the best visualization according to data characteristics? What degree of complexity analysis can be conducted with this service? Then learnability step should probably become necessary to use the interface with the best performance. Other tests have also to be done to evaluate the use of the system with a greater number of data. The time performance of the system is also a crucial point to guaranty an efficient use. Lastly, technical tests have to be done to guaranty the use with a distant database.

### *4.6.6 Recommendations*

High level:

- Improve speed of response time
- Connection with distant database

Low level

- Put a visual marker to indicate when the system works
- Indicate more clearly the selection done in the cube navigator

## 4.7 Ingestion: Workflow management

This service provides automatic enrichment of metadata records capabilities to the ASSETS platform.

### *4.7.1 Introduction*

This chapter presents the evaluation of the usability of Ingestion service through GUI. User test are perform in operating phase: this evaluation is conducted in real conditions of use during the enrichment process conducted by Content Providers on their own data.

This evaluation is focussed on the usability of the system offered to the professional users who will have to enrich their metadata. It deals with the analysis of the service offered to the professional user to achieve his task but will not deals with the quality of the enrichment.

***Evaluation elaborated by***: A.Saulnier ( Ina)

***Technical contact for the service***: Andrea Esuli (CNR), GUI: Sergiu Gordea(AIT)

***Remote users test performed***: one user of CVCE

***Date***: March 28th – 30th 2012

**Objectives**

Test the usability in terms of

- Accuracy and completeness to achieve the tasks
- Easiness of achievement
- User satisfaction related of the usage of GUI functionality:
  - Knowledge extraction
  - Metadata classification

Limitation of this evaluation:

No analysis of the quality result (see 4.8).

**Methods**

User tests are performed during operating phase**.** Questionnaires are filled by the users after task achievement.

### *4.7.2 Ingestion service presentation*

The Ingestion services allows ASSETS professional users : 1) to automatically identify and extract, within metadata records, pieces of text that denote relevant entities; 2) to automatically classificy the metadata records according to a set of categories, possibly organized into a taxonomy, relevant for the domain.

The knowledge extraction involves three steps:

1. Definition of an annotation schema for a specific information extraction process (relevant type of information to be extracted from its records).
2. Definition of a training set (manually annotated records following the annotation schema ) for a specific knowledge extraction process.

3. Automatically enrich metadata records by extracting information from unstructured text.

The automatic classification involves three steps:

1. Definition of a classification schema for a specific metadata classification process.
2. Definition of a training set (at least one thousand manually classified records ) for a specific metadata classification process.
3. Classify a record according to a given taxonomy.

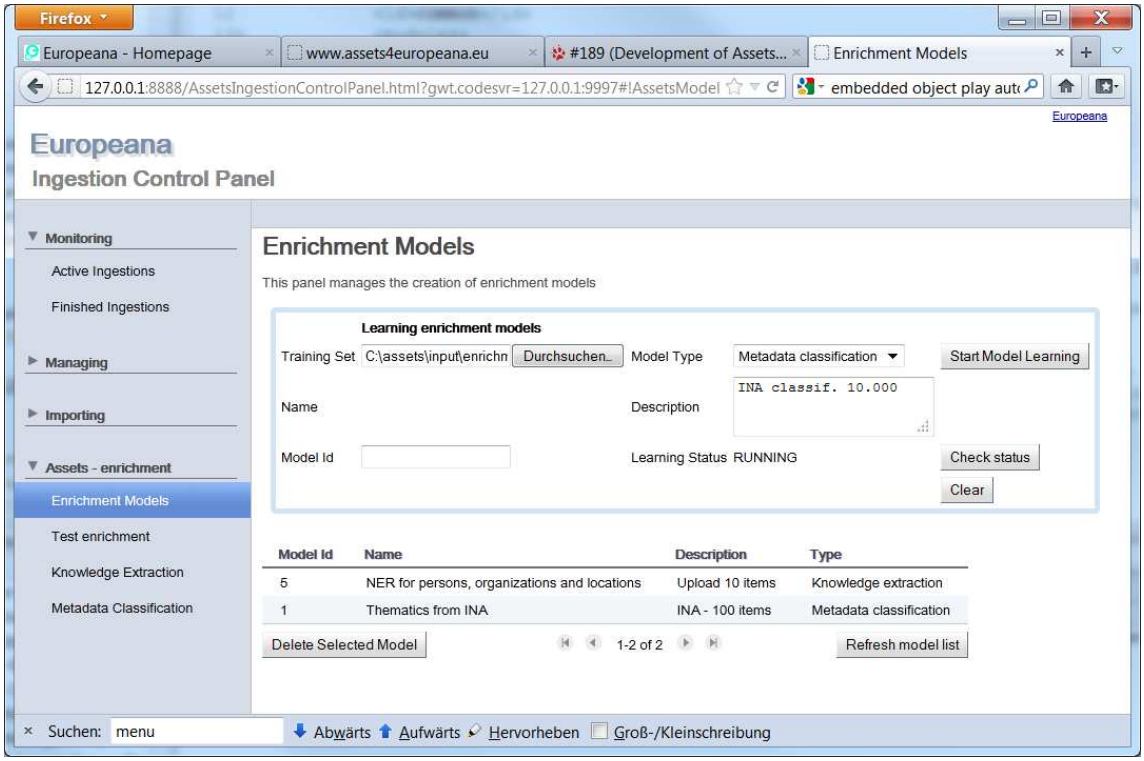



**Figure 39: Enrichment GUI**

### *4.7.3 Scenario*

Three scenarios are proposed one for knowledge extraction and one for automatic classification with three tasks each:

1. Training set definition.
2. Test enrichment.
3. Execution of knowledge extraction/metadata classification.

### *4.7.4 Results*

The users think to have mostly succeeded to achieve their task. They have encountered some running errors and interaction errors, essentially formatting of training sets, data encoding and some operational problems related to the generation of the Linux paths on the server. Nevertheless, even if they had difficulties to achieve their task, they have finally produced enriched metadata.

### 4.7.5 Analysis and conclusions

This evaluation was performed during an operational phase of the project. Some improvement of the GUI had already been done according to the user problems.

To summary the reactions of the users, the training set definition is the most delicate task to realize. The GUI is not always intuitive but could be used with some help. Input are sometimes not vey fault-tolerant. It is sometimes difficult to find errors in the input (no helpful responses).

Users expressed some difficulties but were not enough experts in this domain to propose some recommendations.

## 4.8 Ingestion: Perceived quality of metadata classification and knowledge extraction

The objective of this evaluation is to analyze the human-perceived quality of the produced enrichment. Users having their own enriched metadata on the ASSETS database are invited to analyze and comment their enriched metadata, that is to comment the enrichment done by knowledge extraction and metadata classification services. The analysis can be done for one or both of these services.

The objectives are not to evaluate the effectiveness of the service with an exhaustive quantitative analysis but to analyze a set of results and comment them. The analysis should be concentrated to conclude where the services are fine and where they should be improved.

### *4.8.1 Evaluation of Named Entities automatic extraction on Ina corpus.*

***Evaluation supervised by***: A.Saulnier (INA)

***Evaluation performed and reported by***: P.Courounet (INA)

***INA scripts done by***: S.Lalande (INA)

***Date***: May 2012

The ingested Ina corpus in ASSETS includes around 100 000 records.

To prepare the ingestion process Ina provided two learning set of ~10 000 examples: one for metadata extraction and one for metadata classification, with a list of around 60 thematic concept.

The present evaluation was leaded on the results of Named Entities (NE) extraction, for two reasons:

1. This topic presents for Ina a bigger asset than the classification one,
2. The delivered metadata were also more consistent at this level. The classification list which was given being too general to allow sure interpretation of what is a “mistake”.

As a matter of fact, results coming “speech to text”, or from OCR on manual notes increase in importance nowadays. AS much good reasons to check the efficiency of a “language independent” method on difficult texts. Despite real and constant progress, texts coming from transcription remains difficult to work with, and those coming from OCR presents by construction many syntax and grammar structure approximations due to their status of quick informal notes.

In the present case our targeted text despite (because?) manually produced by professional annotators may be considered as good sample of such “difficult” texts, for various reason:

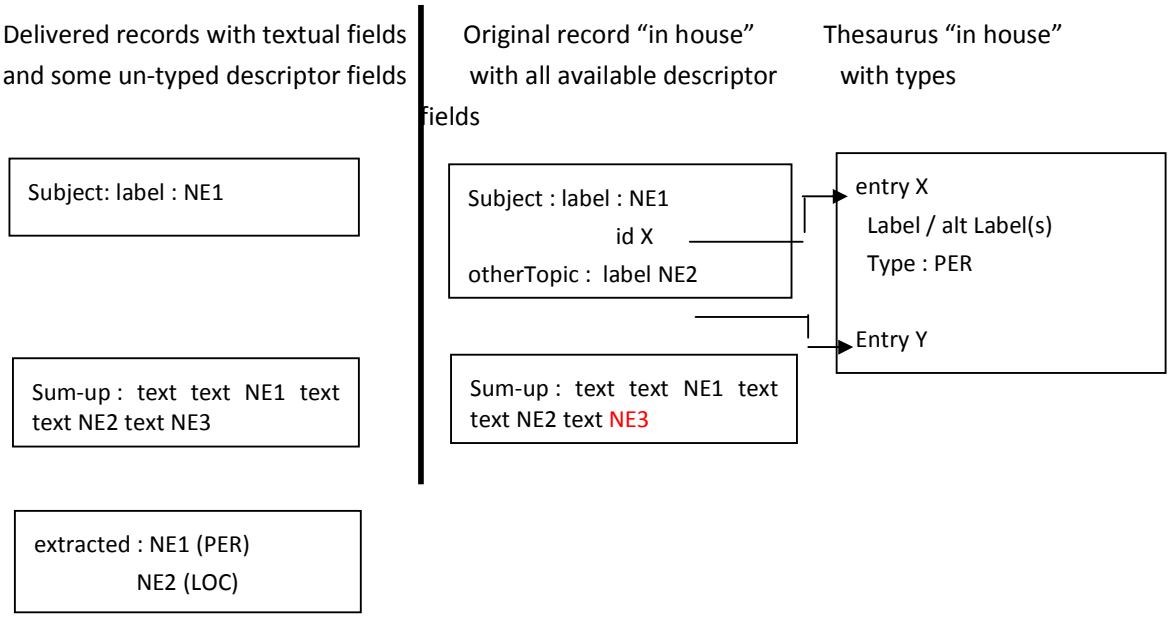
- The sentences often present a very telegraphic syntax, quite far from “good French language” and with weird punctuation convention. This is especially true for “news” description.
- Syntactically the use of capital letters obey to conventions which are far from French usual tradition.
- The adopted conventions as the one quoted above, evolved during the long history of the archives and despite some periodic corrections left different strata in the textual fields (much less maintained than the thesaurus itself). So, for example, if capital letters cannot generally be trusted for what they mean generally, any use of diacritic character cannot be trusted too.

- Finally textual field being not controlled by thesaurus, it's possible to find here various misspelling , or even alternative spellings.

### *a) First pass: automatic checking*

In a first step the resulting xml package of enriched records was transformed in RDF, and some prolog scripts were created to compare enriched records coming from ingestion with original « in house » metadata (translated also in RDF) taken as a ground truth allowing automatic check with Ina thesaurus to get the type. It must be noticed this thesaurus was not delivered to CNR who had to infer types by their own way.

This first approach was adopted as we know that many of the NE quoted in textual fields, may also be found as descriptors (thesaurus controlled), thus providing a first simple automatic checking of many extractions (no information of type are given in the delivered records and thesaurus wasn't accessible to CNR)




In the scheme above, the left part describe the material available to CNR to conduct their work, the right part shows the material available in Ina to perform any checking. From this scheme only the validity of NE1 label could have been be checked by CNR but nothing could help them to find its type.

But in Ina:

1. The validity of both NE1 & NE2 label can be automatically checked by comparison in Ina with the original record, as well as their type by following the link given between the in-house record and thesaurus.
2. Nothing can be said on NE3 label : is it a true named entity? is it really an organization ?

The following table sum-up the results of this first automatic checking: the green line corresponds to the first above case, and the yellow line to the second case.

| | occurrences | % |
|---|---|---|
| Total number of extracted NE | **212 396** | |

| | | |
|---|---|---|
| NE corresponding to already existing one and given as descriptors | 135 462 | => 58% OK |
| NE which were not given as descriptor | 76 934 | => 52% ? |

From this first pass we can consider that 58% of the ~210 000 extractions are relevant and correctly typed according the three categories targeted by this application: person, organization and location.

In this first pass we don't make any difference between a label and its occurrences. Each label were compared within the context of its own record. Thus aiming to answer the question : does "Paris" belong to the record N with Location type.

If the same label "Paris" is recognized in another record X, but not confirmed in X by a descriptor this occurrence of Paris fall in the "yellow line". No matter this label corresponds to a "valid" Label.

### *b) Second pass: Manual checking*

According the "yellow" line, among these 76 934 occurrences (of 25 805 different labels), we can have :

1. A new EN missed (or ignored ) by the annotators and/or which was never entered in the Ina Thesaurus;
2. A mistake (either in extraction or in type): the label (even if coherent) was improperly found, or interpreted ("Paris" extracted from Parisian; "Madona" typed as a location).
3. A repetition of an already existing NE label with a slightly different label (misspelling, or alternative one not included in Ina thesaurus): *Niolas* Sarkozy in place of Nicolas Sarkozy.

The purpose of the manual checking was to weight these three categories.

Important remark: We are not statisticians, there is probably more valid mathematical approach to this kind of problem. We used essentially common sense, and we describe hereafter the method to let the reader appreciate at what point our conclusions can be trusted.

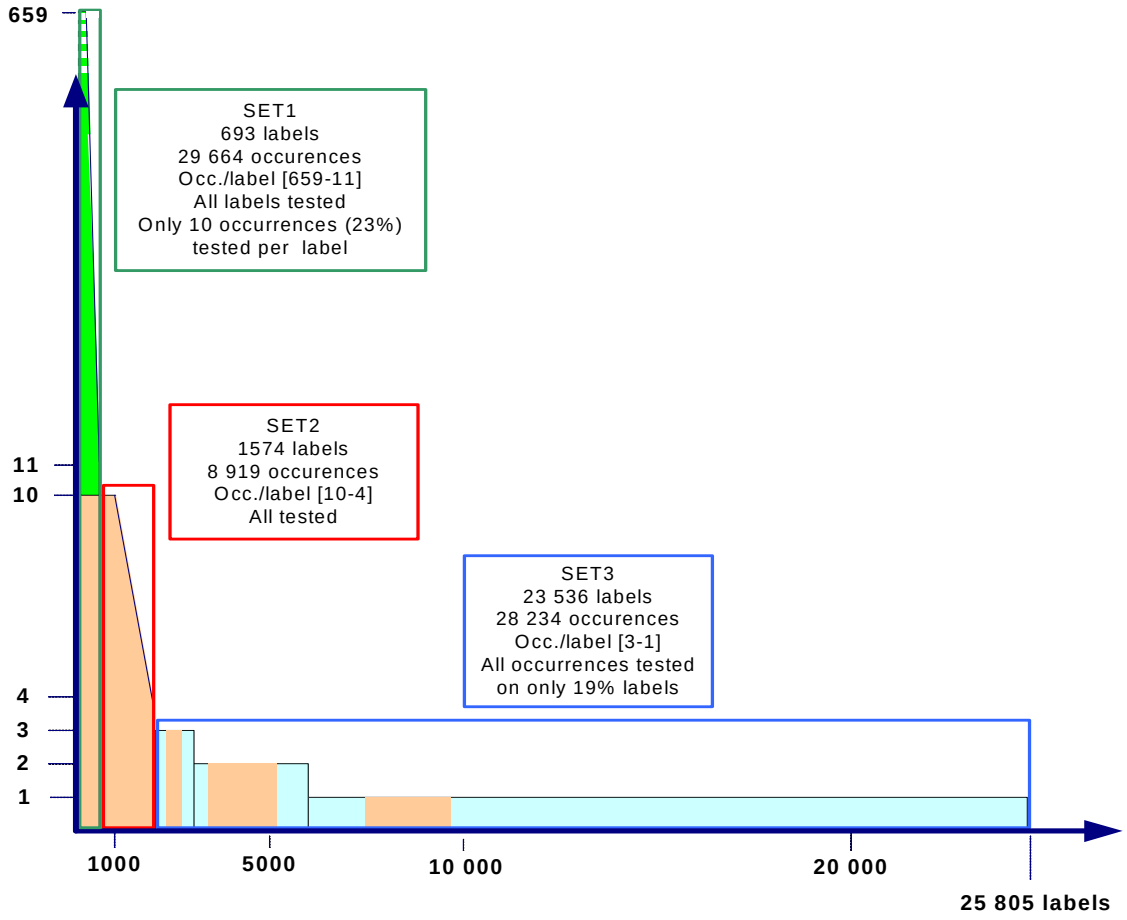
<u>The evaluation sets:</u>

This evaluation has to give some diagnostic upon ~77 000 occurrences covering around 23 000 different Named Entities labels (NE).

The entities ordered first by their number of occurrences were divided in three corpus tables:

- An excel table with all NE having more than 10 occurrences of which only the 10 extracted examples where checked, but for all the NE
- An excel table with NE between 4 and 10 occurrences which where all manually checked
- An excel table with a bunch of NE between 1 and 3 occurrences which where manually checked

Each example is linked (in house) with the original complete notice, but happily the context sentence example, is generally sufficient to check the validity of the extraction.

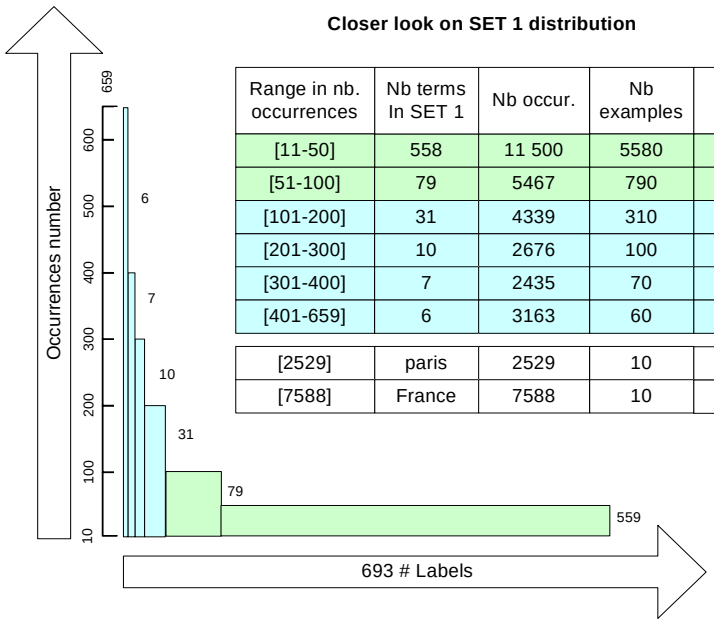



So the test corpus characteristics may be sum up in the following table (in green the 100% checked):

| SETS (nb occurrences) | Number of terms | Number of tested terms | Number of occurrences | Number of tested occurrences |
|---|---|---|---|---|
| France | 1 | 0 | 7 588 | 10* |
| Paris | 1 | 0 | 2 529 | 10* |
| **SET 1:** [11-649] | 693 | 693 (100%) | 29 664 | 6930 (23%)** |
| **SET 2:** [4-10] | 1 574 | 1 574 (100%) | 8 919 | 8919 (100%) |
| **SET 3:** [1-3] | 23 536 | 4 495 (19%) | 28 234 | 7 625 (27%) |
| total | 25 805 | 6 762 (26 %) | 76 934 | 23 474 (30%) |

* Two terms were not considered in the occurrences statistics: Paris (**2 529** occurrences) and France (**7588** occurrences).

** For terms with high numbers of occurrences (above 100) the dispositive give too few examples to be really representative. This will concern more or less 60 NE of the SET 1 as it may be seen in the table below showing the distribution for terms presenting occurrences strictly greater to 10 (the maximum number of examples we could manage in our dispositive).




**Closer look on SET 1 distribution**

| Range in nb. occurrences | Nb terms In SET 1 | Nb occur. | Nb examples | % |
|---|---|---|---|---|
| [11-50] | 558 | 11 500 | 5580 | 48,6 |
| [51-100] | 79 | 5467 | 790 | 14,5 |
| [101-200] | 31 | 4339 | 310 | 7 |
| [201-300] | 10 | 2676 | 100 | 3,7 |
| [301-400] | 7 | 2435 | 70 | 2,8 |
| [401-659] | 6 | 3163 | 60 | 1,8 |
| [2529] | paris | 2529 | 10 | 0,3 |
| [7588] | France | 7588 | 10 | 0,01 |

Structure of the excel checking extraction tables:

Three columns were generated to display:

- the label of the extracted term
- the code of the extracted type (LOC: location; PER: person: ORG: organisation),
- the number of occurrences of this entity.

And ten more columns were generated to display up to 10 examples of extracted context around the label. More examples became quickly not manageable for operators even on a wide screen for a productive work.

Nine others columns were created to track the events during the manual checking; namely:

**Events on examples (max 10 per NE)**

- ERR 1: examples: good / bad context

Needs to have many occurrences to begin a (light) problem, which seems logical. The most a term is used, the most it risks to be misused. That's why we didn't check the two highest represented terms : France and Paris as (at least in France) they are used in multiple way : in organisation name, as adjective, they can also be found (and extracted) in many composed words according the case (we saw a paris extracted from parisian in the 10 examples). Moreover 10 examples for such common words were not significant.

This kind of errors remains rather limited and may touch some very famous people: Georges Pompidou considered as a person when the annotation spoke about the hospital or the famous modern art centre, in the same way Charles de Gaulle (also recognized as a person) is more often quoted in modern documents as a location associated to the square or the airport ….

Some famous town were extracted too short "Marseille olympique" => "Marseille", thus not recognising it was a sportive club (ORG) and not the town itself (LOC). In the same way "real de Madrid", "Bayern de Munich" , "AS Monaco" were missed as well and seen as towns.

In these case Marseille cannot be seen as a bad extraction as a label , but used improperly according the context. So these errors belongs to occurrences and not to the label extraction itself which is generally coherent.

**Events on NE level:** for the following errors all the occurrences were rejected without further checking: the label being not acceptable in itself.

- ERR 2: error of bad extraction coming generally:

**When** the named entity is accompanied with irrelevant words

- AVEUX KHOL (PER) for Helmut Khol … Even if this kind of extraction could be interesting to track "events"
- Oratorio Samson de Haendel what is aimed the oratorio "SAMSON" or the composer "Haendel" ?
- The French politician Georges Pasqua were especially cherished by the software :
  - CHARLES PASQUA EN CONVERSATION AVEC JACQUES CHIRAC (2 PER in conversation …)
  - CHARLES PASQUA HIER SOIR (Charles Pasqua yesterday evening ..)
  - Charles PASQUA RPR (the man and is political party (RPR))

We may note that often, the text was in capital letters (even when it should not have been).

One UFO in this chapter (the only one): all the words are irrelevant !

