

## ANNEX T1

# HIGH LEVEL REQUIREMENTS FOR THE NEW EUR-LEX

---

**Indicative document for information purposes –  
subject to modification after the competitive  
dialogue**

**Important Note :**

Some parts of the functional and technical specifications for the new EUR-Lex were drawn up before the Publications Office had finalised its new layered architecture. Where there is potential conflict between the text used in these specifications and the layered model, **it is the contractor's responsibility** to propose a solution which fully respects the independence and interoperability of the presentation, search and content layers and allows other systems which will form part of the web portal to share user profiles, metadata and content with new EUR-Lex.

Author	PUBLICATIONS OFFICE
Version	4.0
Date	07/07/2009
Status	Reviewed

# Table of content

<b>1</b>	<b>THE TRANSFORMATION PROGRAM OF THE PUBLICATIONS OFFICE.....</b>	<b>4</b>
<b>2</b>	<b>PURPOSE OF THE DOCUMENT.....</b>	<b>7</b>
<b>3</b>	<b>REQUIREMENTS OVERVIEW.....</b>	<b>8</b>
<b>4</b>	<b>FRONT OFFICE.....</b>	<b>9</b>
4.1	ERGONOMICS.....	9
4.2	NAVIGATE.....	10
4.2.1	Home page.....	10
4.2.2	Header.....	10
4.2.3	Footer.....	11
4.2.4	Menu.....	11
4.2.5	Breadcrumb.....	11
4.3	VIEWS.....	11
4.3.1	Official Journal.....	11
4.3.1.1	Official Journal page.....	11
4.3.1.2	Table of content.....	12
4.3.1.2.1	Access an act.....	12
4.3.1.2.2	Latest issues.....	12
4.3.1.2.3	PDF version.....	12
4.3.1.2.4	Electronic certified version.....	12
4.3.1.2.5	Notice.....	12
4.3.2	Collections.....	12
4.3.3	Thematic files.....	13
4.3.4	Other views.....	13
4.4	CONTENT CONSULTATION.....	13
4.4.1	Display formats.....	13
4.4.2	URL.....	14
4.4.3	Notice.....	14
4.4.4	Historical view.....	14
4.4.5	Multilingual display.....	14
4.4.6	Add to my cart.....	14
4.4.7	Extend your search to.....	14
4.4.8	Share link.....	15
4.5	SEARCH SERVICE.....	15
4.5.1	Simple search.....	15
4.5.2	Advanced search.....	16
4.5.3	Expert search.....	17
4.5.4	National search.....	17
4.5.5	Search results.....	18
4.5.5.1	Display.....	18
4.5.5.2	Facets.....	18
4.5.5.3	Export results.....	19
4.6	EXTRA FEATURES.....	19
4.6.1	Get started.....	19
4.6.2	Social network.....	19
4.7	USER ACCOUNT.....	19
4.8	USERS' FEEDBACK.....	19
4.8.1	Preferences.....	20
4.8.1.1	Languages.....	20
4.8.1.2	Search results.....	20
4.8.1.3	Notify me.....	20
4.8.2	Syndication/notification.....	20
4.8.3	Cart.....	20
4.8.4	Exports: web services for re-users.....	20
4.9	PORTABILITY.....	21

<b>5</b>	<b>BACK OFFICE .....</b>	<b>22</b>
5.1	SITE MANAGEMENT .....	22
5.1.1	<i>Content elements</i> .....	22
5.1.2	<i>Site structure management</i> .....	22
5.1.3	<i>Multilingualism</i> .....	23
5.2	CONTENT MANAGEMENT .....	23
5.2.1	<i>Management of content provided by production chains</i> .....	23
5.2.1.1	Regular content management .....	23
5.2.1.1.1	Store/ Import/Flow management.....	23
5.2.1.1.2	Updating processes.....	24
5.2.1.1.3	Exceptions management.....	24
5.2.1.2	Management of early publishing .....	24
5.2.1.3	Indexation.....	24
5.2.1.4	Management of certified documents and authenticity .....	24
5.2.1.5	Tracking document feeding .....	24
5.2.2	<i>External content management</i> .....	25
5.2.2.1	Deep linking .....	25
5.2.3	<i>Web Oriented Contribution (editorial content)</i> .....	25
5.2.3.1	Possible EUR-Lex Site Map and contents .....	26
5.2.3.2	Content Models .....	26
5.2.3.3	Back-Office Structure concerning the editorial parts .....	27
5.2.3.4	Contribution process.....	28
5.2.3.5	Creation/edition .....	29
5.2.4	<i>Migration of present EUR-Lex content and data</i> .....	30
5.3	ADMINISTRATION .....	31
5.3.1	<i>Users management</i> .....	31
5.3.1.1	Front Office users .....	31
5.3.1.2	Back-Office /editorial users.....	31
5.3.2	<i>Web services for re-users</i> .....	32
5.3.3	<i>Statistics</i> .....	33
5.3.3.1	Statistics of use.....	33
5.3.3.2	Production statistics.....	33
5.3.3.1	Statistics availability.....	33
<b>6</b>	<b>OPERATIONS.....</b>	<b>34</b>
<b>7</b>	<b>ANNEXES.....</b>	<b>35</b>
7.1	SEARCH STATE OF THE ART.....	35
7.1.1	<i>Advanced Linguistic Processing</i> .....	35
7.1.1.1	Linguistics and Relevancy.....	35
7.1.1.2	Linguistic Concepts.....	35
7.1.1.2.1	Lemmatization .....	36
7.1.1.2.1.1	What is Lemmatization? .....	36
7.1.1.2.1.2	Lemmatization by Expansion and Reduction.....	36
7.1.1.2.1.3	Lemmatization and its Influence on the Search Experience.....	36
7.1.1.2.2	Synonyms and Spell Variations .....	36
7.1.1.2.3	Spell Checking .....	37
7.1.1.2.4	Anti-Phrasing .....	37
7.1.1.2.5	Wildcard Search.....	37
7.1.2	<i>Facetted search and taxonomies</i> .....	37
7.1.2.1	Facets.....	37
7.1.2.2	Taxonomy .....	39
7.1.2.3	Unsupervised Clustering .....	39
7.1.2.4	Creating Taxonomy on the Fly .....	39
7.2	ABOUT CONTENT MANAGEMENT SYSTEM.....	39
7.3	TABLE OF FIGURES.....	41

# 1 The transformation program of the PUBLICATIONS OFFICE

PUBLICATIONS OFFICE, the Publications Office of the European Union aims to continuously improve its highest-quality services to its customers — the originating departments of the institutions and other bodies of the European Union and to its public — the citizens of the European Union and people throughout the world who are interested in European affairs.

In order to retain and strengthen its publishing position, a Transformation Program has been initiated.

The objectives of the Transformation Program include:

- Improving the public value of the publications office including:
- Easier use of the access portal(s)
- Easier to find and obtain publications
- Support for knowledge management and collaboration
- Thorough re-using valuable business processes, systems and technology
- By aligning change and innovation initiatives, through merging initiatives of the individual pillars into three consolidated initiatives on portal, search and content.

It is never wise to transform an entire organization at once. Therefore, it became obvious to focus on those activities that create the greatest synergies, efficiencies of scale and public value. The optimal focus was found by merging the access, dissemination, cataloguing and archiving processes of the pillars. For the time being, the publication production processes, the pillars itself, are left as is, to preserve the continuity and quality of all publications.

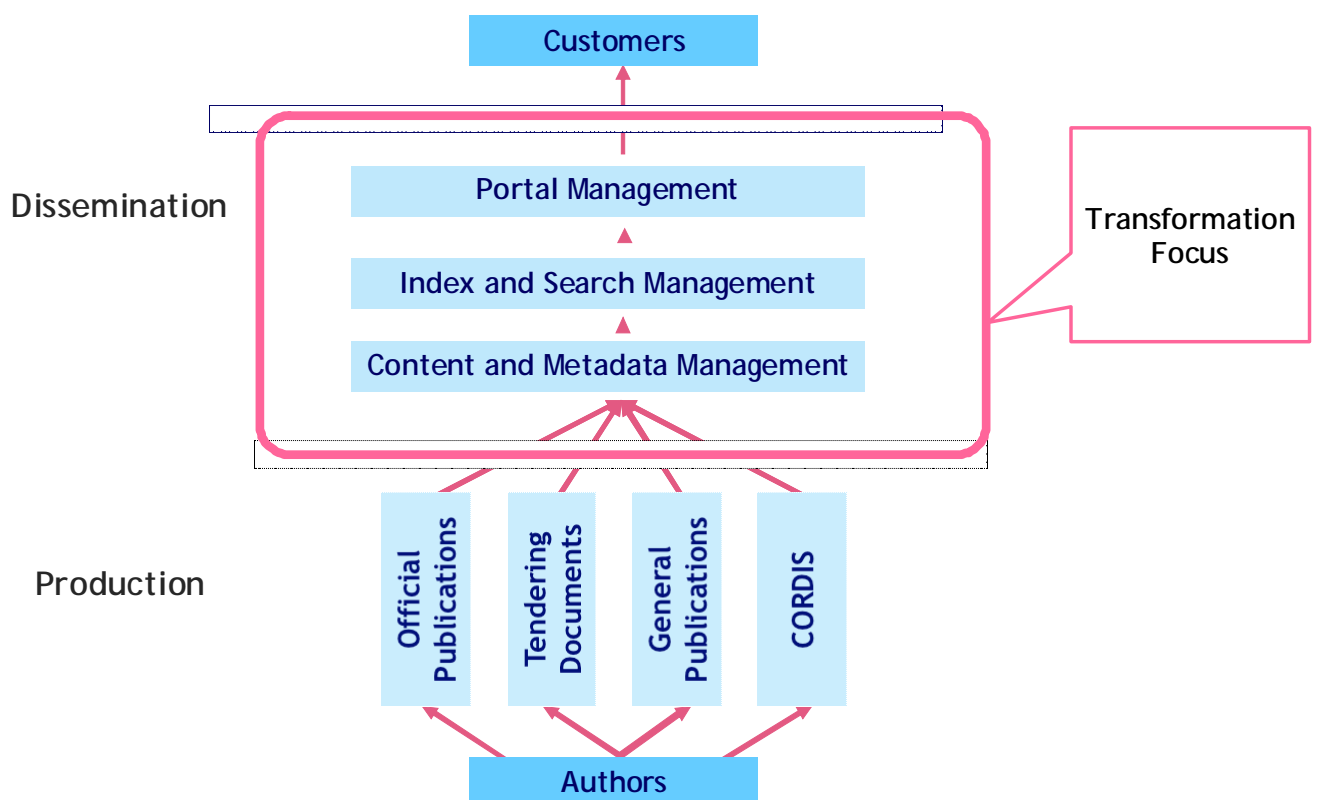


Figure 1. General overview of the Transformation Program

This figure can be regarded as the overall transformation architecture, for business and systems. Electronic versions of publications, resulting from production pillars, are made available (accessible or copied) to be used by the newly created access modules created from the merger of access initiatives of the individual pillars. This transformation architecture enables the harmonization of archiving, cataloguing, and electronic dissemination processes of the Office.