- Celle-ci meurt (PER) : "this one dies"… !!!! (just one occurrence)

**When** the named entity is accompanied with another named entity, sometimes of the same type, but often of a different one:

- Alger Bouteflicka (PER)
- Cergy Ari Vatannen (PER)

The two above may be explain by the telegraphic style of some metadata, which put altogether the location followed by the actor and the action, without verbs or any other kind of transition.

- Alger VG (PER)

Here the extraction was probably fooled by the code VG which means general view ("Vue Générale"). These shots values represent a valuable knowledge for professional but may induce any automatic process in mistake.

But in some case we faced to some interrogations:

- "AUSCHWITZ" (LOC) is good; how to judge "Auschwitz en Pologne" (LOC) (Auschwitz in Poland) good or noisy ?
- "Algérie Alger" (LOC) exists as well as Algérie (LOC) and Alger (LOC) is it noise or interesting ?
- "Algérien BOUTEFLICKA" (PER) Pdt Bouteflicka is indeed Algerian, once again is it a noise or a valuable information …
- "ANTI SARKOZY l'affaire" (PER): "anti_Sarkosy the affair" is definitely not a Person, but an event. Is it noise in this context ?

We finally decide to reject all items of this category as it seems a little bit hazardous and not consistent (generally a name has no nationality, and location no included region). As we change our mind during the test it remains perhaps some of these which were counted accepted but very few.

"We rejected all" but the two following cases:

- Sir name extracted with his title:

  Amiral Ghensoul ; Ayatollah Khomeiny

  As it is rather natural to search such people with their titles. Anyway these cases are very rare (less than 10).

- Extraction of a location followed by its included site:

  Ajaccio en Corse (Ajaccio in Corsica) ; Auschwitz en Pologne" (LOC) (Auschwitz in Poland) …

  In fact each time this kind of extraction can be done it's a way to solve possible synonymies (Paris in Texas, Paris in France …) even if it is not in the announced scope of this application.

**When** the extracted term is incomplete:

- Charles VI instead of Charles VII: the only mistake of this kind (despite many other numbered pope and kings).

This incomplete extractions leads generally to wrong type…

- First name of a PER alone typed generally as LOC: Anne-Marie, Charles; Eric, Brahim… most of the time the extraction is somewhat good, it is the term which is not defined. This generally happens:
  - in fiction description where the (fiction) hero is only referenced by first name,
  - in documentaries when full identity is not given : Charles, cuisiner (coocker), etc…
  - in quotation of peoples having the same last name: Jean-Louis et Marie Trintignan (actor and actress).

  Thus Cecilia [Sarkozy] after having been the former wife of the French current president, became a Location in ASSETS… Anyway Cecilia is not a good extraction.
- Other examples of incomplete values: this time the term is complete but only the first part is extracted:

  Cac (40) and so, typed as LOC instead of CPT (concept which are not tracked by the software).

  Buena (Vista) and typed as LOC instead of ORG

  (Ku) Klux Klan typed as PER instead of ORG

… but may lead, here again, to some cases difficult to appreciate:

DOM [TOM] (overseas french department and Territories) is quoted one time as a LOC .

The location "Bande de Gaza" (the territory around Gaza in French) was extracted as a location in the following three cases:

a. "bande de Gaza"
b. "Bande de Gaza"
c. but also as "Bande", this word very common, were extracted only once in the right place.

According operator this could be judged differently: we tried in these case to think what a user could enter for a search: For a French DOM has no signification without TOM so he probably never have idea to search DOM by type, so we rejected this truncation.

For "bande de Gaza" we accepted all the cases, as "Bande" as a location for a French is uncommon enough to be search as is.

We accepted also [val de] Grâce, and [Vaison la] Romaine as these terms were always extracted and used advisedly. But perhaps wrongly in the last case, as a French user will most often search for "Vaison", than "Romaine"…

- ERR 3: Good extraction but wrong type:

Generally it seems that when two words are found a location risks in some (few) cases to be taken as a name, (e.g.: "Arabie Saoudite") and "a contrario" a single person word may become a location.

The best example is probably the singer / actress "Madonna" quoted 6 times she was recognised 4 times as a location and twice as an organisation and never (!) as a person. For someone who took so much care of her image it's really very bad luck.

This kind of error is slightly different from the context error in the sense it was judged as generic : all occurrences were considered false as not depending of the interpretation context.

Fall also in this category all extraction of topics which should have been ignored as being too generic to be Named Entities:

Cinema, bank, town, celebrity, hospital … which are clearly concepts / keywords

- ERR 4: Label ended by ] :

  We created this specific error for the case where the label and type was good except this final inclusion of a separator "]" which can be often used in Ina notices to concatenate various topics in telegraphic style. This is a 90% good error (!).

  We come back on this in the evaluation of repetition which "stricto sensu" is not in the scope of this application but could be a good entry for a cleaning process proposed at the beginning of the project but not tested by default of provider material.

Results:

| Errors types | | SET1: 11 to 649 occurences | SET 2: 4 to 10 occurences | SET 3: 1 to 3 occurrences |
|---|---|---|---|---|
| 1:Context error ** | n.a.<br>nbr of occurrences | n.a.<br>149 (2%) | n.a.<br>1 (# 0 %) | n.a.<br>3 (# 0 %) |
| 2: Bad extractions | nbr of terms<br>nbr of occurrences | 9 (1,3%)<br>222 (0,7%)* | 44 (2,7%)<br>208 (2,3 %) | 208 (4,6 %)<br>289 (3,7 %) |
| 3: Wrong type | nbr of terms<br>nbr of occurrences | 2 (0,3%)<br>32 ( 0,1%) * | 22 (1,3 %)<br>116 (1,3 %) | 102 (2,2 %)<br>148 (1,9 %) |
| 4: Label ended by ] | nbr of terms<br>nbr of occurrences | 28 (4%)<br>555 (1,8) * | 84 (5,3 %)<br>496 (5,2 %) | 219 (4,9 %)<br>405 (5,3 %) |
| total | Wrong term<br>Wrong occurences | 39 (5,6%)<br>958 (~4,6 %) * | 150 (9,3 %)<br>794 (8,8 %) | 529 (11,7 %)<br>845 (10,9 %) |

* These numbers were compared to the global number of occurrences (and not the tested one) as when a type or an extraction is false, it was considered false for any occurrences (without more checking: "Jacques Chirac" is not a location; "Marie" is first name and cannot be a named entity, "cinema" is too generic to be a named entity of type location …).

**In the contrary the context errors were only compared with the tested occurrences; this problem happens with polysemic name (e.g; swiss which can be the country or an adjective or a generic person). The term is good but sometimes wrongly used, that's why the number of term was not taken in account. This approach may be debatable.

As explained when talking about distribution of the SET 1, this result quite representative for most of the terms, becomes very weak for terms having very high number of occurrences (in red in the table).

Considering that results concerns the uncertain 52% part of the checked corpus the figures are quite good.

### *c) Third pass: repetition*

Why ?

Currently on the Ina portal (and Ina / europeana part as well) searching Lance Amstrong give back more or less 20% or the documents: 80% being (correctly) tagged Lance Armstrong.

Having an application which is able to extract with good accuracy Named Entities from free-text can be seen a first interesting step to stress this kind of opportunity and see if it is interesting to put future effort to manage this problem.

So we got interested in the noise which could be generated by repetition of the same entity under different labels.

We took the same three tables (one per SET) , ordered alphabetically to facilitate to the users the detection of this kind of problems (in many case misspelling are alphabetically closed), and added some new fields to track the number and reason of these repetitions.

We distinguished:

1. Repetitions due to an alternate label:

   e.g. : Carla Bruni / Carla Bruni Sarkozy; Ajaccio / Ajaccio en Corse

2. Repetitions due to a typo

   Beyond unusual fault (missing letter): François Mitterrand / *Fraçois* Mitterrand; Nicolas Sarkozy / *Niolas* Sarkozy

   The most common are due to possible double letters (or not): t, r, n, essentially. Here again our former president is specifically unlucky: François Mitterrand: François *Miterrand* / François *Mitterand*

3. Repetitions due to the specific "]" error

   Jacques Chirac / Jacques Chirac]

4. Repetitions due to conventions evolutions

   Use of capital letters:

   Initially all name in the text were in capital letters: ALICE SAUNIER SEITE

   Then only the last name : Alice SAUNIER SEITE

   Then sometimes just the first letter: Marcel Proust

   Ignorance, or not of diacritic character :

   Completely ignored when capital letters were systematically used, but more and more used since. Thus allowing as well:

   Alic Saunier Seite / Alice Saunier Seité

   Recent adoption of "-" in first or last name:

   Alice Saunier-Seité (to be added to variations quoted above…)

   Management of name with a handle:

   Dominique DE VILLEPIN / Dominique de VILLEPIN

5. Repetition due to wrong type
6. Global number of repeated occurrences of

The table beneath gives the number of errors and repartition % of the repetition causes

| Causes | Nb labels | % |
|---|---|---|
| alternate label | 38 | 3 |
| typo | 105 | 9 |
| ] error | 253 | 21 |
| Conventions | 667 | 56 |
| type | 126 | 11 |
| Total number of duplicated labels | 1189 | 100 |

These 1 189 repetitions concern a total 5 828 occurrences i.e. 12,6 % of the tested corpus.

### *d) Conclusion*

All of the errors described in the second pass, often funny, ***are anyway very sparse***. One remarkable aspect lies in the fact the context is very well taken in account, inducing few propagation of mistakes. When we found the first checking time a term like “announce” we were afraid to be overflow by this very common term, in fact it appears to have been confounded with a LOC only two times in the whole corpus. The same for “bank” , “cinema” “ville” (town) (5), “célébrité” (celebrity) (1) or first name like “brahim”(1), “angeline” or even words without real signification like “san” (6), or “santa” (5), …

On another hand the very common French word “Centre” was extracted advisedly just once as a location (the region “Centre” is effectively an administrative official entity).

So, even if this contextual analysis is sometimes faulty, for example the same single word label “Madonna” is extracted thrice (with lower/upper case differences) and seen once as an organisation and twice as a location, the advantages overcome greatly any inconveniences.

The number of case given above, must not hide the fact the software is really impressive. The ground field was a difficult one: sentences in telegraphic style, inclusion of codes, use of capital letters according some rules which have nothing to see with the French syntax …

We have still many data to analyse, this can just be seen as a first overview, but the result is excellent so far.

If any recommendation could be done from a naïve technical point of view (this kind of work is not our speciality) it would be to advocate processing in order to suppress the problems caused by gathering separator within the extraction (at least the “]”) as a rule or a post processing. The software being language independent this rule is perhaps more difficult to integrate that we can imagine (value of separators in different languages or, at least, alphabet…).

Anyway we can possibly imagine that this application, as is, may prepare the ground for some supervised extended cleaning process based on alert system, to help management of true and false synonyms, alternate labels or usual misspellings.

## 4.9 Digital preservation: Notification

### *4.9.1 Introduction*

This chapter presents the evaluation of the Preservation Notification service that has been conducted by using a GUI access. This service has been evaluated with ISO 9126-2 criteria because it could not be evaluated in real conditions of use like the previous systems. For the test execution, this service has been tested using the ASSETS taxonomy.

***Evaluation elaborated by***: M.Nigrelli (ENG),

***Technical contact for the service***: M.Nigrelli (ENG),

***7 Remote users test performed***: two users from FLM, three users from HASC, one user from CVCE and one user from ANSC,

***Date***: March 5th- 22th 2012.

**Objectives**

The objectives of the evaluation are:

- To evaluate the adherence of the provided functionalities to the agreed service requirements,
- To evaluate the software freedom of failures caused by faults existing in the software itself,
- To evaluate the behaviour of the service during testing (such as the occurrence of unsatisfying operation).

**Use case**

The Preservation Notification is a service of the ASSETS Framework, and it covers a very important role by providing Data Curators with the following features:

- Exchange vital knowledge of changes - that knowledge is really important for the data curator which will react in order to properly maintain the information and guarantee accessibility and usability;
- So whenever events - potentially impacting the digital preservation - are identified by system, those events are notified to the data curator…

...and this allows to trigger corrective actions (based on the suitable preservation strategy).

**Method**

This service is evaluated according to ISO 9126 evaluation framework (3.4.4) with the aid of remote tests. Users have a scenario which explains a actions to be taken (e.g. steps to followed) and the expected system beahviour (in terms of output to obtain); in order to report their experience, evaluators have to fill in matrices/tables with data related to the service that they have evaluated.

**Users**

Users involved in the tests work with some of ASSETS content providers (FLM, HASC, CVCE and ANSC).

Few users had previous knowledge about other notification service.

### 4.9.2 Preservation Notification service presentation

The purpose of this service is to support the adequate communication (notification) of events and changes which can potentially impact on long-term accessibility and usability of the digital library objects (i.e. long-term digital preservation). Notifications have to be addressed to data curators which have expressed interest and capabilities for reacting and properly maintaining the digital library objects. Asynchronous interaction is fundamental in preservation process, due to the fact that sender of notification can unknown the interested addresses. That means that senders (publishers) have to notify events/changes to a system (the preservation notification service). And consequently the system dispatches messages to interested addressees (subscribers). Subscribers register their interest for receiving messages/alerts concerning specific "topics/terms of interest" (subscriptions), and those latter are used by the dispatching rules of the service.

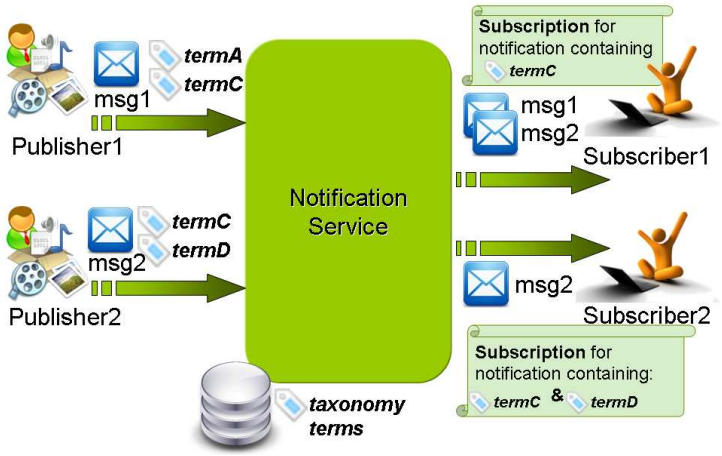



**Figure 40: Notification scenarios and concepts**

A similar, at least in the basic principles, activity is carried out by the Taxonomy-based Notification Service (Community Services), and indeed those services share common concepts.

Finally, the Preservation Notification provides common features for supporting the Taxonomy-based Notification.

### 4.9.3 Scenario description

The following tasks have to be followed in order to complete the test:

Step 1

- A subscription on topics/terms of interest of the ASSETS taxonomy is registered by a user previously authenticated on the ASSETS portal (thus becoming a subscriber); by registering a subscription, he/she will receive alerts whenever contents related to the terms are published by a content provider (a publisher);

Step 2

- All the ASSETS subscription and subscribers will be displayed;

Step 3

- A publisher (e.g. a content provider) is created;

Step 4

- The already created publisher (e.g. a content provider) creates notification channels on specific terms from the ASSETS taxonomy;

Step 5

- A publisher (e.g. a content provider) publishes messages on the already created notification channel/s;

Step 6

- Messages are delivered by selecting the filtering rules according to which subscribers will receive messages related only to the terms they have previously subscribed for; messages will be delivered through different interfaces (ASSETS Notification Service home page, REST URL, Google Reader,)

### *4.9.4 Results and metrics*

*Functional adequacy*

All the users have found that, with regards to the tasks performed, all the evaluated functionalities were adequate.

| **SCENARIO STEPS**<br>**Suitability** *Functional adequacy*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br>**FUNCTIONALITY** | | Step 1 – Register a subscription on the ASSETS portal | Step 2 – View subscriptions and subscribers (in ASSETS) | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Publishing messages | Step 6.1 – Delivery of messages (from the Preservation Notification service home page) | Step 6.2 – Delivery of messages (via a REST URL on a browser) | Step 6.3 – Delivery of messages (via a RSS reader on a web browser) |
|---|---|---|---|---|---|---|---|---|---|
| create a new service | FA 1 | | | | | | | | |
| create a new taxonomy (for a newly created or for an existing service) | FA 2 | | | | | | | | |
| view an existing taxonomy | FA 3 | | | | | | | | |
| create a new publisher | FA 4 | | | 7 | | | | | |
| create a new notification with topics/terms | FA 5 | | | | 7 | | | | |
| publishing messages for an existing notification | FA 6 | | | | | 7 | | | |
| create a new subscriber | FA 7 | 7 | | | | | | | |
| create a new subscription for terms of an existing taxonomy | FA 8 | 7 | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| view subscribers / subscriptions | FA 9 | | 7 | | | | | | |
| dispatch messages according to the delivery rules / deliver messages for a specific set of terms, according to delivery rules | FA 10 | | | | | | 7 | 7 | 7 |

**Table 9: Notification Functional adequacy - summary users answers**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A= Number of functions in which problems are detected in evaluation: A=0*

*B= Number of functions evaluated: B=7*

So **X=1**

*Functional implementation completeness*

All the users have found that, with regards to the tasks performed, all the evaluated functionalities were adequate.

| **SCENARIO STEPS**<br><br>**Usability** *Function understandability*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Step 1 – Register a subscription on the ASSETS portal | Step 2 – View subscriptions and subscribers (in ASSETS) | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Publishing messages | Step 6.1 – Delivery of messages (from the Preservation Notification service home page) | Step 6.2 – Delivery of messages (via a REST URL on a browser) | Step 6.3 – Delivery of messages (via a RSS reader on a web browser) |
|---|---|---|---|---|---|---|---|---|---|
| create a new service | FA 1 | | | | | | | | |
| create a new taxonomy (for a newly created or for an existing service) | FA 2 | | | | | | | | |
| view an existing taxonomy | FA 3 | | | | | | | | |
| create a new publisher | FA 4 | | | 7 | | | | | |
| create a new notification with topics/terms | FA 5 | | | | 7 | | | | |
| publishing messages for an existing notification | FA 6 | | | | | 7 | | | |
| create a new subscriber | FA 7 | 7 | | | | | | | |
| create a new subscription for terms of an existing taxonomy | FA 8 | 7 | | | | | | | |
| view subscribers / subscriptions | FA 9 | | 7 | | | | | | |
| dispatch messages according to the delivery rules / deliver messages for a specific set of | FA 10 | | | | | | 7 | 7 | 7 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| terms, according to delivery rules | | | | | | | | | |

**Table 10: Notification Functional implementation - summary of user answers**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A = Number of missing functions detected in evaluation: A=0*

*B = Number of functions described in requirement specifications: B=6*

So **X=1**

*Function understandability*

All the users have found that they were able to understand all the functionalities correctly.

| **SCENARIO STEPS**<br>**Usability** *Function understandability*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br>**FUNCTIONALITY** | | Step 1 – Register a subscription on the ASSETS portal | Step 2 – View subscriptions and subscribers (in ASSETS) | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Publishing messages | Step 6.1 – Delivery of messages (from the Preservation Notification service home page) | Step 6.2 – Delivery of messages (via a REST URL on a browser) | Step 6.3 – Delivery of messages (via a RSS reader on a web browser) |
|---|---|---|---|---|---|---|---|---|---|
| create a new service | FA 1 | | | | | | | | |
| create a new taxonomy (for a newly created or for an existing service) | FA 2 | | | | | | | | |
| view an existing taxonomy | FA 3 | | | | | | | | |
| create a new publisher | FA 4 | | | **7** | | | | | |
| create a new notification with topics/terms | FA 5 | | | | **7** | | | | |
| publishing messages for an existing notification | FA 6 | | | | | **7** | | | |
| create a new subscriber | FA 7 | **7** | | | | | | | |
| create a new subscription for terms of an existing taxonomy | FA 8 | **7** | | | | | | | |
| view subscribers / subscriptions | FA 9 | | **7** | | | | | | |
| dispatch messages according to the delivery rules / deliver messages for a specific set of terms, according to delivery rules | FA 10 | | | | | | **7** | **7** | **7** |

**Table 11: Notification understandability - summary users answers**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of functions whose purpose is correctly understood by the user: A=7*

*B= number of available functions: B=7*

So **X=1**

*Understandable input and output*

All the users have found that they were able to understand completely what is required by the system both as input and as output.

| **SCENARIO STEPS**<br><br>**Usability** *Understandable input and output y*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Step 1 – Register a subscription on the ASSETS portal | Step 2 – View subscriptions and subscribers (in ASSETS) | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Publishing messages | Step 6.1 – Delivery of messages (from the Preservation Notification service home page) | Step 6.2 – Delivery of messages (via a REST URL on a browser) | Step 6.3 – Delivery of messages (via a RSS reader on a web browser) |
|---|---|---|---|---|---|---|---|---|---|
| create a new service | FA 1 | | | | | | | | |
| create a new taxonomy (for a newly created or for an existing service) | FA 2 | | | | | | | | |
| view an existing taxonomy | FA 3 | | | | | | | | |
| create a new publisher | FA 4 | | | **7** | | | | | |
| create a new notification with topics/terms | FA 5 | | | | **7** | | | | |
| publishing messages for an existing notification | FA 6 | | | | | **7** | | | |
| create a new subscriber | FA 7 | **7** | | | | | | | |
| create a new subscription for terms of an existing taxonomy | FA 8 | **7** | | | | | | | |
| view subscribers / subscriptions | FA 9 | | **7** | | | | | | |
| dispatch messages according to the delivery rules / deliver messages for a specific set of terms, according to delivery rules | FA 10 | | | | | | 7 | 7 | 7 |

**Table 12: Notification understandability input output - summary users answers**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of input and output data items which user successfully understands: A=7*

*B= number of input and output data items available from the system: B=7*

So **X=1**

*Maturity*

No user has encountered failures during the evaluation of the proposed scenario.

| **SCENARIO STEPS** / **Maturity** **Table 8.2.1 Maturity metrics** of the Norm ISO_IEC 9126-2 / **MEASUREMENTS** | Step 1 – Register a subscription on the ASSETS portal | Step 2 – View subscriptions and subscribers (in ASSETS) | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Publishing messages | Step 6.1 – Delivery of messages (from the Preservation | Step 6.2 – Delivery of messages (via a REST URL on a browser) | Step 6.3 – Delivery of messages (via a RSS reader on a web |
|---|---|---|---|---|---|---|---|---|
| Number of failures | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| Number of tests performed | **9** | **9** | **9** | **10** | **10** | **10** | **9** | **9** |

**Table 13: Notification maturity - summary users answers**

Measurement, formula and data element computations: **X= A1 / A2** where:

A1 = number of detected failures: A1=0

A2 = number of performed test cases: A2=75

So **X=0**

*Functionality compliance*

| **SCENARIO** / **Functionality compliance** **Table 8.1.5 Functionality compliance metrics** of the Norm ISO_IEC 9126-2 / **REQUIREMENT TO BE COMPLIANT WITH** | | | Test Scenario | Number of users |
|---|---|---|---|---|
| FC1 | OAIS Preservation Planning | Any change event needs to be properly identified (i.e. monitored and captured), notified (i.e. published and delivered) in order to exchange the information of change within the archive environment.<br>This component is mainly an implementation of the publish-subscribe pattern, and for its responsibility it may be considered a component of the OAIS Preservation Planning.<br>The publish/subscribe interaction paradigm has two main actors: users of a digital library, or subscribers, with the ability | 1<br>2<br>3 | <br><br>5 |

| | | | | |
|---|---|---|---|---|
| | | to express their interest in classes of events generated by content providers, or publishers.<br>A system that supports this paradigm must be able to find the subscriptions that match each incoming event, in order to determine which subscribers should be notified. | | |
| FC2 | OAIS RM ( ISO:14721:2003) – Provenance | By gathering and tracing the notifications for occurring events on a content, it is possible to build part of the Preservation Description Information (the so called Provenance) This is a key element of the OAIS Reference Model ISO:14721:2003 and it is important for providing evidence of what occurred on any archived content. | 1<br>2<br>3 | <br><br>4 |
| FC3 | OAIS RM (ISO:14721:2003) – Reference | Any object in a OAIS-compliant system has to be univocally referenced through a Persistent Identifier. The Notification is able to generate identifiers and to use external identifiers. Those identifiers can be managed by a Persistent Identifier Service (e.g. purl, doi). | 1<br>2<br>3 | <br><br>5 |
| FC4 | CASPAR-POM | During the CASPAR Project (which dealt with Digital Preservation and OAIS Reference Model), a Preservation Orchestration Manager (acronym CASPAR-POM or simply POM) has been modelled and developed. This component was responsible for submitting notifications from Data Preservers (role of publishers). Data Holders/Curators can receive alerts based on their expressed interests/expertises. The CASPAR-POM models can be reasonably considered for the implementation of the ASSETS Preservation Notification. | 1<br>2<br>3 | 1<br><br>4 |
| FC5 | Digital Libraries | Digital Libraries (DL) maintain (in a repository) descriptions of documents and pointers to documents' contents. In this context a publisher is an author that provides the DL with descriptions of his/her documents and with ways to access their contents (e.g. their URIs), whereas a subscriber is a user wanting to be informed of any event affecting a document that relates to his/her topics of interest. We consider a DL model with the following characteristics:<br>• There is a taxonomy to which the authors of documents and the subscribers of the library both adhere; this taxonomy is just a set of keywords (or terms) structured in a hierarchy. An example of a taxonomy is the well known ACM Computing Classification System.<br>• A document is represented in the DL repository by a description of its content together with an identifier (e.g. the document's URI) allowing to access the document's content; the description is just a set of terms from the taxonomy.<br>• A query is just a set of terms from the DL taxonomy (i.e. a conjunctive query)<br>• A user is represented by an identifier together with a subscription; a subscription is just a query defining (intentionally) the documents of interest to the user. | 1<br>2<br>3 | 2<br>2<br>1 |

**Table 14: Notification functionality compliance - summary users answers**

level of compliance (column) of the evaluated service and its features (1 = low, 2=medium, 3=high).