Horizontal layering (in short: 'Portal', 'Search' and 'Content') enables independent change per layer. This means e.g. that content formats and storage can be changed without affecting the search and portal layers other than enhanced functionality and services. Also, this architecture allows for multiple portals and multiple search engines and multiple content storage facilities to co-exist, thus enabling controlled, incremental and risk-averse migration.

The key mission of the Office is to provide access to law and publications. The horizontal layering enables the Office to respond easily and swiftly to changes in the demands of the environment and to provide continuous improvement of access to law and publications.

The Transformation Program is a collection of aligned projects that together deliver the desired results to reach the objectives. Alignment is ensured through governance organization and procedures. Major decisions are taken by the steering group headed by the director general and supported by an enterprise architecture board.

This Transformation Program aims to leverage and consolidate existing and projected initiatives to further enhance and innovate the access capabilities (Portal, Search and Content) of the publications office.

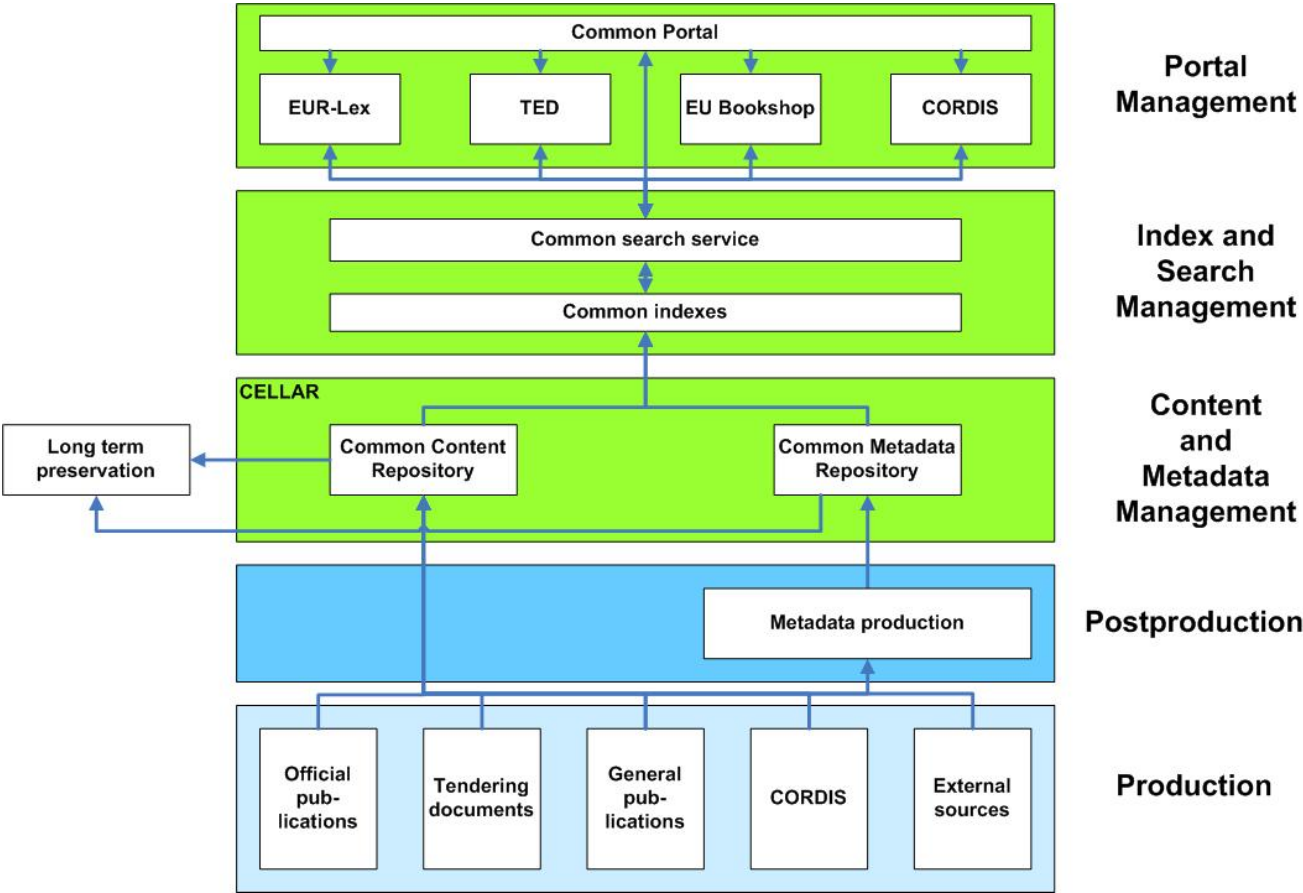


Figure 2. Transformation focus : from individual pillars to common access layers

As the new EUR-Lex project is an important application of the strategy of the PUBLICATIONS OFFICE, the following technical and functional specifications are to be fully in line with the directions of the Transformation Program.

## 2 Purpose of the document

---

EUR-Lex' mission is to become the public service of reference for the European Union's legal documents, the central access point for EU law and all other related documents.

EUR-Lex repository contains currently about 3.650.000 documents in 23 languages and receives up to 100 000 users search queries per day.

This document is the analysis of the requirements for the new conception of EUR-Lex.

The main reasons for a re-design and development of new EUR-Lex are:

- The transformation and development plan of the Information Architecture at the PUBLICATIONS OFFICE,
- The need for a better integrated and consolidated information base, in order to provide a complete and coherent portal for access to EU law and related documents (preparatory acts, case law, transposition into national law...);
- Improvement of usability and enhancement of functionality (This analysis is based on results of "Analysis of the existing system"<sup>1</sup> and of "EUR-Lex web usability assessment"<sup>2</sup>) :
  - Though, in general, the users strongly appreciate the content of the EUR-Lex website, most novices, and even some experienced users, do not intuitively find what they need: not enough help is offered on the site, the search is not quite satisfactory, either for functionalities or for results. The website "as such" is not perceived as a very usable one, according to the current standards.
  - Though main functionalities on legislative content are available in EUR-Lex, some have been developed in other applications:
    - The historical view (from a proposal to legally binding act published in the OJ) is available in PreLex,
    - N-Lex enables users to search through national legislative bases,
    - Users should connect to LexAlert to manage notifications on EUR-Lex content and get Web Service feature (currently not available to the general public.)
- These features should be better highlighted and much easily used when integrated in EUR-Lex new conception.
- Improvement of site management and contribution support:
  - Due to technologies used with EUR-Lex, the contribution and modification process involves many actors (editorial section, infrastructure section, external contractor, etc.) and therefore is very slow and heavy. Contributors should benefit of new functionalities to make edition and publishing of pages and contents easier and faster
  - The current back-office infrastructure for administration, maintenance, statistics, monitoring, diagnostic and basic data extraction tasks is quite light. The EUR-Lex team depends too much on other units (Communication Direction, external contractors, DG DIGIT, etc.).
- Improvements of reliability, robustness, performance and scalability of the system.

All these statements justify a new conception of EUR-Lex taking into account the requirements described in this document.

---

<sup>1</sup> Cf. PUBLICATIONS OFFICE\_10005\_DLX-0101\_V1.0.doc

<sup>2</sup> Cf. EUR-Lex web usability assessment\_Final\_v2.0\_en.pdf

### 3 Requirements overview

New EUR-Lex will offer:

- A better structured and richer Front Office, giving an easy access to all law resources, and providing advanced services to consult, search, interact and work with EUR-Lex content;
- The ability to access to:
  - Legal information, either processed by PUBLICATIONS OFFICE or managed by external sources
  - Editorial content produced by PUBLICATIONS OFFICE contributors
- A more robust and streamlined Back Office, allowing administrators to:
  - Organize site views, directories and content
  - Produce or access to editorial content
  - Manage back office processes in a better controlled way.

In the continuation of the document, the requirements have been organized following this operating view split into:

- Front-office features for end-users
- Back-office features for management activities

	EUR-Lex Front Office				EUR-Lex Back Office		
	Display	Search	Extra features	Personalisation / User account	Site Management	Administration	Identity Access Management
Portal	<ul style="list-style-type: none"> <li>• Navigate</li> <li>• Consult</li> <li>• Visualise</li> </ul>	<ul style="list-style-type: none"> <li>• Simple</li> <li>• Advanced</li> <li>• Expert</li> <li>• Results</li> </ul>	<ul style="list-style-type: none"> <li>• Social Networks</li> <li>• Help guides</li> <li>• FAQ</li> <li>• Guided tour</li> <li>• Feedback/ePolls</li> </ul>	<ul style="list-style-type: none"> <li>• Preferences</li> <li>• Syndication</li> <li>• Cart</li> <li>• Export (Web Services)</li> </ul>	<ul style="list-style-type: none"> <li>• Manage content</li> <li>• Manage content publication</li> <li>• Access to external sources of data (Business Data Catalog)</li> </ul>	<ul style="list-style-type: none"> <li>• Manage Statistics</li> <li>• Operate (supervision, checking, logging)</li> </ul>	<ul style="list-style-type: none"> <li>• Access Rights</li> <li>• Manage Profiles</li> </ul>
Index & Search	Index & Search Layer (Autonomy IDOL or similar)						
Content	EUR-Lex Specific Repository (specific editorial content, pages, css, xml, templates, etc.)		CELLAR (OP Official Documents of all kind: legal, publications, notices, etc. – type: documents, xml, video, images, metadata, etc.)				