This last matrix shows that the Preservation Notification has been considered capable to adhere with a high level of compliance to

- OAIS Preservation Planning (a model),
- OAIS RM ( ISO:14721:2003) – Provenance (a standard),
- OAIS RM (ISO:14721:2003) – Reference (a standard),
- CASPAR-POM (an EU project with similar prescriptions relating to functionality).

On the contrary, the not too good results obtained for the last "item" of compliance can be partly attributed to the vastness of the subject; possibly, the aspects of the Digital Libraries on which the evaluation was focussed were not too relevant with the functionalities tested throughout the scenario.

### *4.9.5 Conclusions*

The functional testing of the Preservation Notification service has arisen no issues according to the predefined criteria and to the tested scenario:

- All the evaluated functions are suitable to perform the specific tasks,
- There were not missing functions to perform the specific tasks,
- Users understand easily the purpose of each functions,
- Users understand what is required as input and what is provided as output at each step of the tests,
- The maturity metrics (belonging to reliability characteristic) is very good because there was no failure caused by faults existing in the software itself.

An external functionality compliance table (Table 14) has also been filled. It shows that the notification service has been considered to be highly compliant with most of the standards, models and similar EU projects selected for the evaluation of functionality compliance.

## 4.10 Digital preservation : Risk management & normalization

### *4.10.1 Introduction*

This chapter presents the evaluation of the Preservation Risk management service that has been conducted by using a GUI access.

***Evaluation elaborated by***: Andrew Lindley, Roman Graf, Sergiu Gordea

***Technical contact for the service***: Andrew Lindley, Roman Graf, Sergiu Gordea (AIT)

***Remote users test performed***: CVCE

***Date***: March 28th-30th 2012.

**Objectives**

The objectives of the evaluation are:

- To evaluate the adherence of the provided functionalities to the agreed service requirements,
- To evaluate the software freedom of failures caused by faults existing in the software itself,
- To evaluate the behaviour of the service during testing (such as the occurrence of unsatisfying operation).

**Use case**

Technical Use Case based on D2.3.1 "Prototype Preservation Services". The Preservation Risk Management is a service of the ASSETS Framework, and it covers a very important role by providing metadata analysis and preservation risks estimation for Europeana collections.

**Method**

This service is evaluated according to remote user tests. Users have a scenario which explains a actions to be taken (e.g. steps to followed) and the expected system behavior (in terms of output to obtain); in order to report their experience.

### *4.10.2 Preservation Risk management service presentation*

The purpose of this service is to support Europeana collections content classification and profiling. The service computes risk analysis for given data and provides support for preservation planning. The Risk Management service leverages Linked Open Data repositories to provide preservation technology watch.

The service aims at mitigating the risk of digital obsolescence by providing risk management reports to content providers - which is done by analyzing the contributed content. The service performs object inspection and statistical analysis of the content formats at hand to categorize them based on their preservation risk. The service is robust in terms of reliably classifying the underlying data on basis of available metadata and giving solid preservation recommendations. The component beyond this addresses the topics of technology watch and (semi)automated preservation policies. It will make use of available preservation community resources such as technical registries (like PRONOM,

Freebase and DBPedia) for policy extraction and the Assets Normalisation Service for object identification and policy execution.

The service addresses the problem of:

- format obsolescence and limited support for proprietary formats;
- unstructured and unknown digital collections;
- automated collection profiling and recommendation of fine-tuned preservation actions.

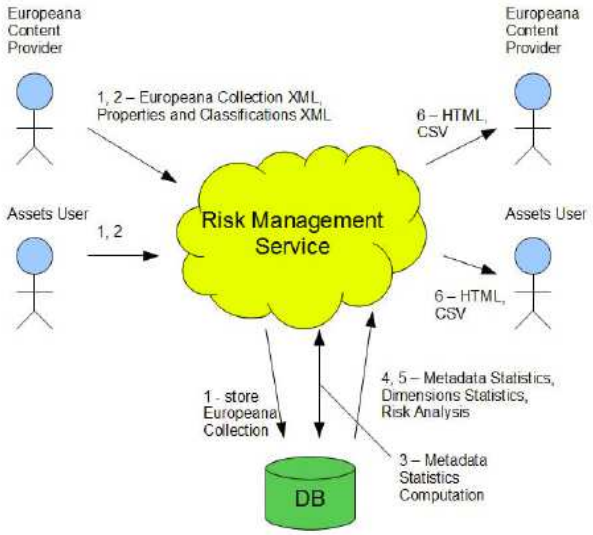



**Figure 41: Activity diagram for Risk Management user evaluation scenario 1 "Europeana collection preservation risks evaluation"**

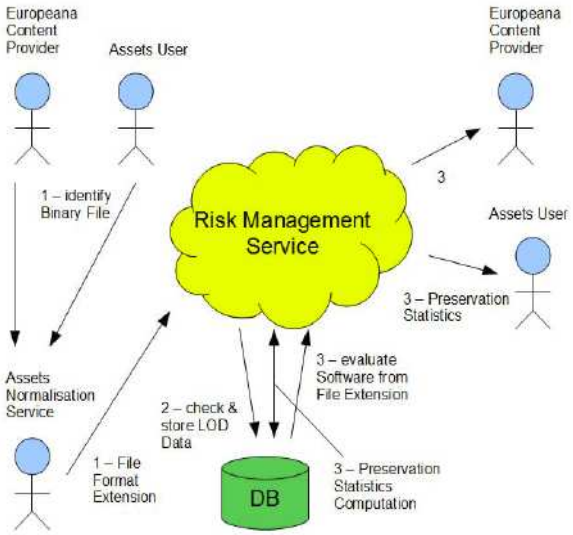



**Figure 42: Activity diagram for Risk Management user evaluation scenario 2 "Evaluation of software that supports preservation plan for particular file format"**

### 4.10.3 Scenario

This section presents the basic idea on how to use the Preservation Risk Management APIs for supporting the test scenarios: “The Preservation Risk Management Test Scenario”.

The following preparations tasks have to be followed in order to perform the test:

- Step 1 – The risk and LOD properties containing classifications, preservation rules and repository queries are defined in XML property files;
- Step 2 – The file formats and associated data is automatically retrieved from LOD repositories;
- Step 3 – A Risk Management service is instantiated for accessing the REST risk management services;
- Step 4 – A risk analysis calculation model is instantiated based on acquired properties.

### 4.10.4 Results and analysis conducted by CVCE

CVCE performs an evaluation of the preservation risk management service in ASSETS. Only the functions available through the on line graphic user interface have been tested.

***Evaluation of software and vendor that support preservation plan for particular file format***

The concerned services are providing information about:

- A file format
- Software solutions associated to a file format.
- Vendors associated to a file format.

| | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| **1** | **Usefulness** | | | | X | |
| **2** | **Completeness** | | | | X | |
| **3** | **Quality** | | X | | | |
| **4** | **Understandability of the service** | | | | | X |
| **5** | **Understandability of the output** | | X | | | |

1 = poor

5 = rich

**Additional comments**

These services are quite easy to understand and could be very useful to people who work with digital media. It seems that the system could not provide some information for some formats like .pages (Apple) and .docx (when a request for ‘doc’ is giving valuable answers). Sometimes, the system states for one result without filling out the associated table (which is blank). The results are also produced

in a very rough way.

***Europeana collection preservation risks evaluation***

The concerned services are providing information about:

- A metadata analysis for a collection
- Metadata statistics report for a collection
- Preservation dimension (Provenance, Context, Accessibility) risk score for a collection
- Overall risk score for a collection

| | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| **1** | **Usefulness** | | | | X | |
| **2** | **Completeness** | | | | X | |
| **3** | **Quality** | | | X | | |
| **4** | **Understandability of the service** | | | X | | |
| **5** | **Understandability of the output** | | X | | | |

1 = poor

5 = rich

**Additional comments:**

There were first some difficulties to understand the context and usage of this service. With the support by AIT we were finally able to use the service in a constricted way and to make use of our own collections.

The proposed services are very interesting but they depend a lot on the completeness of your ESE records for the fields that are analysed to provide the results. As these fields are not mandatory in ESE, the service will be only useful in the rare case of the presence of these fields. Like for the 'file format' services, the results are very rough and the length of the results may generate sometimes confusion. The server is sometimes also taking a very long time to answer (depending on the number of records in a collection).

**Sample results of the metadata analysis of CVCE contributions collection:**

{ "DomainObjectName" : "CollectionAnalysisReport" , "ComponentName" : "preservation-riskmanagement" , "BrokenIsShownAtCount" : 0 , "BrokenIsShownByCount" : 0 , "MissingDcCreatorCount" : 12 , "MissingDcFormatCount" : 0 , "MissingDcTypeCount" : 0 , "TotalBrokenObjects" : 12 , "TotalAnalyzedObjects" : 12 , "CollectionId" : "182900" , "CollectionName" : "07711_AV_CVCE_ASSETS_contribution" , "CreateTime" : 1333092348314 , "LastUpdateTime" : 1333092348314 , "_id" : { "$oid" : "4f755ffc9f50b0e48794d827"}}

**Sample results of the overall risk score of CVCE contributions collection:**

***Collection Analysis Report***

| ID | NAME | BrokenIsShownAtCount | BrokenIsShownByCount | MissingDcTypeCount | MissingDcCreatorCount | MissingDcFormatCount | TotalBrokenObjects | TotalAnalyzedObjects |
|---|---|---|---|---|---|---|---|---|
| 1829 00 | 07711_AV_CVCE_ASSETS_contribution | 0 | 0 | 0 | 12 | 0 | 12 | 12 |

Overview

***Preservation Score: RiskAnalysis***

| Preservation Dimension | Average Risk Value | Risk Score (min=0.0, max=1.0) | Risk Level | Total Risk Score | Total Risk Level |
|---|---|---|---|---|---|
| BrokenObjectsScore | 1.0 | 1.0 | High | 1.0 | High |

# 5. Evaluation of professional services using an API access

The evaluations of professional services to be conducted by using an API access have been elaborated from a selection of criteria extracted from ISO 9126-2. The shared objectives are to test these services from the functionality (suitability, functionality compliance), from the reliability (maturity) and from the usability (understandability) point of view (see 3.1 and 3.4.4):

- Functionality: **Suitability**: functional adequacy.
  How adequate are the evaluated functions?
  This measure is obtained calculating the occurrence of unsatisfying function during testing.
- Functionality: **Suitability:** functional implementation completeness
  How complete is the implementation according to requirement specifications?
  This measure is obtained calculating the number of missing functions detected in the evaluation.
- Functionality**: Functionality compliance**:
  How compliant is the functionality of the service to other requirement (*here requirements list extracted from other projects*)
  This measure is obtained calculating the number of compliance items that have been satisfied.
- Reliability**: maturity**:
  How many failures were detected during the test?
  This measure is obtained calculating the number of detected failure.
- Usability**: understandability**: Functions understandability
  What proportion of the service functions will the user be able to understand correctly?
  This measure is obtained calculating the number of functions whose purpose is correctly understood by the user.
- Usability**: understandability**: understandable input and output
  Can users understand what is required as input data and what is provided as output by the system?
  This measure is obtained calculating the number of input and output data items which user successfully understands.

The evaluation protocol of each of these services has been elaborated by the service provider. Each service has been evaluated by developers who were not involved in its development. A scenario has been defined to describe the tests the users had to perform in order to be able to calculate the above mentioned measures.

## 5.1 Preservation Notification service

### *5.1.1 Introduction*

**Evaluation elaborated and completed by**: Massimiliano Nigrelli (ENG),

**Evaluation performed by**: CNR,

**Date of evaluation performance:** March 21th 2012.

The Preservation Notification service is presented in subsection 4.9.

The objectives and methodology of this evaluation are similar to the ones that are described for the evaluation of Preservation Notification by GUI access.

The evaluation of this service using an API access is intended to developers who might want to add the above mentioned features to their digital libraries.

### *5.1.2 Results*

*Functional adequacy*

| **SCENARIO STEPS**<br><br>**Suitability** *Functional implementation completeness*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONAL REQUIREMENTS** | | Step 1 – The instance of a dummy client service | Step 2 – Creating a dummy taxonomy | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Creating a dummy subscriber | Step 6 – Creating a dummy subscription | Step 7 – Publishing messages | Step 8 – Delivery of alerts |
|---|---|---|---|---|---|---|---|---|---|
| The service has to make available its features to external services. Each service could have different taxonomies, subscribers and subscriptions. | FIC 1 | X | | | | | | | |
| The service has to be fed with a taxonomy that have only one parent/broader and many children/narrower | FIC 2 | | X | | | | | | |
| A registered Publisher is enabled to submit notification whenever impacting event occurs | FIC 3 | | | X | | | | | |
| A registered Publisher is enabled to submit more than one notification (is able to create multiple notification channels). | FIC 4 | | | | X | | | | |
| Each notification may have more than one topic/term. | FIC 5 | | | | X | | | | |
| For one or more terms of a pre-loaded Taxonomy (within a notification channel) a | FIC 6 | | | | | | | X | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| registered Publisher is enabled to submit a series of messages, each of which would notify an event (alerts). | | | | | | | | | |
| Delivery of alerts (notifying messages) has to be based on filtering rules such as AND, OR and EXACT. | FIC 7 | | | | | | | | X |
| A subscription for one or more terms of a pre-loaded taxonomy can be created | FIC 8 | | | | | | | X | |
| A registered Subscriber is enabled to receive alert related to impacting events for carrying out planned procedures | FIC 9 | | | | | | X | | |

**Table 15: Notification Functional adequacy**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A= Number of functions in which problems are detected in evaluation: A=0*

*B= Number of functions evaluated: B=9*

So **X=1**

*Functional implementation completeness*

| **SCENARIO STEPS**<br><br>**Suitability** *Functional implementation completeness*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONAL REQUIREMENTS** | | Step 1 – The instance of a dummy client service | Step 2 – Creating a dummy taxonomy | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Creating a dummy subscriber | Step 6 – Creating a dummy subscription | Step 7 – Publishing messages | Step 8 – Delivery of alerts |
|---|---|---|---|---|---|---|---|---|---|
| The service has to make available its features to external services. Each service could have different taxonomies, subscribers and subscriptions. | FIC 1 | X | | | | | | | |
| The service has to be fed with a taxonomy that have only one parent/broader and many children/narrower | FIC 2 | | X | | | | | | |
| A registered Publisher is enabled to submit notification whenever impacting event occurs | FIC 3 | | | X | | | | | |
| A registered Publisher is enabled to submit more than one notification (is able to create multiple notification channels). | FIC 4 | | | | X | | | | |
| Each notification may have more than one topic/term. | FIC 5 | | | | X | | | | |
| For one or more terms of a pre-loaded | FIC | | | | | | | X | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Taxonomy (within a notification channel) a registered Publisher is enabled to submit a series of messages, each of which would notify an event (alerts). | 6 | | | | | | | | |
| Delivery of alerts (notifying messages) has to be based on filtering rules such as AND, OR and EXACT. | FIC 7 | | | | | | | | X |
| A subscription for one or more terms of a pre-loaded taxonomy can be created | FIC 8 | | | | | | | X | |
| A registered Subscriber is enabled to receive alert related to impacting events for carrying out planned procedures | FIC 9 | | | | | | X | | |

**Table 16: Notification Functional implementation**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A = Number of missing functions detected in evaluation: A=0*

*B = Number of functions described in requirement specifications: B=9*

So **X=1**

*Function understandability*

| **SCENARIO STEPS**<br><br>**Usability** *Function understandability*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Step 1 – The instance of a dummy client service | Step 2 – Creating a dummy taxonomy | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Creating a dummy subscriber | Step 6 – Creating a dummy subscription | Step 7 – Publishing messages | Step 8 – Delivery of alerts |
|---|---|---|---|---|---|---|---|---|---|
| create a new service | FA 1 | X | | | | | | | |
| create a new taxonomy (for a newly created or for an existing service) | FA 2 | | X | | | | | | |
| view an existing taxonomy | FA 3 | | | | | | | | |
| create a new publisher | FA 4 | | | X | | | | | |
| create a new notification with topics/terms | FA 5 | | | | X | | | | |
| publishing messages for an existing notification | FA 6 | | | | | | | X | |
| create a new subscriber | FA 7 | | | | | X | | | |
| create a new subscription for terms of an existing taxonomy | FA 8 | | | | | | X | | |
| view subscribers / subscriptions | FA 9 | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| dispatch messages according to the delivery rules / deliver messages for a specific set of terms, according to delivery rules | FA 10 | | | | | | | | X |

**Table 17: Notification understandability**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of functions whose purpose is correctly understood by the user: A=8*

*B= number of available functions: B=8*

So **X=1**

*Understandable input and output*

| **SCENARIO STEPS**<br><br>**Usability** *Understandable input and output*<br>*y*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Step 1 – The instance of a dummy client service | Step 2 – Creating a dummy taxonomy | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Creating a dummy subscriber | Step 6 – Creating a dummy subscription | Step 7 – Publishing messages | Step 8 – Delivery of alerts |
|---|---|---|---|---|---|---|---|---|---|
| create a new service | FA 1 | 1/1 | | | | | | | |
| create a new taxonomy (for a newly created or for an existing service) | FA 2 | | 3/3 | | | | | | |
| view an existing taxonomy | FA 3 | | | | | | | | |
| create a new publisher | FA 4 | | | 1/1 | | | | | |
| create a new notification with topics/terms | FA 5 | | | | 2/2 | | | | |
| publishing messages for an existing notification | FA 6 | | | | | | | 2/2 | |
| create a new subscriber | FA 7 | | | | | 1/1 | | | |
| create a new subscription for terms of an existing taxonomy | FA 8 | | | | | | 3/3 | | |
| view subscribers / subscriptions | FA 9 | | | | | | | | |
| dispatch messages according to the delivery rules / deliver messages for a specific set of terms, according to delivery rules | FA 10 | | | | | | | | 2/2 |

**Table 18: Notification understandability input output**

This table shows the number of significant (IMO) Input/Output parameters given as arguments to Java
classes: N/M
N: number of significant parameters

M: total number of parameters

Measurement, formula and data element computations: **X = A / B** where:

*A= number of input and output data items which user successfully understands: A=15*

*B= number of input and output data items available from the system: B=15*

So **X=1**

*Maturity*

| **SCENARIO STEPS**<br><br>**Maturity**<br>**Table 8.2.1 Maturity metrics** of the Norm ISO_IEC 9126-2<br><br>**MEASUREMENTS** | Step 1 – The instance of a dummy client service | Step 2 – Creating a dummy taxonomy | Step 3 – Creating a dummy publisher | Step 4 – A dummy publisher creates notification(s) | Step 5 – Creating a dummy subscriber | Step 6 – Creating a dummy subscription | Step 7 – Publishing messages | Step 8 – Delivery of alerts |
|---|---|---|---|---|---|---|---|---|
| Presence of failures | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Number of tests performed | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

**Table 19: Notification maturity**

Measurement, formula and data element computations: **X= A1 / A2** where:

A1 = number of detected failures: A1=0

A2 = number of performed test cases: A2=40

So **X=0**

*Functionality compliance*

| **SCENARIO**<br><br>**Functionality compliance**<br>**Table 8.1.5 Functionality compliance metrics** of the Norm ISO_IEC 9126-2<br><br>**APPLICABLE REGULATIONS, STANDARDS AND CONVENTIONS TO BE COMPLIANT WITH** | | | Test Scenario for the "Preservation Notification" |
|---|---|---|---|
| FC1 | OAIS Preservation Planning | Any change event needs to be properly identified (i.e. monitored and captured), notified (i.e. published and delivered) in order to exchange the information of change within the archive environment.<br>This component is mainly an implementation of the publish-subscribe pattern, and for its responsibility it may be considered a component of the OAIS Preservation Planning.<br>The publish/subscribe interaction paradigm has two main actors: users of | 1<br>**2**<br>3 |

| | | | |
|---|---|---|---|
| | | a digital library, or subscribers, with the ability to express their interest in classes of events generated by content providers, or publishers.<br>A system that supports this paradigm must be able to find the subscriptions that match each incoming event, in order to determine which subscribers should be notified. | |
| FC2 | OAIS RM ( ISO:14721:2003) – Provenance | By gathering and tracing the notifications for occurring events on a content, it is possible to build part of the Preservation Description Information (the so called Provenance) This is a key element of the OAIS Reference Model ISO:14721:2003 and it is important for providing evidence of what occurred on any archived content. | 1<br>2<br>**3** |
| FC3 | OAIS RM (ISO:14721:2003) – Reference | Any object in a OAIS-compliant system has to be univocally referenced through a Persistent Identifier. The Notification is able to generate identifiers and to use external identifiers. Those identifiers can be managed by a Persistent Identifier Service (e.g. purl, doi). | 1<br>2<br>**3** |
| FC4 | CASPAR-POM | During the CASPAR Project (which dealt with Digital Preservation and OAIS Reference Model), a Preservation Orchestration Manager (acronym CASPAR-POM or simply POM) has been modelled and developed. This component was responsible for submitting notifications from Data Preservers (role of publishers). Data Holders/Curators can receive alerts based on their expressed interests/expertises. The CASPAR-POM models can be reasonably considered for the implementation of the ASSETS Preservation Notification. | 1<br>**2**<br>3 |
| FC5 | Digital Libraries | Digital Libraries (DL) maintain (in a repository) descriptions of documents and pointers to documents' contents. In this context a publisher is an author that provides the DL with descriptions of his/her documents and with ways to access their contents (e.g. their URIs), whereas a subscriber is a user wanting to be informed of any event affecting a document that relates to his/her topics of interest. We consider a DL model with the following characteristics:<br>• There is a taxonomy to which the authors of documents and the subscribers of the library both adhere; this taxonomy is just a set of keywords (or terms) structured in a hierarchy. An example of a taxonomy is the well known ACM Computing Classification System.<br>• A document is represented in the DL repository by a description of its content together with an identifier (e.g. the document's URI) allowing to access the document's content; the description is just a set of terms from the taxonomy.<br>• A query is just a set of terms from the DL taxonomy (i.e. a conjunctive query)<br>• A user is represented by an identifier together with a subscription; a subscription is just a query defining (intentionally) the documents of interest to the user. | 1<br>2<br>**3** |

**Table 20: Notification functionality compliance**

level of compliance (column) of the evaluated service and its features (1 = low, 2=medium, 3=high).