## 4 Front office

---

### 4.1 Ergonomics

**HLR01.** EUR-Lex new conception should be an opportunity to redesign the “look and feel” following the conclusions of the report “EUR-Lex web usability assessment” published on may 2007, in accordance to the new corporate image to be applied to all websites managed by the Publications Office:

- Consistency in layout: header, menus, breadcrumb, pop-up management, etc.
- Consistency in visual identity: fonts, colours, graphical elements, etc.
- Consistency in action elements: buttons, hyperlinks, images, etc.

All textual parts of the web user interface of EUR-Lex's Front Office, due to their multilingual and synoptical nature, should be parameterized with the translation tables. Whenever a modification in the translation tables is introduced, the system will automatically update all templates with the new translated terms. In order to facilitate linguistic maintenance, the system will also provide a mechanism to manage the lists of terms for different parts of the web user interface.

”

**HLR02.** The new EUR-Lex website should also follow up-to-date WAI guidelines to make sure that it could be viewed, navigated and interacted with by anyone, particularly disabled people:

- Web Content Accessibility Guidelines (WCAG) 2.0 :
  - Perceivable
    - Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, Braille, speech, symbols or simpler language.
    - Provide alternatives for time-based media.
    - Create content that can be presented in different ways (for example simpler layout) without losing information or structure.
    - Make it easier for users to see and hear content including separating foreground from background.
  - Operable
    - Make all functionality available from a keyboard.
    - Provide users enough time to read and use content.
    - Do not design content in a way that is known to cause seizures.
    - Provide ways to help users navigate, find content, and determine where they are.
  - Understandable
    - Make text content readable and understandable.
    - Make Web pages appear and operate in predictable ways.
    - Help users avoid and correct mistakes.
  - Robust
    - Maximize compatibility with current and future user agents, including assistive technologies.
- Authoring Tool Accessibility Guidelines (ATAG) 1.0 :
  - Support accessible authoring practices
  - Generate standard mark-up.
  - Support the creation of accessible content.
  - Provide ways of checking and correcting inaccessible content.
  - Integrate accessibility solutions into the overall "look and feel".
  - Promote accessibility in help and documentation.
  - Ensure that the authoring tool is accessible to authors with disabilities.

It is true that WAI compliancy might considerably limit the browsing/visualization capabilities in the future EUR-Lex. The aim is to find a solution to make the site available to WAI users but without sacrificing usability.

## 4.2 Navigate

### 4.2.1 Home page

On the home page, the user should have a good overview of the website content and functionalities:

**HLR03. Sign in form.** PUBLICATIONS OFFICE wants to develop the user account. The login form should therefore be immediately visible for regular users to sign in, but also to incite occasional ones to create their own account (cf. 4.7 User account).

**HLR04. Get started.** EUR-Lex is a web site with very specific contents and rich functionality. Today, many occasional users have difficulties exploring its full potential, therefore some improvements in this area are necessary. (cf. 4.6.1 Get started).

**HLR05. Search through EU law.** Most users come to the website to search for legal texts. So, this functionality must be accessible on the home page.

**HLR06. News.** Current news is composed of a newsletter, thematic files and selection of new documents. It is not quite clear to the user: website news, though important, is not what one first think of when talking about news, and thematic files are more a content view than real news. News should only be legal ones (selection of new documents).

**HLR07. Recently published.** This feature automatically lists all documents (OJ, COM, SEC, Case-law, etc.) released recently.

**HLR08. Our services.** This part gives direct accesses to major website services:

- **Last published OJ.** This link leads to the Official Journal latest issues page (cf.4.3.1.1 Official Journal page).
- **Buy the OJ on CD-ROM.**
- **Get informed by RSS.** This link leads to the predefined notifications page (cf. 4.8.2 Syndication/notification).
- **Search on national bases.** This link leads to the national bases search page (cf. 4.5.4 National search).
- **Enter social network.** This link leads to the EUR-Lex' social network page (cf. 4.6.2 Social network)

**HLR09. Practitioner's corner.** This part gives direct access to extra-content practitioners may need.

**HLR10. What's new about EUR-Lex.** This part gives to PUBLICATIONS OFFICE the possibility to add editorial content on EUR-Lex new functionalities.

### 4.2.2 Header

The header should identify the website, but should also give direct access to common functionalities such as direct access to home page, search form, etc, provided that this configuration is not against the IPG.

**HLR11.** EUR-Lex' header, though already quite complete, should be updated in consideration of the new conception.

**HLR12.** It also should be taken into consideration that EUR-Lex is a multilingual website and the navigation through languages is one of the major functionalities. Users can easily change the pages' language from any website page and don't have to navigate again from home page.

### 4.2.3 Footer

**HLR13.** The footer gives direct access to contact and help tools.

### 4.2.4 Menu

Currently, there's only a first-level left menu which is not visible on all the pages (ex: advanced search, Official Journal, etc.).

A persistent menu is really important to reassure users on:

- Their ability to find what they are looking for: a well-organized menu gives an efficiency feeling.
- Their ability to discover what they are not looking for: a well-organized menu shows the website content and gives the user an idea on what he can find in it.

**HLR14.** This menu should put forward the different ways to access content:

- **Official Journal.** The Official Journal gathers legislative text depending on their publication date.
- **Collections.** The EU law texts are gathered in collections depending on their types: treaties, international agreements, etc.
- **Thematic files.** Thematic files are collections of legal documents on specific fields of EU activity (legislative instruments, preparatory acts, case law, parliamentary questions, etc.). The idea is not to be exhaustive but rather to group together the most relevant current legislation on a specific subject.

### 4.2.5 Breadcrumb

**HLR15.** A breadcrumb should enable users to go back with a one-click action to all the pages up in the information structure. It is also an effective way for users to find out where he is after navigating through many pages.

## 4.3 Views

### 4.3.1 Official Journal

The Official Journal of the European Union (OJ) is the only periodical published every working day in all official languages of the European Union (EU).

It consists of two related series (L for legislation and C for information and notices) and a supplement (S for public procurement). The C series have also an electronic section, known as the OJ CE, which contains documents published exclusively in an electronic way. There are also the specific series CA and LM and finally the Special Editions in the languages of accession countries (the Special Editions published on the occasion of the two last enlargements (2004 and 2007) are available on EUR-Lex).

#### 4.3.1.1 Official Journal page

From the home page of the Official Journal collection, the user should access to the following features:

**HLR16.** The display of OJ L and OJ C editions should be clearly separated and listed in chronological order.

**HLR17.** A dedicated OJ search will be available (cf. 4.5.2 Advanced Search)

**HLR18.** The latest editions should be listed and immediately accessible by a simple-click. action on edition's name. The number (ex: the 10 latest editions) or the time span (ex: the last week editions) of displayed editions should be defined in the application settings.

**HLR19.** Access to previous issues should be provided. Access to Special editions where available should be provided

#### **4.3.1.2 Table of content**

When selecting an Official Journal issue, the user will access the issue's table of content and the following features:

##### **4.3.1.2.1 Access an act**

**HLR20.** Each act should be accessible by a simple-click action via the table of content

##### **4.3.1.2.2 Latest issues**

**HLR21.** A link to the previous page (cf. 4.3.1.1 Official Journal page) should be displayed.

##### **4.3.1.2.3 PDF version**

**HLR22.** A link gives access to the PDF version of the whole OJ number. The Official Journal can be downloaded or viewed directly in the browser's window via the Acrobat Reader plug-in.

##### **4.3.1.2.4 Electronic certified version**

PUBLICATIONS OFFICE will produce in 2009 electronically certified authentic PDF versions of OJs. The details corresponding to the certificate should be accessible for each authentic document. Additionally, it might be necessary to make available for download the digital signature (in XML).

**HLR23.** Information regarding authenticity of documents should be clearly displayed, to allow the users to distinguish authentic documents from ordinary ones.

**HLR24.** The system should allow accessing the digital signature itself, to make it possible to verify the certificate on the clients' side (with a service or software independent from the EUR-Lex.)

##### **4.3.1.2.5 Notice**

**HLR25.** The OJ notice can be viewed from the table of content (cf. 4.4.3 Notice for information on notice feature).

## **4.3.2 Collections**

The EU law texts are gathered in collections depending on their types:

- Treaties
- International Agreements
- Legislation
- Preparatory acts
- Case-law
- Parliamentary Questions

**HLR26.** Collections should be accessible from a sub-menu.

**HLR27.** Each collection will have its own page similar – as much as possible - to the Official Journal page (dedicated search, latest issues list, etc.). The dedicated search of each collection will be detailed in the specifications.

**HLR28.** Other collections may be considered.

### 4.3.3 Thematic files

Thematic files are collections of legal documents on specific fields of EU activity (legislative instruments, preparatory acts, case law, parliamentary questions, etc.). The idea is not to be exhaustive but rather to group together the most relevant current legislation on a specific subject.

**HLR29.** Thematic files could be organized by chapters and subchapters according for instance to EUROVOC or Directory Code (subject area). A possibility to include only acts in force/no longer in force should be foreseen. It is recommended that all views are created dynamically query-based, using metadata at the time the user opens the view. However, possible performance issues should be carefully analysed.

**HLR30.** From the user's query, the system should display more convenient thematic files.

### 4.3.4 Other views

**HLR31.** The list of content views is not final yet. PUBLICATIONS OFFICE should be able to add or suppress some.

For example, a "Legislative procedures" view which enables users to get a list of all documents waiting for an opinion of the EESC or a list of documents waiting for an approval of the Council, etc.

## 4.4 Content consultation

Users will be able to access to contents by navigation through views or by search interfaces. This chapter describes the functionalities available when displaying a given content.

### 4.4.1 Display formats

Currently, the legal documents may be available in many formats (HTML, PDF, TIFF, Word, etc.). But there's no difference between them, except for the HTML rendition which may be visually different. Depending on the view, the display format may be different: PDF for Official Journal, HTML notice for treaties, HTML content for search results, etc.

HTML format is a very convenient format for website for the following reasons:

- It doesn't need extra plug-in or software to be displayed,
- It doesn't generate extra opening delay,
- It can easily be displayed in a website frame, which enables the use of the website menus and direct access to extra functionalities (ex: notice, multilingual display, etc.).
- Website style sheets can be applied

Unfortunately, it is not available for each type of content; neither do other formats like PDF, TIFF or Word.