Here are some notes from the evaluator:

- **For FC1**: The complete implementation of this functionality needs the cooperation of many components of the ASSETS system. Tests clearly show that the preservation-notification component implements the publish-subscribe pattern, but we should need to test interactions with other components to check if the ASSETS implements the OAI preservation planning specifications.
- **For FC4**: In order to be compliant with this functionality, the implementation of the notification routing is missing. Currently, only asynchronous notifications can be sent.

### *5.1.3 Conclusions*

Conclusions are similar to those stated in 4.9.5**Errore. L'origine riferimento non è stata trovata.**.

As in 0 the functionality compliance table (Table 20) shows that the notification service has been considered to be compliant with most of the standards, models and similar EU projects selected for the evaluation of functionality compliance.

## 5.2 Content by re-use service

### *5.2.1 Introduction*

**Evaluation elaborated and completed by**: N. Aloia (CNR)

**Evaluation performed by**: Exalead,

**Date of evaluation performance:** March 26t$^{h}$ 2012.

The evaluation of this service using an API access is intended to developers who might want to add the above mentioned features to their digital libraries.

**Objectives**

The objectives of the evaluation are:

- To evaluate the adherence of the provided functionalities to the agreed service requirements,
- To evaluate the software freedom of failures caused by faults existing in the software itself,
- To evaluate the behaviour of the service during testing (such as the occurrence of unsatisfying operation).

**Use case**

Technical use case the "Mona Lisa Seminar Scenario".

A user of ASSETS decides to create a UGC object describing a seminar on Mona Lisa and to submit the newly created object to the EUROPEANA digital library.

**Scenario**

The scenarios we describe consist of portions of code to be used by developers who want to benefit from the services provided by the UGC back end. We describe three scenarios of increasing complexity

1. (Low complexity) ASSETS server uploading of a media file from the user's workstation and definition of some metadata values.
2. (Medium Complexity) Adding some metadata values to an object found in Europeana, using a query.

(High Complexity) Creating a complex object, combining an Europeana object and a user uploaded object, providing appropriate metadata values.

**Data for testing**

A generic file chosen by the user, containing a "fake" seminar program (e.g. the supplied MonaLisaSeminarProgram.pdf file).

**Method**

This service is evaluated according to ISO 9126 evaluation framework (3.4.4) with the aid of remote tests. Users have a scenario which explains a actions to be taken (e.g. steps to followed) and the

expected system beahviour (in terms of output to obtain); in order to report their experience, evaluators have to fill in matrices/tables with data related to the service that they have evaluated.

**Users**

Users involved in the tests have a technical role (e.g. developers / application architects).

This service is targeted to:

- developers who want to add the above mentioned features to Europeana.
- content providers who want to understand the behaviour of the deployed service.

### *5.2.2 Service presentation*

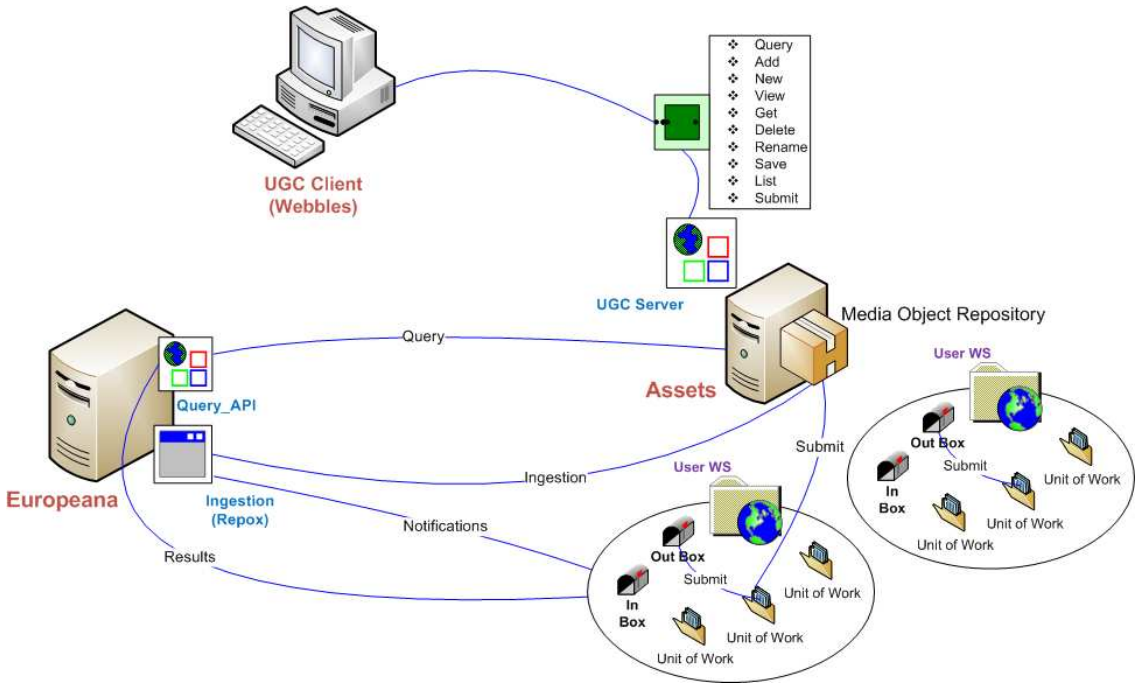



**Figure 43: UGC Architecture**

The User Generated Content (UGC) service is aimed at facilitating the information exchange between the users and the digital library: users find in ASSETS and Europeana the sources of information needed for their work and, in turn, they enrich the library with the results of their own work. In particular, the service allow ASSETS users to contribute to the contents of the digital library in several different ways, such as uploading simple media objects along with their descriptions, annotating existing objects, enriching existing descriptions or creating new complex content by extracting and recombining in various forms existing or new digital objects or parts thereof. Rather than focusing on a specific set of UGC applications, ASSETS has developed a general purpose, back end component that aims at supporting any UGC service Europeana will want to offer to its users. To this end, the ASSETS back end component implements an Application Programming Interface (API) for creating, storing and manipulating UGC instances, and for submitting these to Europeana, in the form of Europeana Submission Information Packages (SIPs) (Figure 43Three demonstrators have been

implemented to verify the functionality of the UGC services, The first one is a simple HTML page with examples of use of the API provided by the UGC service. This demo is useful for developers who want to implement GUI using HTML and is available at http://assetstest.atc.gr/assets/communityservices-ugc/. The second consists of a GUI that uses the "webble" technology developed at the MEME Media Lab and is available at http://www.meme.hokudai.ac.jp/WebbleWorld/WebbleWorldIndex.html. The third demonstrator is for java developers and is available for distribution on the ASSETS server.

### *5.2.3 Results*

Functional adequacy

| **SCENARIO STEPS**<br><br>**Suitability** *Functional adequacy*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | ScenarioStep1 | ScenarioStep2 | ScenarioStep3 |
|---|---|---|---|---|
| Upload a media file | FA 1 | X | | X |
| Create a new object with metadata. | FA 2 | X | | X |
| Query Europeana | FA 3 | | X | X |
| Select an object from the query result | FA 4 | | X | X |
| Change a metadata value in the selected object. | FA 5 | | | |
| Add one metadata field to the selected object | FA 6 | | X | X |
| Add an annotation/tag to a selected object | FA 7 | | X | X |
| Create a new complex object | FA 8 | | | X |
| Add an uploaded object as part of the new created complex object | FA 9 | | | X |
| Add the object selected in the result as part of the new created complex object | FA 10 | | | X |
| Submit the Unit of Work containing the created complex object. | FA 11 | | | X |

**Table 21: UGC functional adequacy**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A= Number of functions non offered by the evaluated service: A=2*

*B= Number of functions evaluated: B=18*

So **X=0,89**

Functional implementation completeness

| **SCENARIO STEPS**<br><br>**Suitability** *Functional implementation completeness*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONAL REQUIREMENTS** | | ScenarioStep1 | ScenarioStep2 | ScenarioStep3 |
|---|---|---|---|---|
| Support information exchange and enrich digital libraries through the involvement of the user community. | FIC 1 | X | X | X |
| **Composition of new content** by extracting and recombining existing digital objects | FIC 2 | | | X |
| Create a new object possibly with associated media files | FIC 3 | X | | X |
| Create a new object part of, or including existing Europeana objects | FIC 4 | | | X (including) |
| Add a new description to an existing object | FIC 5 | | X | X |
| Enrich the description of an existing object by adding a new metadata field | FIC 6 | | X | X |
| Add a new value to an existing metadata field | FIC 7 | | X | X |
| Annotate/Tag an existing object | FIC 8 | | X | X (UserTag) |

**Table 22: UGC functional implementation completeness**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A = Number of missing functions detected in evaluation: A=0*

*B = Number of functions described in requirement specifications: B=15*

So **X=1**

Function understandability

| **SCENARIO STEPS**<br><br>**Usability** *Function understandability*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | ScenarioStep1 | ScenarioStep2 | ScenarioStep3 |
|---|---|---|---|---|
| Upload a media file | FA 1 | X | | X |
| Create a new object with metadata. | FA 2 | X | | X |
| Query Europeana | FA 3 | | X | X |
| Select an object from the query result | FA 4 | | X | X |
| Change a metadata value in the selected object. | FA 5 | | X | X |
| Add one metadata field to the selected object | FA 6 | | X | X |
| Add an annotation/tag to a selected object | FA 7 | | X | X |
| Create a new complex object | FA 8 | | | X |
| Add an uploaded object as part of the new created complex object | FA 9 | | | X |
| Add the object selected in the result as part of the new created complex object | FA 10 | | | X |
| Submit the Unit of Work containing the created complex object. | FA 11 | | | X |

**Table 23: UBC function understandability**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of functions whose purpose is correctly understood by the user: A=18*

*B= number of available functions: B=18*

So **X=1**

Understandable input and output

| **SCENARIO STEPS**<br><br>**Usability** *Understandable input and output*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | ScenarioStep1 | ScenarioStep2 | ScenarioStep3 |
|---|---|---|---|---|
| Upload a media file | FA 1 | 1.0 | | 1.0 |

| | | | | |
|---|---|---|---|---|
| Create a new object with metadata. | FA 2 | 1.0 | | 1.0 |
| Query Europeana | FA 3 | | 0.5 | 0.5 |
| Select an object from the query result | FA 4 | | 1.0 | 1.0 |
| Change a metadata value in the selected object. | FA 5 | | 0 | 0 |
| Add one metadata field to the selected object | FA 6 | | 1.0 | 1.0 |
| Add an annotation/tag to a selected object | FA 7 | | 1.0 | 1.0 |
| Create a new complex object | FA 8 | | | 1.0 |
| Add an uploaded object as part of the new created complex object | FA 9 | | | 1.0 |
| Add the object selected in the result as part of the new created complex object | FA 10 | | | 1.0 |
| Submit the Unit of Work containing the created complex object. | FA 11 | | | 1.0 |

**Table 24: UGC understandable input output**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of input and output data items which user successfully understands: A=15*

*B= number of input and output data items available from the system: B=18*

So **X=0,83**

Maturity

| **SCENARIO STEPS**<br><br>**Maturity**<br>**Table 8.2.1 Maturity metrics** of the Norm ISO_IEC 9126-2<br><br>**MEASUREMENTS** | ScenarioStep1 | ScenarioStep2 | ScenarioStep3 |
|---|---|---|---|
| Presence of failures | | | |
| Number of tests performed | 1 | 1 | 1 |

**Table 25: UGC maturity**

no failure has occurred in any running test so **X=0**

Functionality compliance matrix

| | Feature | UGC service | Tester feedback |
|---|---|---|---|
| **FC1** | **Display** | The display functionalities are usually implemented in the presentation layer, the ASSETS UGC functionalities are invoked in case an interaction with the server is needed. | 3 |
| **FC2** | **Ease of annotation** | The UGC service provides and stores an annotated object in a single step and information objects are serialized in XML or JSON format, making it easy for a developer to design an ergonomic UI. | 3 |
| **FC3** | **Manage "Anonymity"** | In the ASSETS platform the control of the user credentials is performed by a specific component, the ASSETS UGC service interacts with this component for checking users permissions. | 3 |
| **FC4** | **Control of content**. | The Europeana ingestion workflow is a human supervised process. The user generated content submitted to Europeana using the ASSETS UGC service then will be checked before being accepted. The Assets UGC service interacts with the Europeana Ingestion tools and notify to the user the result of the submission (accepted/rejected). | 3 |
| **FC5** | **Harvesting user generated content** | Annotation, tags, metadata fields, added or updated by users will be considered part of the Object description and can be searched or accessed as any other Europeana content. | 3 |
| **FC6** | **Ease of retrieval** | A specific data space (WS) is associated to every user by ASSETS UGC service, all the content created by a user is stored in her/his WS and the UGC provide simple functionalities to retrieve and manage it. The content created by a user is stored in his/her WS even if it has been accepted by Europeana ingestion tools. | 3 |
| **FC7** | **Traffic and network effects** | One of the declared goal of Europeana is to enhance users contribution to the digital library. The Europeana community is more and more growing and we expect that a significant number of members of this community will use the ASSETS UGC service. | 3 |
| **FC8** | **Notification and sharing** | The UGC service interacts with the ASSETS Notification module to share user generated content within the ASSETS community. However the functionalities provided by UGC can be easily used by developers to create applications for sharing the data also with other Web 2.0 communities. | 2 (not sure about how easy data sharing with other communities would be) |

**Table 26: UGC functionality compliance**

### *5.2.4 Conclusions*

The functional testing of the Content by re-use service has arisen no issues according to the predefined criteria and to the tested scenario:

- Nearly all the evaluated functions are suitable to perform the specific tasks (x=0,89),

- There were not missing functions to perform the specific tasks,
- Users understand easily the purpose of each functions,
- Users understand nearly all what is required as input and what is provided as output at each step of the tests (x=0,83),
- The maturity metrics (belonging to reliability characteristic) is very good because there was no failure caused by faults existing in the software itself.

An external functionality compliance table (Table 20) has also been filled. It shows that the Content by re-use service has been considered to be highly compliant with most of features.

## 5.3 Query logs service

### *5.3.1 Introduction*

The evaluation of Query logs service using an API is intended to developers in charge of analysis of the query logs.

**Evaluation elaborated and completed by**: Diego Ceccarelli, Claudio Lucchese (CNR),

**Evaluation performed by**: CNR (other team)

**Date of evaluation performance:** April 6th 2012.

**Objectives**

The objectives of the evaluation are:

- To evaluate the adherence of the provided functionalities to the agreed service requirements.
- To evaluate the software freedom of failures caused by faults existing in the software itself
- To evaluate the behavior of the service during testing (such as the occurrence of unsatisfying operation)

**Use case**

The corresponding use case is the technical use case the "Query Log Exploration".

A power-user of ASSETS decides to analyze the query log to retrieve some statistics about the portal usage.

**Scenario**

The proposed scenario has the following goals:

1. Collect some general statistics;
2. Find the country of a specific user;

The following tasks have to be executed to complete the scenario:

1. Find the most frequent queries in submitted during Dec. 2011;
2. Find the average query length of the queries submitted during Dec. 2011;
3. Find the number of distinct queries submitted during Dec. 2011;
4. Find the number of distinct user sessions occurred during Dec. 2011;
5. Find the number of distinct queries submitted during Dec. 2011;
6. Find the country from which the user `'553f3e158f3108cfac092e4ec679cd96'` submitted his queries;

**Data for testing:** Query logs

**Method**

This service is evaluated according to ISO 9126 evaluation framework (3.4.4) with the aid of remote tests. Users have a scenario which explains a actions to be taken (e.g. steps to followed) and the

expected system beahviour (in terms of output to obtain); in order to report their experience, evaluators have to fill in matrices/tables with data related to the service that they have evaluated.

**Users**

Users involved in the tests are developers.

This service is targeted to:

- developers for the analysis of the query logs;
- decision makers for better targeting the ASSETS users;

### *5.3.2 Service presentation*

A query log keeps track of historical information regarding past interactions between users and the retrieval system. It usually contains tuples $\langle q_i, u_i, t_i, V_i, C_i \rangle$ where for each submitted query $q_i$, the following information is available: i) the anonymized identifier of the user $u_i$, ii) the submission timestamp $t_i$, iii) the set $V_i$ of documents returned by the search engine, and iv) the set $C_i$ of documents clicked by $u_i$. Therefore, a query log records both the activities conducted by users, e.g. the submitted queries, and an implicit feedback on the quality of the retrieval system, e.g. the clicks.

The goal of this service is to devise a set of query log processing tools needed by other services, in particular for extracting user behavioural patterns needed for improving the ASSETS ranking function and for providing the model used by the query recommendation service. This includes non-trivial activities such as query log cleaning, analysis and indexing.

### *5.3.3 Results*

*Functional adequacy*

| **SCENARIO STEPS**<br><br>**Suitability** *Functional adequacy*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Find the most frequent queries in submitted during Dec. 2011; | Find the average query length of the queries submitted during Dec. 2011; | Find the number of distinct queries submitted during Dec. 2011; | Find the number of distinct user sessions occurred during Dec. 2011; | Find the number of distinct users in Dec. 2011; | Find the country from which the user '553f3e158f3108cfac092e4ec679cd96' submitted his queries; |
|---|---|---|---|---|---|---|---|
| Find the most frequent queries in submitted in a given month | FA 1 | X | | | | | |
| Find the average query length of the queries submitted in a given month | FA 2 | | X | | | | |
| Find the number of distinct queries submitted in a given month | FA 3 | | | X | | | |
| Find the number of distinct user sessions | FA 4 | | | | X | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| occurred in a given month | | | | | | | |
| Find the number of distinct users in a given month | FA 5 | | | | | X | |
| Find the country from which a given user submitted his queries | FA 6 | | | | | | X |

**Table 27: Query log service Functional adequacy**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A= Number of functions in which problems are detected in evaluation: A=0*

*B= Number of functions evaluated: B=6*

So **X=1**

*Functional implementation completeness*

| **SCENARIO STEPS**<br><br>**Suitability** *Functional implementation completeness*<br>**Table 8.1.1 Suitability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONAL REQUIREMENTS** | | Find the most frequent queries in submitted during Dec. 2011; | Find the average query length of the queries submitted during Dec. 2011; | Find the number of distinct queries submitted during Dec. 2011; | Find the number of distinct user sessions occurred during Dec. 2011; | Find the number of distinct users in Dec. 2011; | Find the country from which the user '553f3e158f3108cfac092e4ec679cd96' submitted his queries; |
|---|---|---|---|---|---|---|---|
| The service provides information about the queries | FIC 1 | X | X | X | | | X |
| The service provides information about sessions | FIC 2 | | | | X | | X |
| The service provides information about the users | FIC 3 | | | | | X | X |

**Table 28: Query log service Functional implementation completeness**

Measurement, formula and data element computations: **X = 1 – (A / B)** where:

*A = Number of missing functions detected in evaluation: A=0*

*B = Number of functions described in requirement specifications: B=6*

So **X=1**

*Function understandability*

| **SCENARIO STEPS**<br><br>**Usability** *Function understandability*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Find the most frequent queries in submitted during Dec. 2011; | Find the average query length of the queries submitted during Dec. 2011; | Find the number of distinct queries submitted during Dec. 2011; | Find the number of distinct user sessions occurred during Dec. 2011; | Find the number of distinct users in Dec. 2011; | Find the country from which the user '553f3e158f3108cfac092e4ec679cd96' submitted his queries; |
|---|---|---|---|---|---|---|---|
| Find the most frequent queries in submitted in a given month | FA 1 | X | | | | | |
| Find the average query length of the queries submitted in a given month | FA 2 | | X | | | | |
| Find the number of distinct queries submitted in a given month | FA 3 | | | X | | | |
| Find the number of distinct user sessions occurred in a given month | FA 4 | | | | X | | |
| Find the number of distinct users in a given month | FA 5 | | | | | X | |
| Find the country from which a given user submitted his queries | FA 6 | | | | | | X |

**Table 29: Query log service Function understandability**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of functions whose purpose is correctly understood by the user: A=6*

*B= number of available functions: B=6*

So **X=1**

*Understandable input and output*

| **SCENARIO STEPS**<br><br>**Usability** *Understandable input and output y*<br>**Table 8.3.1 Understandability metrics** of the Norm ISO_IEC 9126-2<br><br>**FUNCTIONALITY** | | Find the most frequent queries in submitted during Dec. 2011; | Find the average query length of the queries submitted during Dec. 2011; | Find the number of distinct queries submitted during Dec. 2011; | Find the number of distinct user sessions occurred during Dec. 2011; | Find the number of distinct users in Dec. 2011; | Find the country from which the user '553f3e158f3108cfac092e4ec679cd96' submitted his queries; |
|---|---|---|---|---|---|---|---|
| Find the most frequent queries in submitted in a given month | FA 1 | X | | | | | |
| Find the average query length of the queries submitted in a given month | FA 2 | | X | | | | |
| Find the number of distinct queries submitted in a given month | FA 3 | | | X | | | |
| Find the number of distinct user sessions occurred in a given month | FA 4 | | | | X | | |
| Find the number of distinct users in a given month | FA 5 | | | | | X | |
| Find the country from which a given user submitted his queries | FA 6 | | | | | | X |

**Table 30: Query log service Function understandable input output**

Measurement, formula and data element computations: **X = A / B** where:

*A= number of input and output data items which user successfully understands*

*B= number of input and output data items available from the system*

*Here A=B*

So **X=1**

*Maturity*

| SCENARIO STEPS / Maturity / MEASUREMENTS | Find the most frequent queries in submitted during Dec. 2011; | Find the average query length of the queries submitted during Dec. 2011; | Find the number of distinct queries submitted during Dec. 2011; | Find the number of distinct user sessions occurred during Dec. 2011; | Find the number of distinct users in Dec. 2011; | Find the country from which the user '553f3e158f3108cfac092e4ec679cd96' submitted his queries; |
|---|---|---|---|---|---|---|
| Presence of failures | | | | | | |
| Number of tests performed | 3 | 3 | 3 | 3 | 3 | 3 |

**Maturity**

**Table 8.2.1 Maturity metrics** of the Norm ISO_IEC 9126-2

**Table 31: Query log service Maturity**

Measurement, formula and data element computations: **X= A1 / A2** where:

A1 = number of detected failures: A1=0

A2 = number of performed test cases: A2=18

So **X=0**

### *5.3.4 Conclusions*

The functional testing of the Query log service has arisen no issues according to the predefined criteria and to the tested scenario:

- All the evaluated functions are suitable to perform the specific tasks (x=1),
- There were not missing functions to perform the specific tasks (x=1),
- Users understand easily the purpose of each functions (x=1),
- Users understand all what is required as input and what is provided as output at each step of the tests (x=1),
- The maturity metrics (belonging to reliability characteristic) is very good because there was no failure caused by faults existing in the software itself (x=0).