The default formats can be as follows:

- Editorial content (search, navigation and results screens), notices, Treaties, etc. → HTML
- Legal texts linked to the notices → HTML, PDF, Word or TIFF.

Whenever a format is available for a document, a link towards this format should appear on the document page. At least, the formats available today should be proposed.

**HLR32.** The administrator should be able to set preferred display formats in the website preferences.

**HLR33.** Access to other formats may be proposed afterwards. In case a document is available only in a HTML format, the user might ask for an "on-the-fly" conversion to a PDF document, which could then be downloaded by the user.

## 4.4.2 URL

**HLR34.** Preferably all pages should be accessible by a unique and simple URL (including search, where the URL will include the search query). There should be unified syntax for URLs to access all types of contents in the system.

## 4.4.3 Notice

**HLR35.** On new EUR-Lex, the notice page should allow progressive displaying of different groups of metadata. The user's choice could be remembered for the next session.

## 4.4.4 Historical view

**HLR36.** Views of the "life-cycle" of a document should be provided, such as the one given currently by PreLex. PreLex follows all Commission proposals (legislative and budgetary files, conclusions of international agreements) and communications from their adoption through their transmission to the other institutions.

Links enable users to access directly available electronic texts (COM documents, Official Journal, Bulletin of the European Union, documents of the European Parliament, press releases, etc.).

This view could be completed with other subsequent acts: modifiers; corrigenda; consolidated versions; case-law affecting a specific item of legislation; national implementing measures of directives, etc.

## 4.4.5 Multilingual display

**HLR37.** EUR-Lex gives the possibility to compare two linguistic versions, i.e. to make both versions appear side by side on the screen, both documents having a similar structure, page layout, and being scrolled in parallel. However, some users want to compare more linguistic versions. Larger screens and higher screen resolution make it possible for at least three linguistic versions, so it should be available on new EUR-Lex.

## 4.4.6 Add to my cart

**HLR38.** Signed on users could add content to their cart, which should be saved for next visits. All contents displayed through lists (list of documents, of search results) may be selected, in a "select all" mode or manually, in order to add them to the cart. The "add to the cart" link appears only for signed users.

**HLR39.** The administrators should be able to define the number of bookmarks that can be stored in a user's cart.

## 4.4.7 Extend your search to

**HLR40.** Leaning on the notice metadata, some search extensions may be proposed to users. For instance, the user could extend his/hers search by:

- **The Eurovoc classification concepts.** By clicking on one of the content's Eurovoc concept, the user should perform a search on all contents matching the selected concept and be led to the simple search results interface.
- **Legal basis.** Search all acts which have a specific treaty article as legal basis; search all comitology acts; search all acts adopted in co-decision; etc.)
- **Document affected/concerned** (e.g.: all case-law related to a specific legislative act; national implementing measures of a directive; national case-law)

#### 4.4.8 Share link

**HLR41.** EUR-Lex should provide the ability to send by mail a link to a content page. This feature should be available to all pages in EUR-Lex to which access is not restricted by a user account.

### 4.5 Search service

The search engine should be able to perform a full text search in the content body, but also in certain metadata

The search will be performed by default on non editorial content. The extension of search to editorial content should be an option.

**HLR42.** Further functionalities should be available (cf. 7.1 Search state of the art):

- Boolean search. The search engine should enable users to easily, intuitively and precisely specify what they are searching (all words, any words, exact phrase, etc.)
- Search help:
  - Semantic and linguistic analysis. The search should be extended to semantically connected terms and to all linguistic versions.
  - Orthographic correction and phonetic search. The search engine should take into account hypothetical mistakes in writing the search terms.
- Search completion from:
  - The integration of Eurovoc multilingual thesaurus is to be considered. This thesaurus may be used to:
  - Control values of certain parameters used in search forms, to assure that these values match with the acknowledged vocabulary provided by the thesaurus. This vocabulary control should be applied to expert search queries building
  - Provide a reference to find semantically connected terms and linguistic versions
- History of previous searches performed
- Sort by relevance. From semantic, linguistic, statistic and structured analyses, the search engine should be able to evaluate the relevancy of results to a search, to display only results with a relevancy greater than a level defined by administrators and to sort them from the most relevant to the less one. It should be possible to configure relevance parameters like number of appearances in the text, position of the search term in the text (title, abstract, conclusion...), etc., as far as possible.

For the sake of rationality and efficiency, the Publications Office intends to adopt the same search engine for all its sites.

#### 4.5.1 Simple search

The simple search is an interface mainly oriented toward occasional users. So, it should be simple and intuitive to use.

**HLR43.** The Simple Search involves a search box, where one can enter search terms, numbers, and dates. The search is launched by pushing a "search" button

Simple search is available on the EUR-Lex home page, on the header<sup>3</sup> and on the dedicated search page. Simple search is available on every page giving access to legal content (e.g.: OJ; treaties, legislation in force, case-law, etc.)

The simple search should always be displayed together with links leading to:

- Advanced search
- Expert search
- A help file

**HLR44.** One may enter one or more query terms in the Simple Search box (ex: keywords, CELEX numbers, directory code, subject area, dates, etc.).