# 6. Conclusions

As an output of the end-user evaluation, a detailed evaluation report was produced, where all the usability issues and recommendations were reported with a severity ranking. This report has been the input for the ASSETS developers team, who after having analyzed the list of issues, have assigned them a priority based on three factors: the severity of the issue, the estimated cost to fix it and the relevance of the issue for the ASSETS project. Finally, the usability issues have been addressed by the team in priority order, starting with the High Priority issues and then as many of the Medium and Low Priority issues as time and resources have allowed. As result, the final version of the ASSETS portal has been released.

The evaluation of the professional services performed by professional end-user such as content providers, developers and programmers, was executed in two ways: evaluation of the service using their GUIs or using their APIs. The evaluation methods have been chosen according to the access and to the status of the service.

Each service evaluated with *user test* or *cognitive walkthrough* method provide a) a list of recommendations listed in three levels of priority: high, medium, and low for consideration, b) a detailed analysis of the data from questionnaires and comments made by the users when performing the task and c) a usability report based on discussing the effectiveness, efficiency and satisfaction of the service. Services evaluated with *user test* aiming to measures some IS0 9126 metrics provide an evaluation of the quality according to suitability, functionality compliance, maturity and understandability characteristics.

The final report will be the input for the developers of these services in general and specifically it will guide the services developed in ASSETS that will be selected for integration in the Europeana portal.

This deliverable finally provides a detailed and specified feedback as well as guidelines for improvements for professional services and a final description of the final UI for the Assets portal.

# 7. Bibliography

[Bastien 91] « Validation de critères ergonomiques pour l'évaluation d'interfaces utilisateurs », INRIA, rapport de recherche n1427, Programme 3, mai 1991

[Bevan 97] Nigel Bevan, van Veenendaal, E, and McMullan, Quality and usability: A new framework, J (eds) Achieving software product quality, Tutein Nolthenius, Netherlands, 1997

[Bevan 99] Quality in use for all, User interfaces for all. Stephanidis, C (ed), Lawrence Erlbaum, 1999.

[Coutaz and all, 94] J. Coutaz - S. Balbo : Evaluation des interfaces utilisateurs : taxonomie et recommandations, IHM 94, Lille, Déc 1994

[Jeffries 91] R. Jeffries, J. R. Miller, C. Wharton, K. M. Uyeda, « User interface evaluation in the real world: a comparison of four techniques », Proceedings of CHI'91, New Orleans, ACM New York, pp119-124, may 1991.

[Mahatody, 10] Thomas Mahatody, Mouldi Sagar, Christophe Kolski: State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions. Int. J. Hum. Comput. Interaction 26(8): 741-785 (2010)

[Nielsen, 93] Nielsen J. Usability Engineering. Academic Press, Boston, 1993.

[Nielsen and Mack, 1994] Nielsen, J and Mack R. L. Usability Inspection Methods. Wiley, New York, 1994.

[Preece, Rogers and Sharp, 2002] Preece, J., Rogers, Y., and Sharpe, H. Interaction Design. Beyond human-computer interactions. Wiley, New York, 2002.

[Senach 90] B. Senach, "Evaluation ergonomique des interfaces home-machine: une revue de la literature", Rapport de recherché INRIA n1180, Programme 8, Communication home-machine, mars 1990

# 8. Appendixes

## 8.1 Appendix 1: Evaluation instruments for ISO 9126

The evaluation instruments for the users in charge to perform the evaluation (e.g. developers / content providers) represent a set of tools that have been used during the service "walkthrough" of the test scenario.

These consists of:

- Suitability-Functional adequacy Matrix;
- Suitability- Functional implementation completeness Matrix;
- Understandability - Function understandability Matrix;
- Understandability - Understandable input and output Matrix.
- Maturity Matrix;
- Functionality Compliance Matrix;
- Suggestion / Feedback Table.

An example of these matrices can be found in **Errore. L'origine riferimento non è stata trovata.** Appendix 2, where the matrices for the evaluation of the preservation notification service are presented.

**How to use the evaluation instruments**

Suitability-Functional adequacy Matrix

WHEN: This tool shall be used at the end of each step of the scenario.

WHAT TO DO: For each step (column), a cross (X) would need to be placed in the white box whenever the corresponding functionality (row) is offered by the being evaluated service.

Suitability- Functional implementation completeness Matrix

WHEN: This tool shall be used at the end of each step of the scenario.

WHAT TO DO: For each step (column), a cross (X) would need to be placed in the white box whenever the corresponding functional requirement (row) is covered by the functionality being evaluated.

Understandability - Function understandability Matrix;

WHEN: This tool shall be used at the end of each step of the scenario.

WHAT TO DO: For each step (column), a cross (X) would need to be placed in the white box whenever the corresponding functionality's objective (row) offered by the being evaluated service has been understood by the evaluator.

Understandability - Understandable input and output Matrix.

WHEN: This tool shall be used at the end of each step of the scenario.

WHAT TO DO: For each step (column), a number would need to be placed in the white box which represents the number of input and output data items (e.g. parameters, variables correctly valued) which user has successfully understood.

Maturity Matrix

WHEN: This tool shall be used at the end of the scenario.

WHAT TO DO: For each step (column),

- a cross (X) would need to be placed in the white box of the first row ("presence of failures") whenever a failure/error is encountered,
- a "1" would need to be placed in the white box of the second row ("number of tests performed") whenever the step is executed and ended with no failures,

For example, after one test carried forward with no failures, the table should look like the following:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Presence of failures | | | | | | | | |
| Number of tests performed | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Whilst after three tests, last of which ended with a failure at step 2, the table should look like the following:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Presence of failures | | X | | | | | | |
| Number of tests performed | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |

Functionality Compliance Matrix

WHEN: This tool shall be used at the end of the scenario.

WHAT TO DO: For each applicable regulations, standards and conventions listed (rows), please provide a level of compliance (column) of the evaluated service and its features (1 = low, 2=medium, 3=high).

Suggestion / Feedback Table

WHEN: This tool shall be used when either a failure/error/unexpected behaviour or a "non-matching" is encountered by the evaluator

WHAT TO DO: For each failure/error/unexpected behaviour or "non-matching (row):

- fill in the first table column with the name of the matrix where a failure/error/unexpected behaviour or a "non-matching" has been encountered
- fill in the second table column with the row and the column corresponding to the
- fill in the third table column with free-text notes

For example, after a "non matching" encountered while filling the **Suitability** *Functional adequacy* matrix, step 3 in the scenario, FA4, a row in the table should look like the following:

| | | | |
|---|---|---|---|
| **Suitability** *Functional adequacy* matrix | FA4 | Step3 | <Notes from the evaluator> |

## 8.2 Appendix 2: MAT stand alone interface evaluation

### *8.2.1 Reporting of cognitive walkthrough results*

The scenario objective is to view a given video about Metz battle and its annotation, to complete this annotation finding the name of the military chief introduced in this video.

***e) Task 1: select a document***

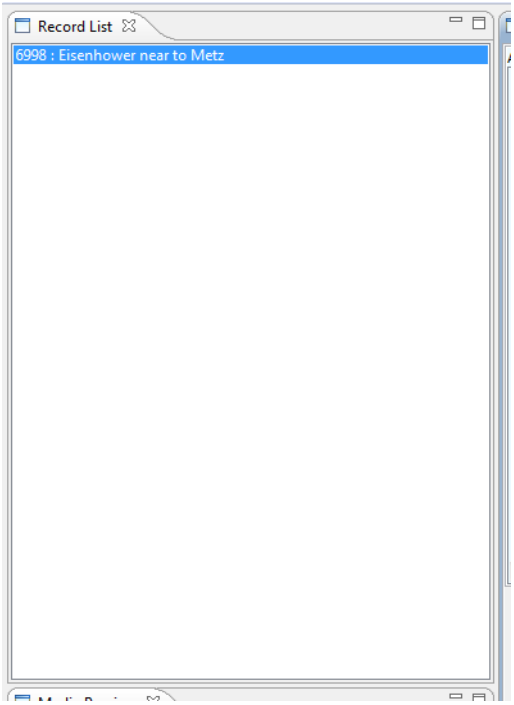



Fig2 : document selection

**Task description:** Select a document in a list of documents

**Right execution**: Click on the document to load the video and annotation

1. **Does the user know what to do to achieve the task?**

   Yes, the selection is a classical task for the users and the interface provides no innovation for this task.

2. **Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?**

   Yes.

3. **Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?**

   yes

4. **Does the user understand the results given by the system? Does the result quality match the user expectation?**

   yes

5. **Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?**

Two users found that it is not easy to modify the size and position of the panel with eclipse. The third one is more accustomed to this kind of manipulation. Nevertheless all the users thought that this panel should be smaller and even perhaps not always displayed.

**Additional user remarks**

Actually there are not a lot of functions but perhaps more functionalities should be necessary to interact with the ASSETS database. For example, how can we select a list?

### *f) Task 2: Read and edit the annotation, test to make one correction*

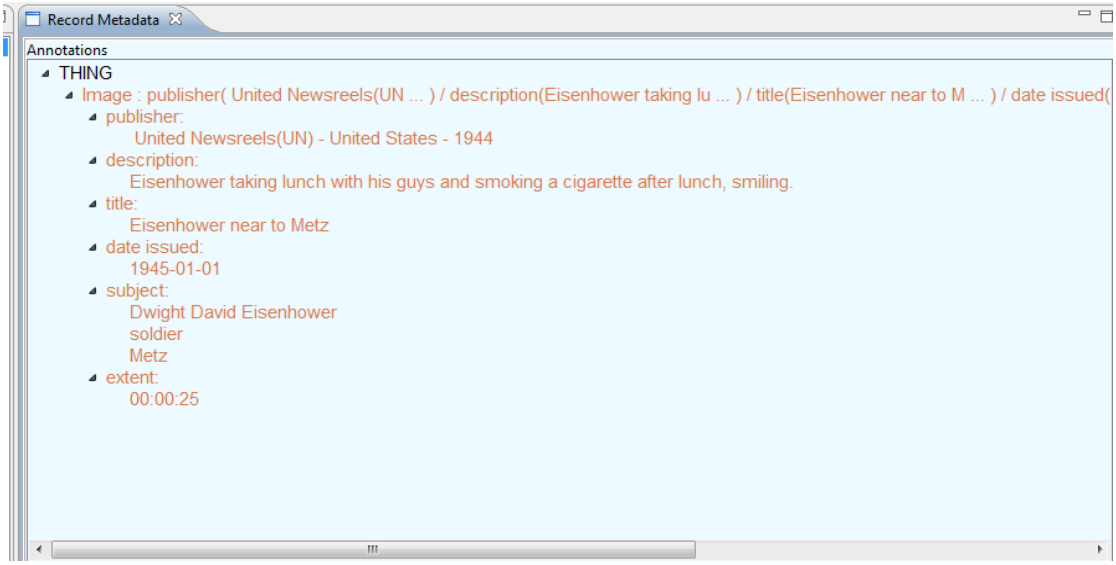



Fig 3: Annotation display

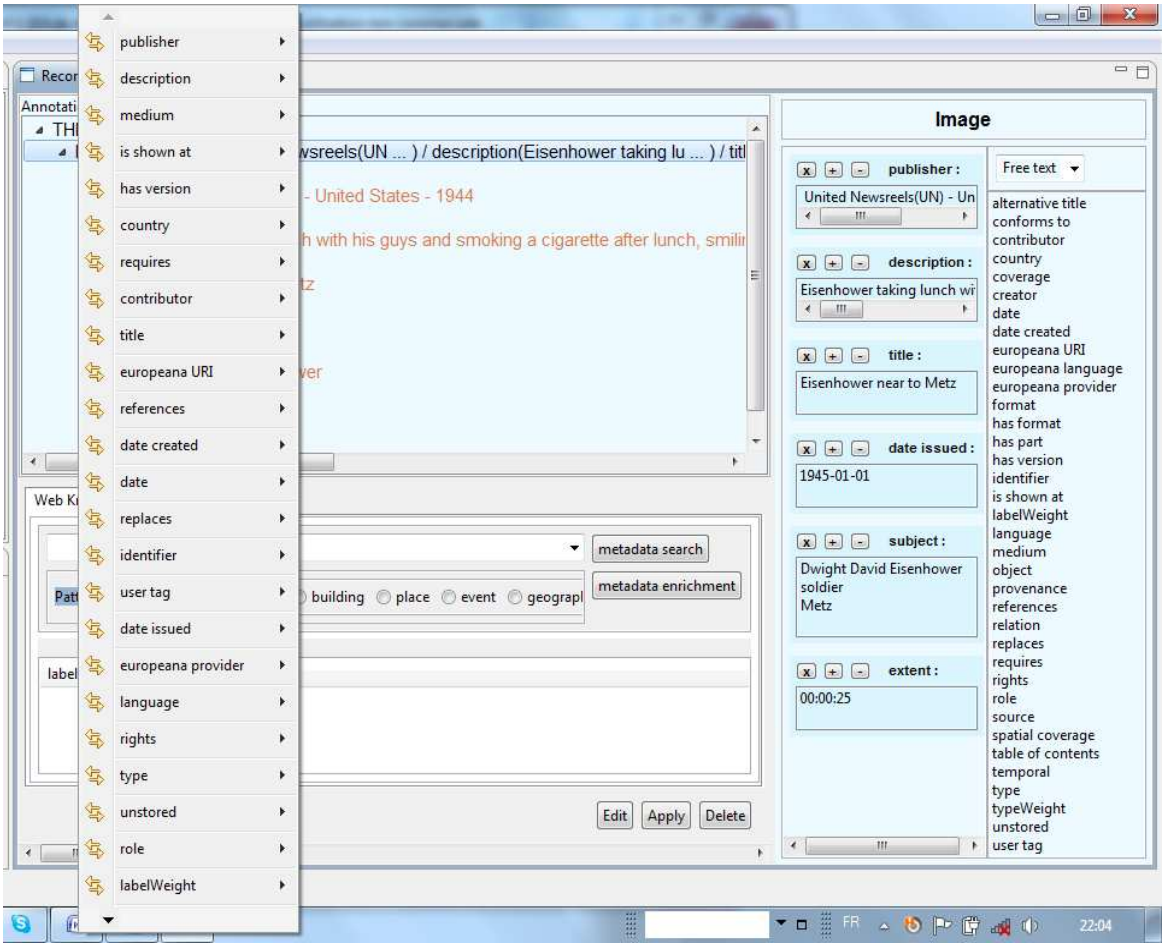



Fig 4: Click on image field provides edition menu

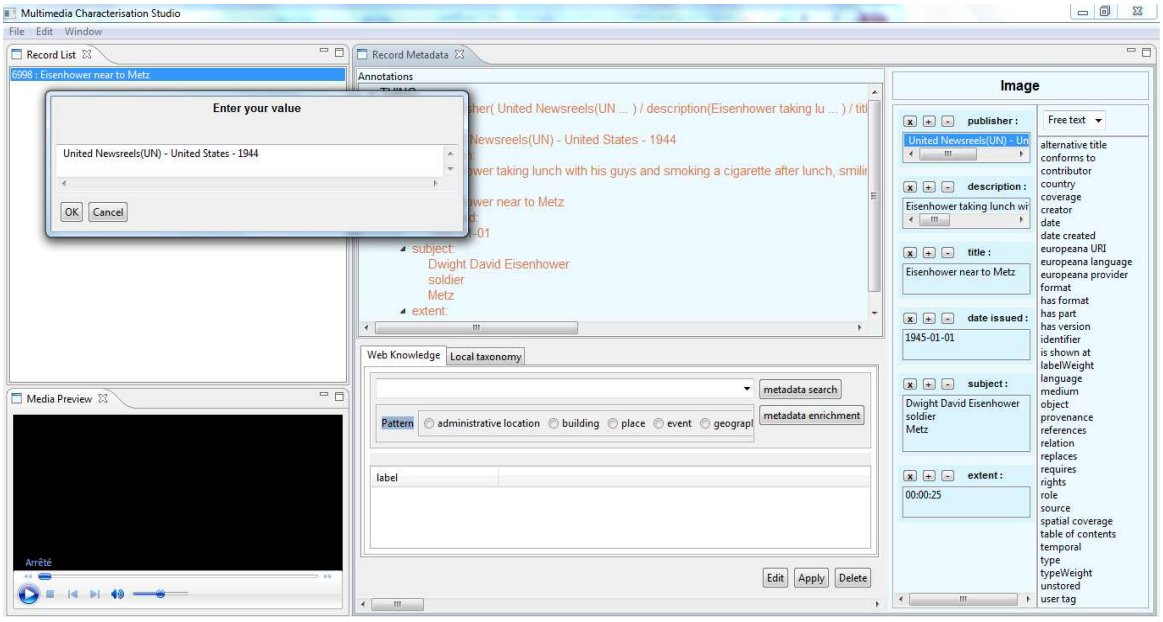



Fig 5: Click on publisher field at the right provides a menu to correct this field

**Task description:** Read and edit the annotation. Test to make one correction.

**Right execution**: correction of the description field

1. **Does the user know what to do to achieve the task?**

   Yes, theoretically no difficulty with the task objectives.

2. **Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?**
   - Not easy to find the edit button
   - Not easy to combine the 3 menus before to change a word in the publisher field.
   - One pop-up menu not useful when clicking on the *image* field.

3. **Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?**

   The interaction to modify notice is not adapted to the user experience and need too much manipulation to realize one action:
   - It is difficult to suppress a piece of information in a notice.
   - There is no confirmation needed to suppress.
   - In the editing menu at the right, why some filed could be modified and other should be displayed in another box to be modified? What is dedicated for the list at the right in the *image* menu?

4. **Does the user understand the results given by the system? Does the result quality match the user expectation?**

   Yes, the users understand the results but they prefer their traditional interface to edit and modify a notice.

5. **Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?**

   Some design problems have been noticed:
   - the notice is not easily readable because there is no distinction between the name and the values of each field
   - police size should be smaller
   - The first line of the annotation (very long) is not useful because it is duplicated just after.
   - In the edition mode, the proposed fields couldn't be displayed by alphabetical order.

**Additional user remarks**

Pertinent functionalities are present (possibility to obtain all the ASSETS allowed field with a mouse click) but too much interactions to make one modification in the annotation.

Furthermore, when users are traditionally doing annotation, they like to work on a set of documents at the same time, saving them temporally and the importing them later in the database. How this will be possible with ASSETS?

### *g) Task 3: View the video and find some entity names missing in the annotation*


Fig 6 : video display

**Task description:** View the video

**Right execution**: Play the video and find Patton name.

1. **Does the user know what to do to achieve the task?**
   Yes, users are accustomed to proceed like this.
2. **Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?**
   Yes.
3. **Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?**
   yes
4. **Does the user understand the results given by the system? Does the result quality match the user expectation?**
   yes
5. **Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?**
   yes

**Additional user remarks**

It is a classical media player with only few functionalities. For example there is no possibility to create extracts but this is presumable not useful to correct annotation of document already prepared.

### *h) Task 4: Find the entity names and their titles in the web*

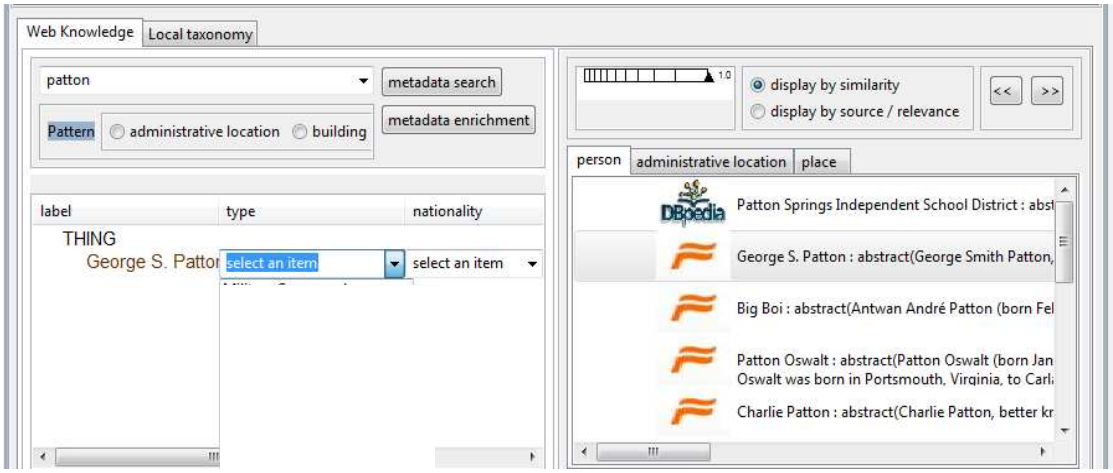



Fig 7 : Web knowledge display

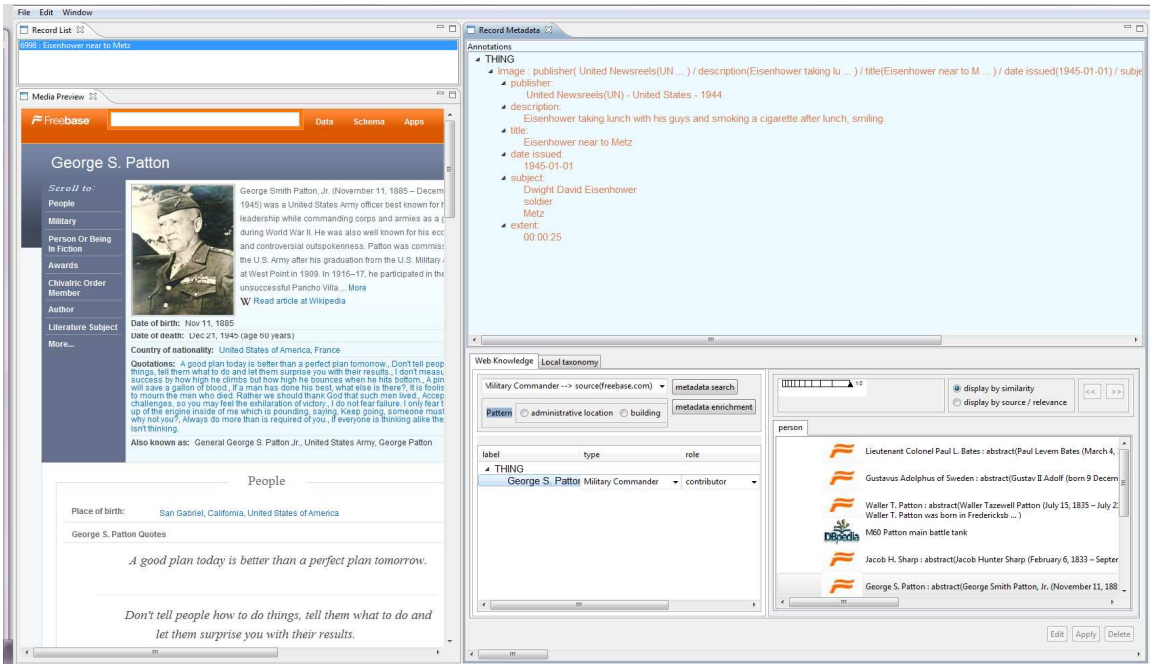



Fig 8 : Web display

**Task description:** Find the entity names and their titles in the web.

**Right execution**: enter "Patton" in the metadata search box, view the list of results, chose one, import it in the working box and verify the pertinence of the answer in the web display.

1. **Does the user know what to do to achieve the task?**

   Yes, metadata search is done in a classical box, then the list of result is easy to distinguish, easy to chose one item inside the list. There is also some pattern options to select for the metadata search.

   The innovation is given by the semantic tag returned by the web search. Also, the working box is less classical and the user had to discover how to use it but it does not present real difficulties to understand how to achieve a task.

2. **Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?**

   Some functionalities are not easy to find:

   - The user must know that he has to drag and drop one answer in the working area.
   - He also has to know that he could drag and drop the result in the media preview panel to visualize the web page.
   - There is only a shortcut (CTRL F) to search a term in the web page. Two users would prefer to have a menu.

   There are some bugs:

   - Text entered in the box is automatically changed and the information given by the user is lost.
   - Sometimes the user can't enter a new word in the search box

   Some users would like to have more functionalities:

   - To have the functionalities of classical web browser. The uses are embarrassed by the fact that they could not navigate in the web display as in a classical web browser (no back and forward function…).
   - To highlight searched term in the web page.
   - To rapidly display the corresponding web page when pointing the items in the results list.

3. **Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?**

   Even if the users have never work with semantic data, the achievement of results in the form of semantic data is easy to understand and manipulate.
   There is a positive feedback with the display of the page of the selected item. The two lines of summary are clear enough to understand the content of the correspondent web pages.
   One remark: users would prefer to have an automatic view of the item page when moving a mouse upon it.
   Some interaction and feedback problems:

   - The second action of drag and drop is more delicate to operate because the user has to find the exact place to copy. This interaction is really no easy to realize.
   - There is no information of time to wait to obtain all the results
   - When a web page is displayed the video display is lost.

4. **Does the user understand the results given by the system? Does the result quality match the user expectation?**

   Yes. The results given in the search results display are enough short and enough explicit to choice one item.

   For the pattern option in metadata search, users find that the categories are not homogeneous. They would prefer to have only "place" in spite of "place", "administrative location" and "geographic location" (because they are not independent). Users were also surprised when they obtain answers in non selected pattern. For example, they ask for Patton with the pattern person and 3 pattern tab were retrieved "person", "administrative location" and place (Fig 7).

5. **Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?**

   Some information is not easily readable:

   - Too much lines to browse:
     - Pattern list in the metadata search box has to be browsed (too long line)
     - Items in the results list are displayed in too long lines that need always browsing.

- For the similarity mode display in the results, the arrow is not enough visible.
- There is some place left in the working box
- Some users would like to test if it is better to move the working box under the results list display.
- The cursor is always displayed as active (with a circle)
- There is a non useful slider at the bottom of the interface.
- Names are missing to recognize the different panels.

**Additional user remarks**

Users ask for more tests to analyze de difference between display by similarity and display by relevance. This option seems to be interesting but need to be more deeply analyzed.

### *i) Task 5: Chose the pertinent field to characterize this person*

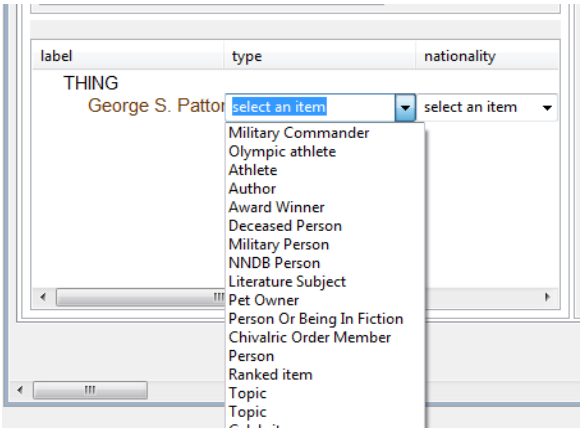



Fig 9 : field modification in the working box

**Task description:** Chose the pertinent field to characterize this person

**Right execution**: the correct label was the first one proposed in the list.

1. **Does the user know what to do to achieve the task?**

   Yes, there is a classical list box to select an answer.

2. **Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?**

   Yes , this functionality is easy to find and to understand.

   There was no problem for this example but two of the users find a problem in this box. It is no possible to select two fields for one entity. For example, it is no possible to select “military commander” and “deceased person” for Patton.

   Users notice that there are also some strange propositions in the list as “Person or being in fiction” for Patton. These problems are inherent in the way of description obtained from web knowledge bases. Indeed, these bases are supplied by crowd sourcing and therefore may contain errors. This is not a critical problem in our case because the user may reject the errors.

3. **Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?**

   yes

4. **Does the user understand the results given by the system? Does the result quality match the user expectation?**

Yes., except the impossibility to have a multi-selection (cf 2)

5. **Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?**

   yes

Some bugs have been noticed in the working area:

- It is not always possible to suppress a field
- When a field has been displaced, the remove action suppresses the previous field which was positioned in the place in spite of the new one.

### *j) Task 6: Import these descriptions in the annotation*

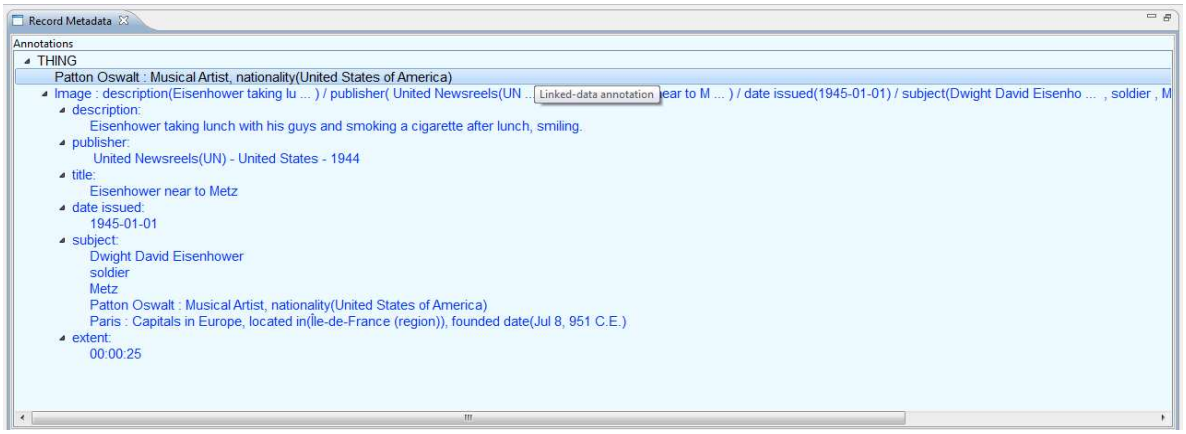



Fig 10 : importation of the information in the annotation

**Task description:** importation of the information in the annotation

**Right execution**: The annotation has been completed.

1. **Does the user know what to do to achieve the task?**

   As there are different possibilities (automatic metadata enrichment or drag and drop) all the users could find a way to do it.

2. **Are the proposed functionalities easy to find, easy to combine? Are they compliant with user desiderata?**

   Yes , this functionality is easy to find and to understand.

   Nevertheless the users find that it was easy to import a new line including all the information of the working box but not easy to create different lines containing the information.

3. **Is user system interaction adapted to user expectations and previous experiences? Is there a positive guide and feedback for the user?**

   yes

4. **Does the user understand the results given by the system? Does the result quality match the user expectation?**

   The user was sometimes a little surprised by the line dispatching in the notice but he said that he could not conclude anything about the results dispatching in the notice because Ina format is very different from ASSETS format and he has no knowledge about the ASSETS fields. Moreover this problem does not belong to the MAT tool but to the Europeana description.

5. **Is the visual design of the interface relevant? Are the different elements of information easily readable and distinguishable?**

yes

**Additional user remarks**

Some users also notice that it is not easy to describe all person role with ASSETS typology (only creator, publisher and contributor fields) but this problem doesn't belong to this evaluation goal.

Bug: The imported line has been placed at the top of the notice up to Image label.

### *8.2.2 Additional user manipulation*

Some users ask for additional tests. They were interested by the link to semantic database and they would understand the future possibilities offered by semantic tagging. Firstly they wanted to analyse more deeply the difference between Google and these databases to search named entity. They try to search the commander's name of Metz battle with MAT and with Google interface.

1. with Google:
   - search for *Metz* gives too much noise
   - search for *Metz battle* provides the Wikipedia page at the beginning of the results. Commanders names are listed in the right box of the page.
2. with MAT
   - search for *Metz* with pattern *event* provides the freebase page of Battle of Metz. The user drags and drops the line in the working area and then click on the Battle name to add the commander's field.

So information are easy to find with MAT (easier or as easy as with Google) but need too much interactions between the panels.

The users were interested to do another test. They are currently confronted to orthographic problem, when they don't know the name writing, when they would find the more common writing of a name, and to verify that it is the right person. They are used to exploit the suggestion mode of Google for this task. They tried to enter *Miterand* in spite of *Mitterrand* with *person* pattern and MAT finds *Mitterrand* in the list of results. So, the users were disappointed to not have the suggestion mode but they were encouraged because they could nevertheless find the correct writing.

### *8.2.3 Compilation of questionnaires*

**This compilation gathers the results of the 3 users.**

## ASSETS evaluation
## Manual annotation & correction task
## QUESTIONNAIRE

1. Entry questionnaire: this will provide us with background information that will help us analyse the answers you give in end of this experiment.

3. Exit questionnaire: this will provide us with your assessment of your overall annotation experience with the ASSETS system. Take into account that we are interested in knowing your opinion: answer questions freely, and consider there are no right or wrong answers. Please remember that we are evaluating the system you have just seen and not you.

# 1. ENTRY QUESTIONNAIRE

## 1.1 Personal details

**1. What is your gender?**

1 Male
2 Female

**2. What is your relationship to the ASSETS project?**

☐ I am an ASSETS partner
☐ I collaborate with an ASSETS partner
3 I have no connection

## 1.2 Annotation experience

**3. Are you involved in annotation activities?**

☐ Learning period of annotation task
3 Practice annotation task
☐ Developing tool for archivist involved in annotation task

**4. Since how long time do you practice annotation task?**

☐ Less than one year
1 1-5 years
☐ 5-10 years
2 more than 10 years

**5. For annotation task, do you have often to find or verify information with external tool not incorporate in your annotation tool?**

☐ Never
☐ Once or twice a year
☐ Once or twice a month
☐ Once or twice a week
1 Once or twice a day
2 More often

**6. Which system do you use currently use to find information for annotation task?**

3 Search engine (Google,..)
3 Encyclopedia (Wikipedia …)
3 Thesaurus
☐ Ontology
☐ Semantic database
☐ Other …….

**7. Which system do you use currently use to find information for other task (personnal,…)?**

3 Search engine (Google,..)
3 Encyclopedia (Wikipedia …)
1 Thesaurus
1 Ontology
☐ Semantic database
☐ Other …….

**EXIT QUESTIONNAIRE**

## 1 ASSETS annotation functionalities

**1. Does this interface allow the user to find information needed to complete / correct an annotation?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| no | ☐ | ☐ | 1 | 1 | ☐ | yes |

**2. How do you find the tools offered in the interface to complete / correct an annotation?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 2 | 1 | useful |
| difficult to use | ☐ | 1 | 2 | ☐ | ☐ | easy to use |
| ineffective | ☐ | ☐ | 2 | 1 | ☐ | effective |

**3. How do you find the system of enrichment proposition?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | 1 | 1 | 1 | ☐ | easy to use |
| ineffective | ☐ | ☐ | 1 | ☐ | 1 | effective |

**4. How do you find the semantic field?** (easier to get information that is already in a structure in the form of semantic fields rather than reading a page to find relevant information?)

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | ☐ | 3 | useful |
| difficult to use | ☐ | ☐ | 1 | 1 | 1 | easy to use |
| ineffective | ☐ | ☐ | ☐ | 2 | ☐ | Effective |

**5. How do you like having all the tools together in one interface?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | 1 | ☐ | 1 | 1 | useful |
| difficult to use | ☐ | 1 | ☐ | 2 | ☐ | easy to use |
| ineffective | ☐ | ☐ | ☐ | 2 | 1 | effective |

## 2 Record metadata window

**6. How do you like annotation display?** (visual presentation)

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | 1 | 1 | ☐ | 1 | easy to use |
| ineffective | ☐ | ☐ | 1 | ☐ | 1 | effective |

**7. What do you think of drag / paste, from the working box, to change a field?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | ☐ | 1 | 1 | 1 | easy to use |
| ineffective | ☐ | ☐ | 2 | 1 | ☐ | effective |

**8. What do you think of the action button "metadata enrichment" to change a field?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | ☐ | 3 | useful |
| difficult to use | ☐ | ☐ | 1 | 1 | 1 | easy to use |
| ineffective | ☐ | ☐ | 1 | 1 | 1 | effective |

## 3 Web knowledge window

**9. What do you think of the feature "search metadata"?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | ☐ | 2 | ☐ | 1 | easy to use |
| ineffective | ☐ | ☐ | 1 | 1 | 1 | effective |

**10. What do you think of the pattern list : administrative location, building, place, event , geographical location, organization, person, any type?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | 2 | ☐ | ☐ | 1 | easy to use |
| ineffective | ☐ | ☐ | 3 | ☐ | ☐ | effective |

**10. What do you think of the Entity/keyword choice for metadata search?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | ☐ | 1 | useful |
| difficult to use | ☐ | ☐ | ☐ | ☐ | ☐ | easy to use |
| ineffective | ☐ | ☐ | ☐ | ☐ | ☐ | effective |

*Comment : interesting but not tested because this menu was hidden.*

**10. What do you think about the choice of labels?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | ☐ | 1 | 1 | 1 | easy to use |
| ineffective | ☐ | 1 | 1 | ☐ | 1 | effective |

**Exemple** :
Person : type, nationality, role
Administrative location : type, founded date, located in
Place: type, located in,
Building: type, located in
Event: type, date, place
Organisation: type

**11. Is there enough labels (properties)?**

| | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|
| **ineffective** | ☐ | ☐ | ☐ | 1 | 2 | **effective** |

**12. Easy to find other labels (properties) ?**

| | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|
| **ineffective** | ☐ | 1 | 1 | 1 | ☐ | **effective** |

**13. Easy to modify labels (properties)?**

| | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|
| **ineffective** | 1 | 1 | 1 | ☐ | ☐ | **effective** |

**14. What do you think of the results extracted from semantic database presented in right side of**

**the working box ?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | 1 | ☐ | 2 | useful |
| difficult to use | ☐ | ☐ | 1 | 1 | 1 | easy to use |
| ineffective | ☐ | 1 | 1 | 1 | ☐ | effective |

**15. Is all the information necessary to see the relevance of the web page displayed, easy to read, quick to read?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | ☐ | 1 | 1 | 1 | easy to use |
| ineffective | ☐ | 1 | 1 | 1 | ☐ | effective |

**16. What do you think of the choice of results display by similarity or by source /relevance?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| difficult to use | ☐ | ☐ | ☐ | 2 | 1 | easy to use |
| ineffective | ☐ | ☐ | ☐ | 1 | 2 | effective |

## 4 Media preview window

**17. What do you think of the « media preview » ?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | 1 | ☐ | 2 | useful |
| difficult to use | ☐ | 1 | 1 | 1 | ☐ | easy to use |
| ineffective | ☐ | 1 | 1 | 1 | ☐ | effective |

**18. Are the displayed web pages always pertinent?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| ineffective | ☐ | ☒ 3 | ☐ | ☐ | ☐ | effective |

**19. What do you think about the display of video and web pages in the same window ?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☒ 2 | ☒ 1 | ☐ | ☐ | useful |
| difficult to use | ☐ | ☒ 2 | ☒ 1 | ☐ | ☐ | easy to use |
| ineffective | ☐ | ☒ 2 | ☒ 1 | ☐ | ☐ | effective |

**20. What do you think about web display outside of classical web browser?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☒ 2 | ☒ 1 | ☐ | ☐ | useful |
| difficult to use | ☐ | ☒ 2 | ☒ 1 | ☐ | ☐ | easy to use |
| ineffective | ☐ | ☒ 2 | ☒ 1 | ☐ | ☐ | effective |

## 4 ASSETS annotation tool overall evaluation

**24. The interface is easy to understand:**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| disagree | ☐ | ☐ | ☐ | ☒ 2 | ☒ 1 | agree |

**25. The interface is easy to manipulate:**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| disagree | ☐ | ☐ | ☒ 1 | ☒ 2 | ☐ | agree |

**26. The visual presentation of the interface is suitable:**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| disagree | ☐ | ☐ | ☒ 2 | ☒ 1 | ☐ | agree |

**27. The response time is fast enough:**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| disagree | ☐ | ☐ | ☒ 1 | ☒ 2 | ☐ | agree |

**28. Overall, the information retrieved for annotation is relevant:**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| disagree | ☐ | ☐ | 2 | ☐ | 1 | agree |

**29. Overall, what did you think about the ASSETS annotation system?**

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | 1 | 2 | useful |
| dull | ☐ | ☐ | ☐ | 3 | ☐ | exciting |
| unoriginal | ☐ | ☐ | ☐ | 1 | 2 | innovative |

**21. What did you LIKE about the system? (please describe)**

- Access to semantic database
- Recovery from the web of formatted data ready to be include in the annotation

**22. What did you NOT LIKE about the system? (please describe)**

- Loss of functionalities for web browsing
- Display of video and web pages in the same window
- Interface organization

**23 What improvement would you suggest?**

- Suggestion for request (useful mostly for orthographic problem)

## 8.3 Appendix 3: MAT connected to ASSETS DB evaluation

### *8.3.1 Compilation of exit questionnaires*

**This compilation gathers the results of the 10 users.**

**1 ASSETS annotation functionalities**

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **1** | How do you find this service to complete and/ or correct an annotation? | not useful | | | **1** | | **9** | useful |
| | | difficult to use | | **1** | **3** | **2** | **4** | easy to use |
| | | ineffective | | | **1** | **5** | **4** | effective |
| **2** | How do you find the basic functionalities of edition of ESE annotation? | not useful | | | **1** | **1** | **8** | useful |
| | | difficult to use | | | **2** | **3** | **5** | easy to use |
| | | ineffective | | | | **2** | **8** | effective |
| **3** | How do you find the possibility to access to open web data base (Freebase, Dbpedia and Geonames)? | not useful | | | **1** | **1** | **8** | useful |
| | | difficult to use | | | **2** | **1** | **7** | easy to use |
| | | ineffective | | | | **5** | **5** | effective |
| **4** | How do you find the management of semantic categories in the working box? (*easier to get information that are already in a structure in the form of semantic fields rather than reading a page to find relevant information*) | not useful | | | | **4** | **6** | useful |
| | | difficult to use | | **1** | **1** | **3** | **5** | easy to use |
| | | ineffective | | | **1** | **3** | **6** | effective |
| **5** | To perform this scenario, do you find that some functions were missing? | Missing functions | | | **2** | **2** | **6** | No missing functions |

## 2 Record metadata panel

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| 6 | How do you like annotation display? (visual presentation) | not useful | | | 3 | 1 | 6 | useful |
| | | difficult to use | | 1 | 2 | 1 | 6 | easy to use |
| | | ineffective | | | | 4 | 6 | effective |
| 7 | What do you think of the "basic" edition functions to manually modify an annotation? | not useful | | | | 2 | 8 | useful |
| | | difficult to use | | 1 | | 4 | 5 | easy to use |
| | | ineffective | | | | 2 | 8 | effective |
| 8 | What do you think of the modification of an annotation with the metadata enrichment functionality? | not useful | | | | 3 | 7 | useful |
| | | difficult to use | | 1 | 2 | 1 | 6 | easy to use |
| | | ineffective | | | 1 | 4 | 5 | effective |

## 3 Web knowledge window

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| 9 | What do you think of the Web knowledge search? | not useful | | | | 2 | 8 | useful |
| | | difficult to use | 1 | | | 3 | 6 | easy to use |
| | | ineffective | | | | 4 | 6 | effective |
| 10 | The pages retrieved with web knowledge are relevant? | not useful | | | 1 | | 8 | useful |
| | | difficult to use | | | 1 | | 8 | easy to use |
| | | ineffective | | | 1 | 4 | 4 | effective |
| 11 | What do you think of the suggestion mode? | not useful | | | 2 | 2 | 6 | useful |
| | | difficult to use | 1 | | | 1 | 8 | easy to use |
| | | ineffective | | 5 | 2 | 2 | 1 | effective |
| 12 | What do you think of the advanced functionalities "search" or "search as" actions applied from the notice to launch a search? | not useful | | | | 1 | 5 | useful |
| | | difficult to use | 1 | | | 1 | 5 | easy to use |
| | | ineffective | | | | 1 | 5 | effective |
| 13 | What do you think of the search results display? | not useful | | | 1 | 2 | 7 | useful |
| | | difficult to use | | | | 2 | 8 | easy to use |
| | | ineffective | | | 3 | 2 | 5 | effective |
| | | Not enough information | 2 | 1 | 1 | 1 | 5 | Enough information |
| | | Too much information | | | | 1 | 6 | Enough information |
| 14 | What do you think of the working box? | not useful | | | 1 | 1 | 8 | useful |
| | | difficult to use | | 1 | 1 | 1 | 7 | easy to use |
| | | ineffective | | | 1 | 3 | 5 | effective |
| 15 | What do you think of the tabs (categories) automatically proposed in the working box? | not useful | | | 1 | 3 | 6 | useful |
| | | difficult to use | | | 5 | | 5 | easy to use |
| | | ineffective | | | 3 | 2 | 5 | effective |
| | | Not enough labels | | 1 | 1 | | 6 | Enough labels |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **16** | What do you think of the action button "metadata enrichment" to modify the annotation? | not useful | | | | **1** | **8** | useful |
| | | difficult to use | | | | | **9** | easy to use |
| | | ineffective | | | | | **9** | effective |

**4 Media preview panel**

| | | | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|---|---|
| **17** | What do you think of the « media preview » ? | not useful | | | **1** | **2** | **7** | useful |
| | | difficult to use | | | | | **10** | easy to use |
| | | ineffective | | | | **2** | **8** | effective |

**4 Web documentation panel**

| | | | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|---|---|
| **18** | What do you think of the displayed web pages in this panel? | not useful | | **1** | | **1** | **8** | useful |
| | | difficult to use | | | **1** | | **9** | easy to use |
| | | ineffective | | | **2** | **1** | **7** | effective |
| | | Not always pertinent | | | **3** | **2** | **5** | Always pertinent |

**4 ASSETS annotation tool overall evaluation**

| | | | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|---|---|
| **19** | This service is pleasant to use? | disagree | **1** | | **1** | **7** | **1** | agree |
| **20** | The interface is easy to understand? | disagree | | | **3** | **2** | **5** | agree |
| **21** | The interface is easy to manipulate? | disagree | | **1** | **4** | | **5** | agree |
| **22** | The visual presentation of the interface is suitable? | disagree | | **2** | **1** | **2** | **5** | agree |
| **23** | The response time is fast enough? | disagree | **3** | **4** | | **2** | **1** | agree |

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **24** | Overall, what did you think about the ASSETS annotation system? | not useful | | | **1** | | **9** | useful |
| | | dull | | | **4** | **2** | **4** | exciting |
| | | unoriginal | | | **1** | **4** | **5** | innovative |

## 8.4 Appendix 4: MAT propagation

### *8.4.1 Compilation of exit questionnaires*

**This compilation gathers the results of the 2 users.**

**1 ASSETS annotation functionalities**

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **1** | How do you find this framework to launch the propagation service? | not useful | | | | **2** | **2** | useful |
| | | difficult to use | | | **1** | **1** | **2** | easy to use |
| | | ineffective | | | **1** | **1** | **2** | effective |
| **2** | How do you find the possibility to access to taxonomy to give semantic tag? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | **1** | | **3** | effective |
| **3** | What do you think of the customization of work area possibilities? | not useful | | | | | **4** | useful |
| | | difficult to use | | | **1** | | **3** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **4** | To perform this scenario, do you find that some functions were missing? | Missing functions | | **1** | **1** | **1** | **1** | No missing functions |

**1 Definition of corpus**

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **5** | What did you think of the functions to create a first query? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **6** | What you think of the function to complete a corpus? | not useful | | | | | **4** | useful |
| | | difficult to use | | | **1** | **1** | **2** | easy to use |
| | | ineffective | | | **1** | **1** | **2** | effective |
| **7** | What you think of the functions of images removal? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | **1** | **3** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **8** | What you think of the access to web pages? | not useful | | | | | **4** | useful |
| | | difficult to use | | | **1** | | **3** | easy to use |
| | | ineffective | | | | **2** | **2** | effective |

## 2 Attribution of semantic tags

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **6** | How do you like documents display? (visual presentation) | not useful | | | **2** | | **2** | useful |
| | | difficult to use | | | | **2** | **2** | easy to use |
| | | ineffective | | | **2** | | **2** | effective |
| **7** | What do you think of the taxonomy display? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | **1** | **3** | easy to use |
| | | ineffective | | | | **1** | **3** | effective |
| **8** | What do you think of the search functionalities in the taxonomy? | not useful | | | | | **4** | useful |
| | | difficult to use | | | **1** | | **3** | easy to use |
| | | ineffective | | | | | **4** | effective |

## 3 Checking of results

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **9** | Could it be useful to have access to the propagation results if the corpus is very large? | not useful | | | | | **4** | useful |
| **10** | How do you like results display? (visual presentation) | not useful | | | **1** | **1** | **2** | useful |
| | | difficult to use | | | **2** | | **2** | easy to use |
| | | ineffective | | | **2** | **1** | **1** | effective |
| **11** | What do you think of the correction functionality? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | **2** | **2** | easy to use |
| | | ineffective | | | | | **4** | effective |

**Free comment: …**

## 4 ASSETS annotation tool overall evaluation

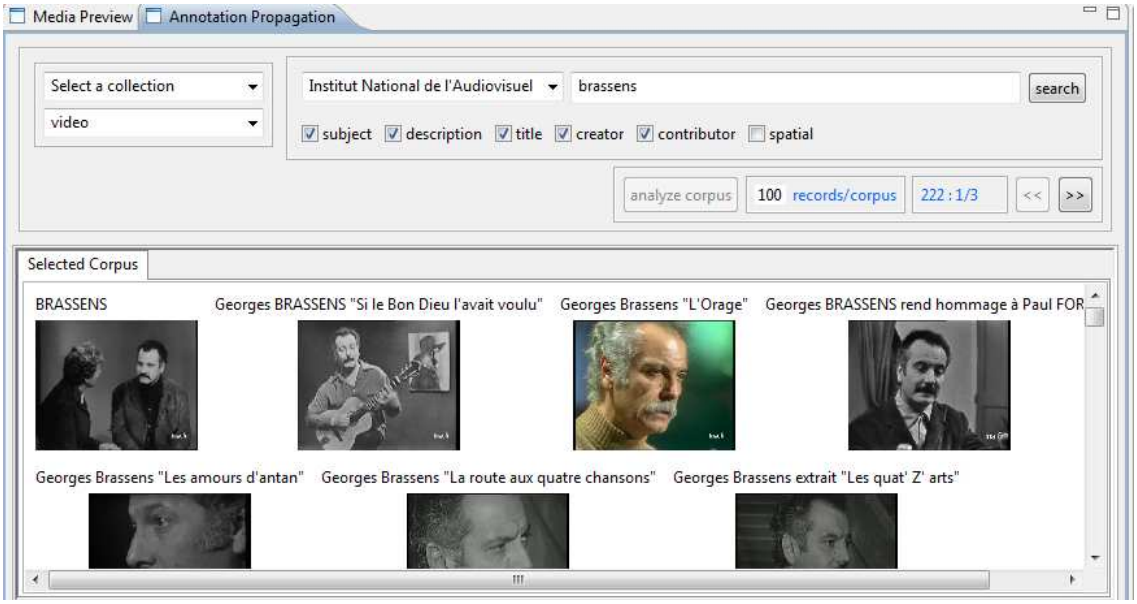
| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **12** | This service is pleasant to use? | disagree | | **1** | **1** | | **2** | agree |
| **13** | The interface is easy to understand? | disagree | | **2** | | **1** | **1** | agree |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **14** | The interface is easy to manipulate? | disagree | | **1** | **1** | | **2** | agree |
| **15** | The visual presentation of the interface is suitable? | disagree | | **1** | **1** | | **2** | agree |
| **16** | The response time is fast enough? | disagree | | **2** | | | **2** | agree |

| | | | **1** | **2** | **3** | **4** | **5** | |
|---|---|---|---|---|---|---|---|---|
| **17** | Overall, what did you think about the ASSETS annotation system? | not useful | | | | **2** | **2** | useful |
| | | dull | | | **3** | | | exciting |
| | | unoriginal | | | **1** | **1** | **1** | innovative |

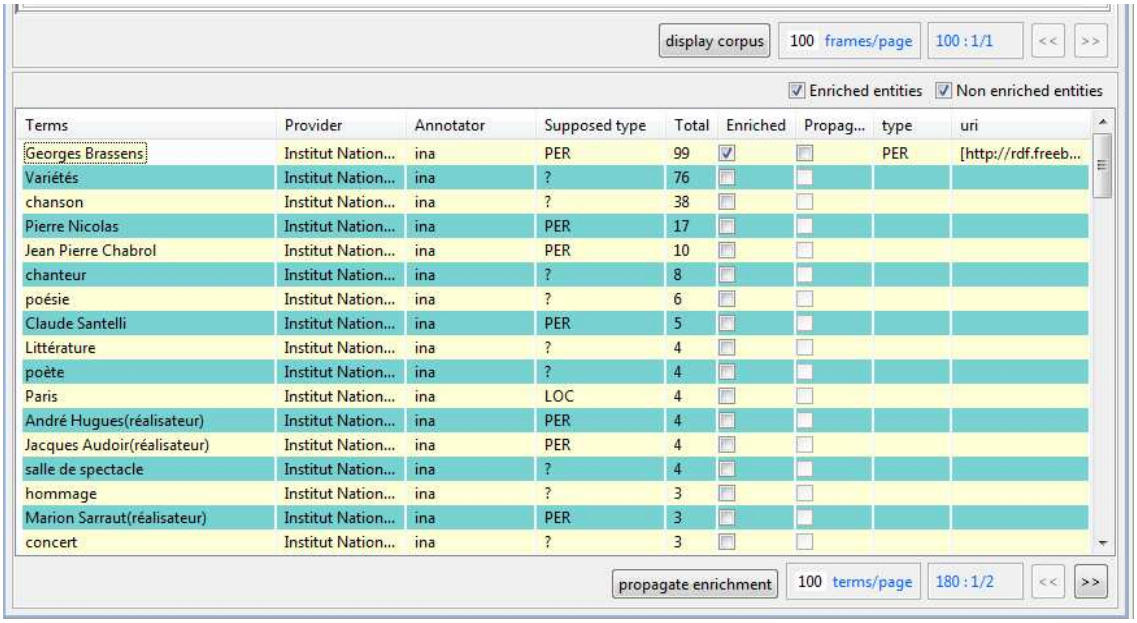
## 8.5 Appendix 5 : MAT enrichment (manual enrichment propagation)

### *8.5.1 Reporting of cognitive walkthrough results*




Task 1: Selection of corpus about Brassens with 177 records.

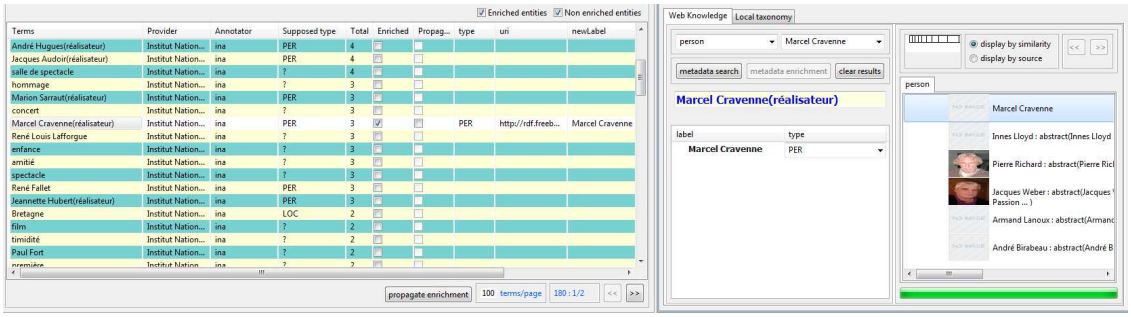
| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | Very classic task. No problem. |
| 2. | Does the user understand the proposed functionalities? | yes |
| 3. | Are the proposed functionalities easy to find, easy to combine? | yes |
| 4. | Does the user understand input and output of the system? | yes |
| 5. | Is the visual design of the interface relevant? | A problem was noticed in the panel displaying the thumbnails. If a title is very long, it will take a very large place and so the adjacent thumbnails will be positioned too far away. |




Task 2: Get the terms

- Click on *analyze corpus* button to process the annotation metadata and display all the named entities and concepts automatically extracted from specific fields of the annotations.
  351 terms have been collected in the displayed table and they are ordered by number of occurrence in the set of annotations.

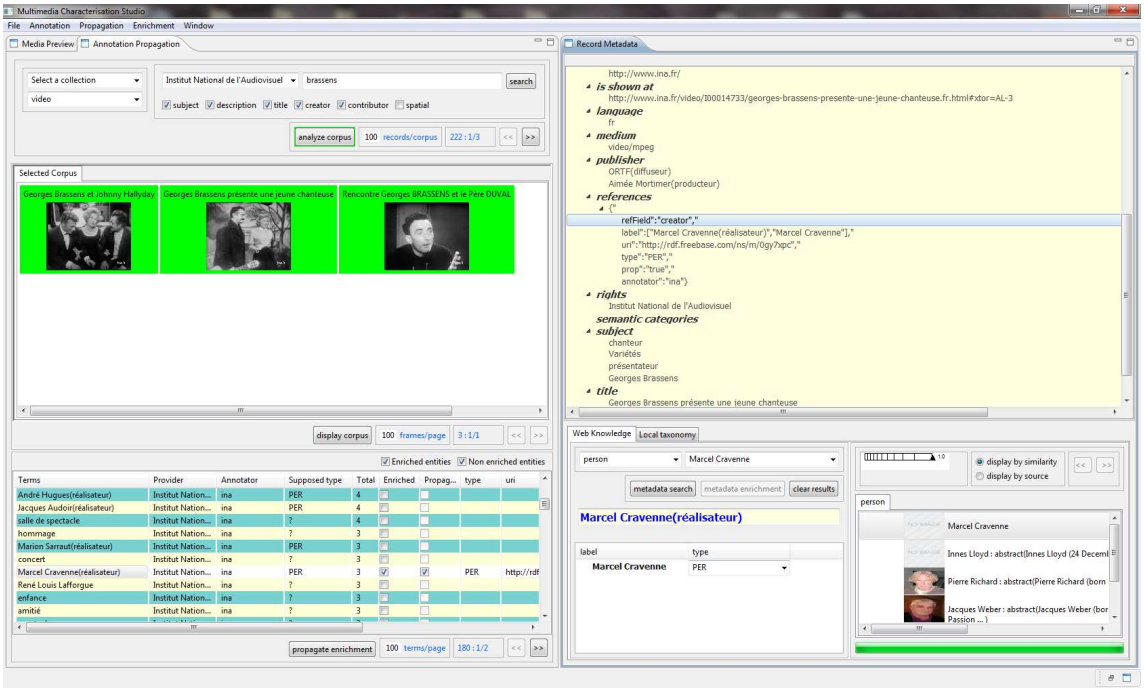
| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | This service is dedicated to professional users to achieve very specific tasks. A learning step is needed to use this service which is very new for annotators used to classical system. |
| 2. | Does the user understand the proposed functionalities? | No difficult to understand this interface goal with a documentation. |
| 3. | Are the proposed functionalities easy to find, easy to combine? | For this task just a button to push. |
| 4. | Does the user understand input and output of the system? | Outputs are displayed in the table. It will not be difficult to understand for annotators. |
| 5. | Is the visual design of the interface relevant? | Same general remarks than for the other evaluations about MAT. Perhaps a title could be added at the top of the table of the term. |




Task3: Manual enrichment

- The user choses one by one the term that he would like to manually enrich. He imports it with drag-and-drop in the web knowledge panel. The web search proposes candidates. The user choses one and drag-and-drop it in the working box. He could verify in the web documentation panel that it is the right page. He selects the type in the menu. Then he drags-and-drops the label in the upper box to update the table.

| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | |
| 2. | Does the user understand the proposed functionalities? | No difficult to understand with a documentation. Experience with other MAT scenario could be useful. |
| 3. | Are the proposed functionalities easy to find, easy to combine? | Drag-and-drop is very easy to do but all the functionalities are not easy to find for the first time. Some functionalities are hidden, for example drag-and-drop a term in the selected corpus panel to display all the documents having this term in their metadata. It is also not intuitive to know that you have to drag-and-drop the label from the working box to the upper box in order to enter it in the table of terms. When you have discovered this action, it is then very efficient. The drag-and-drop is also a consistent interaction for all the actions to do with MAT. |
| 4. | Does the user understand input and output of the system? | The use needs to read the documentation. It is not intuitive to understand the connections between different panels, but after it is easy to use and efficient. |
| 5. | Is the visual design of the interface relevant? | Here the working box has too much place because only one line has to be displayed. |




Task 4: propagate the enrichment

- When he has done this for several terms, he could click on *propagate enrichment* button. Then he could drag-and-drop a term in the *Selected corpus* panel to visualize the corresponding documents having be enriched for this term. He could also drag-and-drop the term or a document in the *Record metadata* panel to visualize the annotation.

| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | yes |
| 2. | Does the user understand the proposed functionalities? | yes |
| 3. | Are the proposed functionalities easy to find, easy to combine? | yes |
| 4. | Does the user understand input and output of the system? | Yes, always drag-and-drop. Enrichment time is very long, and no progress bar displayed. |
| 5. | Is the visual design of the interface relevant? | Same remarks as previous. |

## 8.6 Appendix 6: Content by re-use

### *8.6.1 Interface presentation*


**Figure 44: Accessing the Assets application**

Next, the following image (X and X below) shows how to upload a file from your desktop using the Assets UGC tool and store it in the left pane.

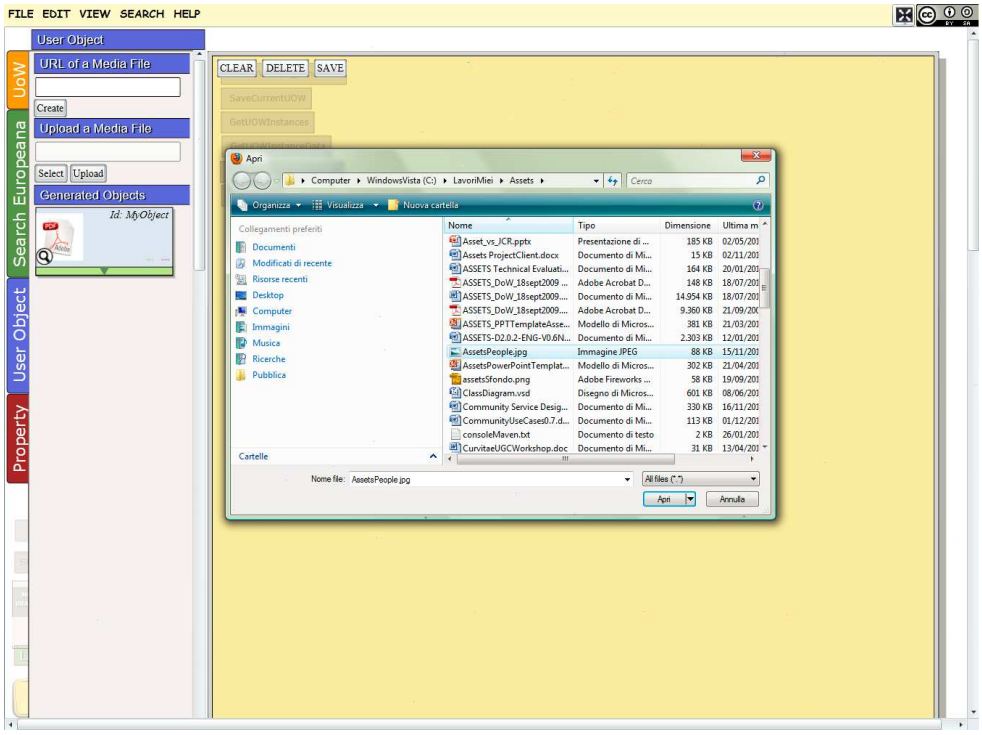



**Figure 45: Upload a file**

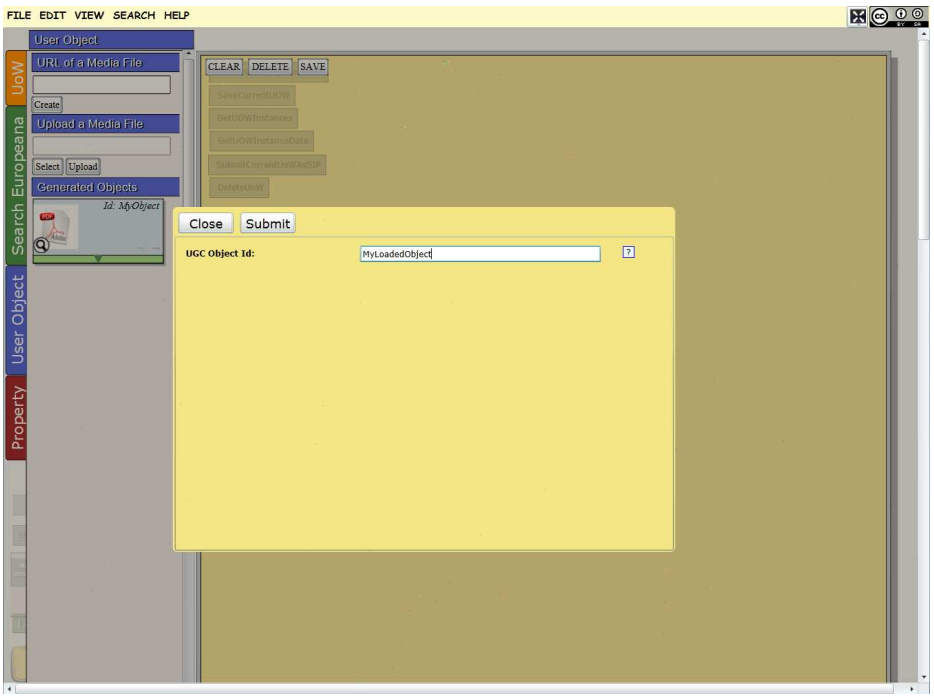



**Figure 46: Submit an uploaded file to UGC tool**

After uploading one or more files, the next function is searching for an object in Europeana digital

repository. The result list is presented to the left in the result pane and each object is represented by a thumbnail and metadata information.

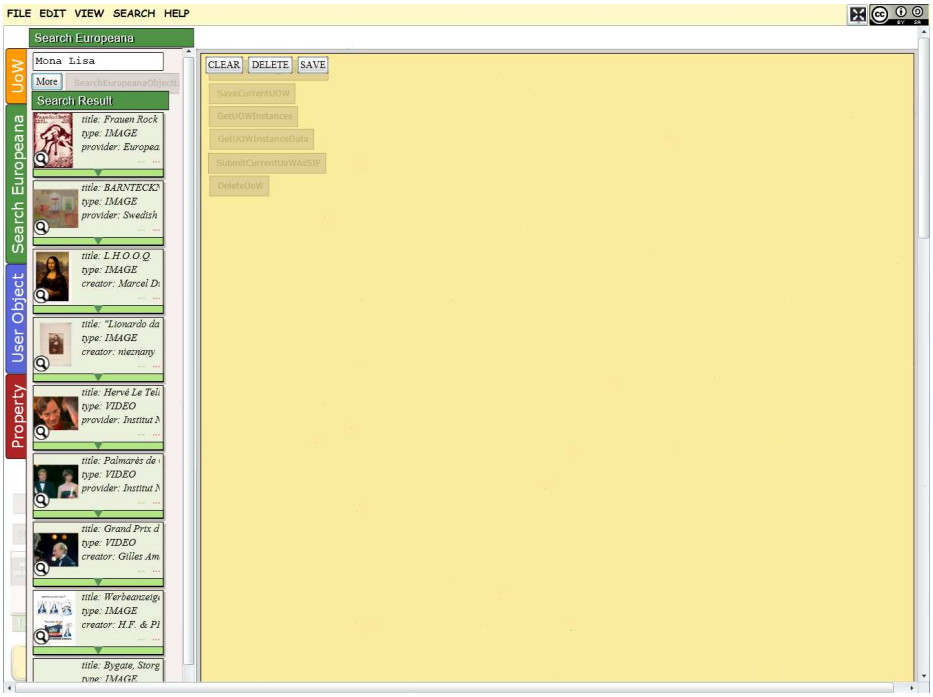



**Figure 47: Search Europeana**

After uploading a file to the UGC tool, the user might be able to add some metadata properties to the Europeana object. The tool then provide a pane to the left with a list of metadata properties to select and using a drag-and-drop functionality, the user can drag a certain property to the object and drop it on the symbol  marked with an 'x'

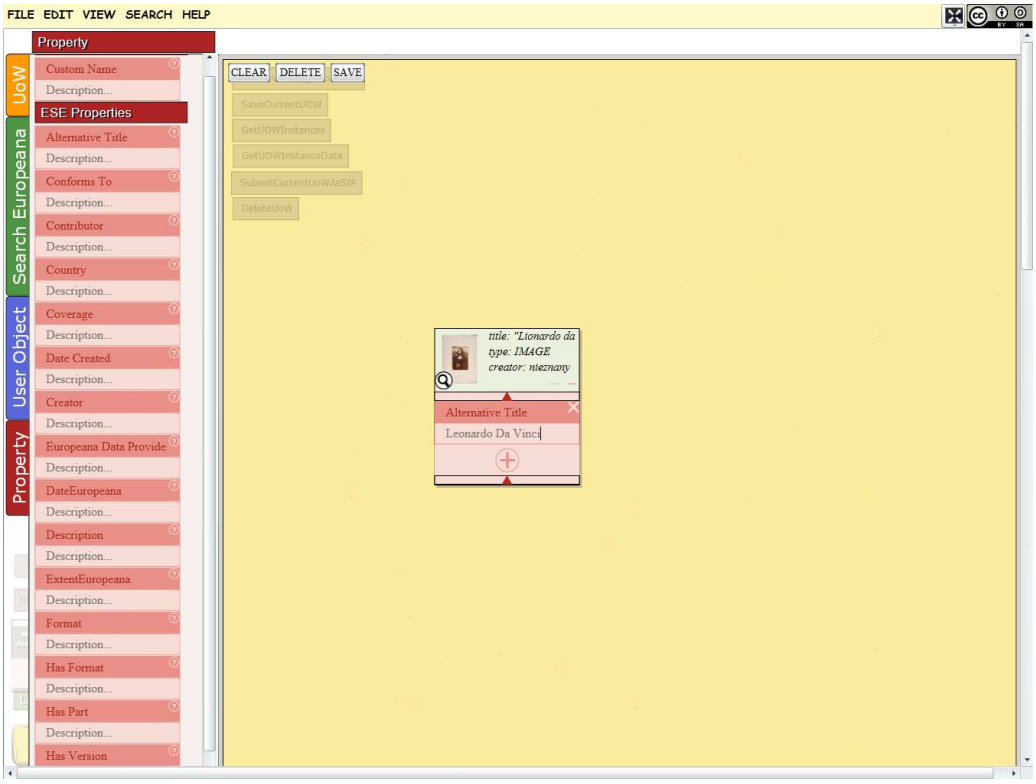



**Figure 48: Add a property to a Europeana Object**

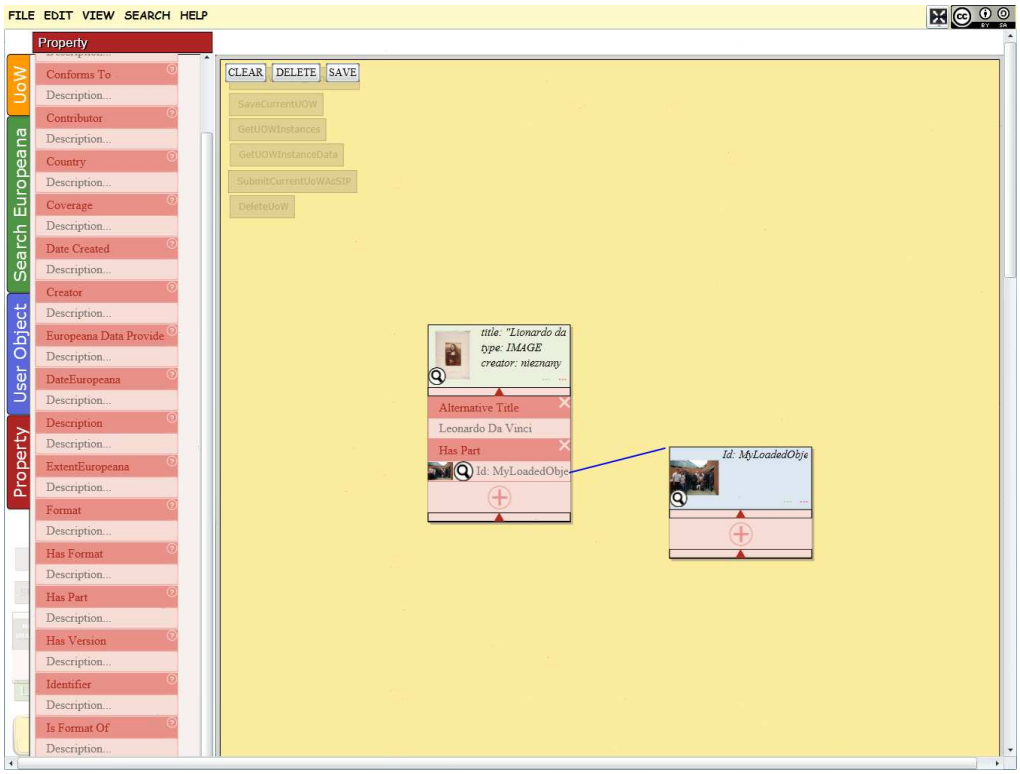



**Figure 49: Create a compound object**

### 8.6.2 Compilation of questionnaires

This appendix is summarizing the data from 13 users and contains of five different questionnaires.

# ASSETS evaluation

# Content Creation by re-use service

# PRE-QUESTIONNAIRE

**USER ID** :

**DATE** :

**LOCATION OF TEST** :

## 1.1 Personal details

**1. Gender**

9 Male 4 Female

**2. Age**

☐ 18-25 ☐26-40 6 ☐41-50 5 ☐51-65 ☐65 – 2

**2. What is your relationship to the ASSETS project?**

☐I am an ASSETS project partner -- 10
☐I collaborate with an ASSETS partner -- 2
☐I have no connection -- 1

**3. What is your qualification?**

☐ I am a content provider --9
☐ Librarian
☐ Archivist
☐ Developer/programmer -- 3
☐ HCI designer
☐ Another, please describe:…… -- 1

**4. Which system do you currently use to find information for other task (educational, leisure, entertainment, etc )?**

☐ Search engine (Google,..) -- 12
☐ Wikipedia …) -- 5
☐ Europeana -- 1
☐ The European Digital Library - EDL
☐ Other…… --3

**5. Have you ever used the Europeana site??**

☐ Yes ☐ No

12 1

**a) If yes, for what purpose?**

**b) If no, what would you expect from a site like the Europeana site?**

## 1.2 Metadata experience

**9. Do you *add information or objects* to different types of sites (collaborative platforms, blogs, flickr etc…)?**

☐ No experience -- 3
☐ Yes, some experience -- 8
☐ Yes, on a regular basis but not that often --2
☐ Yes, I do it in my daily work

**4. Do you have previous experience in *metadata* assignment?**

☐ No experience --2
☐ Yes, some experience --7
☐ Yes, on a regular basis but not that often --2
☐ Yes, I do it in my daily work -- 2

**5. How long experience do you have in assigning metadata to objects?**

☐ Less than one year -- 5
☐ 1-5 years -- 2
☐ 5-10 years -- 2
☐ more than 10 years --4

Please go back to Task 1.

# QUESTIONNAIRE 1 / Task 1

**USER ID** :

**DATE** :

**LOCATION** :

**1. Installation of UGC tool and instructions**

1. Did you find the instructions for installing the UGC tool easy to understand?

   Yes ☐ Mostly ☐ Partially ☐ No ☐

   8 3 1

2. Did you succeed to install the UGC tool/interface on your desktop?

   Yes ☐ Mostly ☐ Partially ☐ No ☐

   10 1 1 1

3. Did you find the length of the task to be difficult?

Yes ☐ Mostly ☐ Partially ☐ No ☐
1 3 9

4. How easy was it to learn to use the UGC tool?

Yes ☐ Mostly ☐ Partially ☐ No ☐
5 5 2

**2.** Uploading a media file/ object and adding metadata value

| | | YES | Mostly | Partially | NO |
|---|---|---|---|---|---|
| | **Uploading an object** | | | | |
| 1 | Did you manage to *select* an object to be uploaded to the UGC tool? | 13 | | | |
| 2 | Did you succeed to upload an object? | 11 | 2 | | |
| 3 | Was it easy to understand were the uploaded object appeared? | 9 | 4 | | |
| 4 | Ddi you find it easy to give the object a name? | 13 | | | |
| | **Adding metadata value to an object** | | | | |
| 5 | Did you find it easy to expand the object for metadata assignment? | 6 | 6 | 1 | |
| 6 | Did you understand the '+' sign on the object? | 9 | 3 | 1 | |
| 7 | Did you understand what was meant by "property"? | 8 | 4 | 1 | |
| | **General comments** | | | | |
| 8 | Did you manage to understand the tabs on the left side? | 5 | 1 | 7 | |
| 9 | Did you find it easy to interact with the objects in the UGC interface? | 6 | 6 | | 1 |
| 10 | Was it easy to drag and drop the objects? | 7 | 1 | 3 | 1 |
| 11 | Was it easy to find the functions in the interface? | 8 | 3 | 1 | 1 |
| 12 | Did you need to repeat any action (s) in order to succeed while performing task 1? | 3 | 2 | 4 | 4 |
| 13 | Did you find the menus/tabs of the interface easy to understand? | 7 | 2 | 3 | 1 |
| 14 | Did you encounter any errors while performing this task? | 5 | | 1 | 7 |
| 15 | Did you need to make any correction during the task? | 2 | 1 | 4 | 6 |
| 16 | Did you find it difficult to complete this task? | 1 | | 2 | 10 |

# QUESTIONNAIRE 2 /TASK 2

## Task 2: Searching Europeana for an object and adding metadata value(s)

| | | YES | Mostly | Partially | NO |
|---|---|---|---|---|---|
| | **Searching Europeana** | | | | |
| 1 | Did you succeed to search for an object in Europeana? | 12 | | 1 | |
| 2 | Did you find the presentation of the objects in the result list to the left easy to understand? | 9 | 2 | 1 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| 3 | Did you understand how to “drag-and-drop” the object to the yellow client area? | 11 | | 2 | |
| | **Adding metadata value(s) to an object** | | | | |
| 4 | Did you find it easy to expand the object for metadata assignment? | 10 | 2 | 1 | |
| 5 | Did you understand the ‘+’ sign on the object? | 12 | | 1 | |
| 6 | Did you understand what was meant by “property” | 12 | | | |
| | **General questions** | | | | |
| 7 | Did you manage to understand the tabs on the left side? | 11 | 2 | | |
| 8 | Did you find it easy to interact with the functions in the GUI? | 7 | 5 | | 1 |
| 9 | Was it easy to drag and drop the objects? | 9 | 3 | 1 | |
| 10 | Was it easy to find the functions in the interface? | 9 | 2 | | 2 |
| 11 | Did you need to repeat any action (s) in order to succeed while performing task 2? | 3 | 1 | 5 | 4 |
| 12 | Did you find the menus/tabs of the GUI easy to understand? | 8 | 2 | 2 | 1 |
| 13 | Did you encounter any errors while performing this task? | 4 | | | 9 |
| 14 | Did you need to make any correction during the task? | 1 | | 1 | 11 |
| 15 | Did you find it difficult to complete this task? | 1 | | | 12 |

# QUESTIONNAIRE 3 / Task 3

**Task 3: Create a complex object involving linking a Europeana object and a user uploaded object and assign metadata value(s).**

| | | YES | Mostly | Partially | NO |
|---|---|---|---|---|---|
| | **Linking Europeana object with user uploaded object** | | | | |
| 1 | Did you find it easy to move between the tabs on left side of the GUI? | 13 | | | |
| 2 | Did you find the presentation of the objects to the left easy to understand? | 9 | 3 | 1 | |
| 3 | Did you find the number of actions satisfactory for creating the links between objects? | 10 | 1 | 1 | 1 |
| 4 | Did you understand how to “drag-and-drop” the object from left pane to the yellow client area? | 11 | 2 | | |
| | **Adding metadata value(s) to an object** | | | | |
| 5 | Did you find it easy to expand the object for metadata assignment? | 12 | | 1 | |
| 6 | Did you understand the ‘+’ sign on the object? | 10 | 3 | | |
| 7 | Did you understand what was meant by “property”? | 12 | 1 | | |
| | **General comments** | | | | |
| 8 | Did you manage to understand the tabs on the left side? | 12 | 1 | | |

| | | | | | |
|---|---|---|---|---|---|
| 9 | Did you find it easy to interact with the functions in the GUI? | 9 | 2 | 1 | 1 |
| 10 | Was it easy to drag and drop the objects? | 8 | 2 | 3 | |
| 11 | Was it easy to find the different functions in the interface? | 11 | | 1 | 1 |
| 12 | Did you need to repeat any action (s) in order to succeed while performing task 2? | 3 | 1 | 3 | 6 |
| 13 | Did you find the menus/tabs of the interface easy to understand? | 10 | 1 | 1 | 1 |
| 14 | Did you encounter any errors while performing this task? | 3 | | 1 | 9 |
| 15 | Did you need to make any correction during the task? | 2 | 1 | 1 | 8 |
| 16 | Did you find it difficult to complete this task? | | 1 | 2 | 9 |
| | | | | | |

# POST-QUESTIONNAIRE

Scale: 1=No; 2=partially; 3= mostly; 4=Yes

## 1. Content creation by re-use functionalities

| | | | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|---|
| **1** | How do you find the overall functions in order to complate the task(s)? | not useful | | **1** | | **11** | useful |
| | | difficult to use | | | **4** | **8** | easy to use |
| | | ineffective | **1** | | **2** | **9** | effective |
| **2** | How do you find the overall number of functions in order to complete the task(s)? | not useful | | | **2** | **8** | useful |
| | | difficult to use | | | **2** | **8** | easy to use |
| | | ineffective | | | **2** | **8** | effective |
| **3** | How did you find the workarea when manipulating the objects and metadata ? | not useful | **1** | **1** | **1** | **10** | useful |
| | | difficult to use | **1** | **2** | **4** | **5** | easy to use |
| | | ineffective | **2** | **1** | **1** | **9** | effective |
| **4** | In order to perform the task(s), did you find that some functions were missing? | Missing functions | **1** | | **4** | **8** | No missing functions |

## 2. Uploading and naming an object

| | | | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|---|---|
| **5** | What did you think of creating an object by the uploading function? | not useful | | **1** | | **12** | useful |
| | | difficult to use | **1** | **2** | **4** | **6** | easy to use |
| | | ineffective | **2** | | **1** | **10** | effective |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **6** | What you think of the function of naming the objects? | not useful | | **1** | **1** | **11** | useful |
| | | difficult to use | **1** | **1** | **2** | **9** | easy to use |
| | | ineffective | **1** | **1** | **1** | **10** | effective |
| **7** | How did you find moving the uploaded object onto the work-area?? | not useful | | | **1** | **12** | useful |
| | | difficult to use | **1** | **3** | **1** | **8** | easy to use |
| | | ineffective | **1** | **2** | | **10** | effective |

## 3. Adding metadata properties

| | | | **1** | **2** | **3** | **4** | |
|---|---|---|---|---|---|---|---|
| **6** | How did you find the display of the metadata properties in the tab to the left? | not useful | | **2** | **1** | **10** | useful |
| | | difficult to use | **1** | **1** | **6** | **5** | easy to use |
| | | ineffective | **2** | | **3** | **8** | effective |
| **7** | How did you find the metadata taxonomy? | not useful | **1** | **2** | **1** | **9** | useful |
| | | difficult to use | **1** | **2** | **3** | **7** | easy to use |
| | | ineffective | **2** | | **5** | **6** | effective |
| **8** | What do you think about adding metadata to an object? | not useful | | | **1** | **12** | useful |
| | | difficult to use | **1** | **2** | **5** | **4** | easy to use |
| | | ineffective | **2** | **2** | **2** | **7** | effective |

## 4. Searching Europeana for an object

| | | | **1** | **2** | **3** | **4** | |
|---|---|---|---|---|---|---|---|
| **9** | How did you find searching an object from Europeana? | not useful | | **1** | **2** | **10** | useful |
| | | difficult to use | | **1** | **3** | **9** | easy to use |
| | | ineffective | **1** | **1** | **3** | **8** | effective |
| **10** | How do you find the presentation (logical and semantic) of the results in the left pane? | not useful | **2** | **2** | **1** | **7** | useful |
| | | difficult to use | **2** | | **7** | **4** | easy to use |
| | | ineffective | **2** | **1** | **4** | **6** | effective |
| **11** | How did you find the object description from an uploaded object from an Europeana search. | not useful | | **3** | | **10** | useful |
| | | difficult to use | **1** | **2** | | **9** | easy to use |
| | | ineffective | **1** | **1** | **4** | **7** | effective |

## 5. Linking two or more objects

| | | | **1** | **2** | **3** | **4** | |
|---|---|---|---|---|---|---|---|
| **9** | **How did you find the function of linking/combining 2 or more objects?** | not useful | | **1** | **1** | **11** | useful |
| | | difficult to use | | **3** | **3** | **7** | easy to use |
| | | ineffective | **2** | **1** | **1** | **9** | effective |
| **10** | How did you find the linking of object in order to enhance the creation of a new object (set of objects)? | not useful | | **1** | **1** | **11** | useful |
| | | difficult to use | | **1** | **8** | **4** | easy to use |
| | | ineffective | **2** | | **1** | **10** | effective |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **11** | Was it easy to understand the procedure of how to link objects? | difficult to use | | **4** | | **9** | easy to use |

## 6. Content creation by re-use overall performance

| | | | **1** | **2** | **3** | **4** | |
|---|---|---|---|---|---|---|---|
| **12** | The service was enjoyable to use. | disagree | **1** | **1** | **1** | **10** | agree |
| **13** | The interface was easy to understand. | disagree | **1** | | **3** | **9** | agree |
| **14** | The interface was easy to manipulate. | disagree | **1** | **1** | **3** | **8** | agree |
| **15** | The visual presentation of the interface was easy to understand. | disagree | **1** | | **7** | **4** | agree |
| **17** | The response time for different action was acceptable. | disagree | **1** | **1** | **4** | **7** | agree |
| **18** | I did understand what was going on within the content creation by re-use service as well as uploading objects from outside into the service. | disagree | | **1** | **3** | **9** | agree |
| **19** | I understood what was happening with the objects when creating a new object. | disagree | **1** | **2** | **1** | **9** | agree |

| | | | **1** | **2** | **3** | **4** | |
|---|---|---|---|---|---|---|---|
| **20** | Finally, what is your overall impression of the content creation by re-use service? | not useful | **1** | **1** | **1** | **10** | useful |
| | | Unexciting | **1** | **2** | **5** | **5** | exciting |
| | | unoriginal | **1** | **1** | **5** | **6** | innovative |

## 8.7 Appendix 7: Access to Query log analysis

### *8.7.1 Reporting of cognitive walkthrough results*

**Task 1**: Define an initial visualization of the number of access by Quarter and Region


| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | He will have to select row and column and visualization will display the chart |
| 2. | Does the user understand the proposed functionalities? | yes |
| 3. | Are the proposed functionalities easy to find, easy to combine? | It is very simpler for this case |
| 4. | Does the user understand input and output of the system? | yes |
| 5. | Is the visual design of the interface relevant? | Yes (but it depends on the size of all the elements to visualize) |

We observe a majority of European users.

**Task 2**: Focus on sub-region in Europe


| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | yes clicking row and column to restrict selection |
| 2. | Does the user understand the proposed functionalities? | yes |
| 3. | Are the proposed functionalities easy to find, easy to combine? | yes |
| 4. | Does the user understand input and output of the system? | Yes, but some information are missing to report when the system is running. When you click on Europe, it is quite long to obtain the results, and the user doesn't know if the system is working or not. |
| 5. | Is the visual design of the interface relevant? | Yes |

**Task 3**: Observe the number of web access during 4 periods (of 6 hours) in a day.

The system gives the average on the month of connections during these periods of day.


| | | |
|---|---|---|
| 1. | Does the user know what to do to achieve the task? | yes |
| 2. | Does the user understand the proposed functionalities? | yes |
| 3. | Are the proposed functionalities easy to find, easy to combine? | Yes, very easy to use |
| 4. | Does the user understand input and output of the system? | Yes, very easy to understand |
| 5. | Is the visual design of the interface relevant? | Yes |

**Task 4**: analysis of the different web browsers.


Association of data of web browser to color: in blue firefox, in green explorer.

We can see for example that in November the majority is firefox during all day and in south of Europe there are more explorer web browser.

| | | |
|---|---|---|
| 6. | Does the user know what to do to achieve the task? | yes |
| 7. | Does the user understand the proposed functionalities? | yes |
| 8. | Are the proposed functionalities easy to find, easy to combine? | Yes, menu is offering a lot of choice |
| 9. | Does the user understand input and output of the system? | Yes, there are some pup up menu to display numbers. |
| 10. | Is the visual design of the interface relevant? | Yes<br>The selection done in the cube navigator could be more visible. |

### *8.7.2 Compilation of questionnaires*

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **1** | How do you find this service? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **2** | How do you find pivoting functionality? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **3** | How do you find slicing functionality? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **4** | How do you find the drill dow/up functionality? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **5** | How do you find cube navigator? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |
| **6** | How do you find chart navigator? | not useful | | | | | **4** | useful |
| | | difficult to use | | | | | **4** | easy to use |
| | | ineffective | | | | | **4** | effective |

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **12** | This service is pleasant to use? | disagree | | | | | **4** | agree |
| **13** | The interface is easy to understand? | disagree | | | | | **4** | agree |
| **14** | The interface is easy to manipulate? | disagree | | | | | **4** | agree |
| **15** | The visual presentation of the interface is suitable? | disagree | | | | | **4** | agree |
| **16** | The response time is fast enough?<br>Not today | disagree | | **2** | **2** | | | agree |

## 8.8 Appendix 8: Ingestion workflow

### *8.8.1 Compilation of exit questionnaires for metadata classification*

**1 ASSETS ingestion workflow functionalities**

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **1** | How do you find this service? | not useful | | | | **1** | | useful |
| | | difficult to use | | | **1** | | | easy to use |
| | | ineffective | | | | **1** | | effective |
| **2** | To perform this enrichment of metadata, do you find that some functions were missing? | Missing functions | | **1** | | | | No missing functions |

**4 ASSETS ingestion workflow overall evaluation**

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **3** | This service is pleasant to use? | disagree | | | **1** | | | agree |
| **4** | The interface is easy to understand? | disagree | | | **1** | | | agree |
| **5** | The interface is easy to manipulate? | disagree | | | **1** | | | agree |
| **6** | The visual presentation of the interface is suitable? | disagree | | | **1** | | | agree |
| **7** | The response time is fast enough? | disagree | | | | **1** | | agree |

| | | | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|---|
| **8** | Overall, what did you think about the ASSETS ingestion system? | not useful | | | | **1** | | useful |
| | | dull | | | **1** | | | exciting |
| | | unoriginal | | | **1** | | | innovative |