- The maximum limit of characters which can be used should be sufficiently high (e.g.: it should be possible to enter a sentence)
- The system should automatically correct wrong spelling (maj/min, accents, s/z), but it should be possible to disable this option. In case of wrong spelling, the system should propose alternative wording ("do you mean...?")
- The simple search scope should include as many metadata fields if possible, without reducing performance. The fields indexed should be at least :
  - Text fields: text, title, summary, keywords, etc.
  - Documentary fields (Eurovoc descriptor, Directory Code, Subject matter, publication reference, etc.)
  - Legal fields (Dates with legal effect, relationships between documents, Author, Type of documents, etc.).
- It will be possible to search for one or more exact expressions (e.g. using inverted commas).
- The system will be looking for all documents which include all the terms typed, regardless of their place and relative order in the document. E.g. when searching for 'swine influenza' the system will interpret the request as "swine AND influenza" and will deliver all documents which will include both terms.
- By default, terms should be entered in the language of the interface (to avoid "noise"), but there should be an option to allow searching in a language other than that of the interface, or even in all languages. For example, an option may be to extend the search to other languages if no result is found in the language of the interface.

## 4.5.2 Advanced search

The advanced search enables users to perform a multicriteria search.

**HLR45.** The "Advanced search" link will open the **Advanced Search** screen. The Advanced Search form includes extra-criteria

- **Search terms:** further requirements should be added (possibility to use Boolean operators, truncation, automatic correction, etc)
- **Date:** the user can filter search with different dates (creation date, publication date, document date, etc.)
- **Author:** the author is the institution that has created the document (Commission, Parliament, etc.)
- **Category/domain:** treaties, legislation, preparatory acts, etc.
- **Sub-category:** Treaty on European Union, Accession Treaties, etc.
- **Keyword/subject matter:** this can be the Eurovoc's classification, the chapters of the directory of legislation, the chapters of the Digest of case law or the subject matter classification used in EUR-Lex

---

<sup>3</sup> as long as it is not against the IPG

- **Language:** The user can choose the language of the documents he wants to search for. By default, it is the navigation language.
- **Number:** this search can be extended to all types of documents, provided that a field “number” is available
- **CELEX identifier**
- **Publication reference**

**HLR46.** It should be possible to combine all search options offered as much as possible without downgrading performance

When passing from the “Simple search” to the “Advanced search”, terms used in the Simple search should be kept, in order to allow the user to refine his search.

**HLR47.** The Advanced Search will provide also a “specific searches” mode: specific search forms will be dedicated to each collection of documents in order to provide advanced search on these collections with relevant criteria search fields. Moreover the search can be limited by type of documents.

- Official Journal
- Collections
  - Treaties: the search will be performed on treaties articles, protocols, annexes, etc.
  - Regulations, directives, decisions, etc
  - National implementing measure: currently the search is performed taking the numbers of the directive as the basis
  - Case-law: case number
  - COM and SEC documents
  - EP question

**HLR48.** A link to “notify me” page should be available from the search result page. All users may have access to a notification service using pre-defined RSS feeds.

**HLR49.** Category selection, standard filtering and sorting options are available for results.

### 4.5.3 Expert search

The expert search could be available to signed on users and will allow constructing and combining complex search queries using all available metadata, visualizing and exporting the results in a very flexible way.

Currently, the expert search (called “advanced search”) is an independent application. It should be redeveloped to be fully integrated in a new EUR-Lex interface.

**HLR50.** The **Expert Search** form works in a similar way as the Advanced Search form, but enables :

- Free text entry using Boolean operators (AND, OR, NOT, AND NOT, etc.)
- Building user specific queries, The user may :
  - Choose any field available through the documents metadata,
  - Enter values attributed to query fields. The values should be chosen into a choice list, whenever the case, or in the EUROVOC vocabulary terms, for keywords values.
  - Assemble these conditions with operators to build complex queries.

**HLR51.** An sign-in user may save his queries or access them as notifications.

**HLR52.** Category selection, filtering and sorting options are available for results pages.

### 4.5.4 National search

**HLR53.** The National Search interface enables users to search through national bases.

This feature is currently available through an experimental website: N-Lex. N-Lex enables users to search in the national bases of 23 UE countries through local or distant connectors (in the future, all connectors should be distant ones).

## 4.5.5 Search results

### 4.5.5.1 Display

**HLR54.** The administrator should be able to set the minimum relevance of displayed search results, depending on the capabilities of search engine.

**HLR55.** The result screen should include :

- **The query.** The search form used by the user to perform his query, filled with the user's query settings should be immediately accessible, for the user to be able to perform at once a new search.
- **The number of results** corresponding to the query
- **The list of search results** (x number of results by default)
- **Facets** allowing for further refinement of results (depending on type of search performed)
- Further options

**HLR56.** For each search result, the following information should be displayed and could be sorted up or down :

- Identifier (CELEX number)
- Title
- Publication reference (e.g. OJ publication number – when available)
- Publication date
- Document type
- Links to :
  - Available text formats (when available)
  - Last consolidated version (when available)
  - Bibliographic notice
  - Life cycle view
  - Multilingual display

The precise list of information depends on the type of search and should be defined later by the PUBLICATIONS OFFICE

**HLR57.** For text search, a dynamic teaser could also be displayed with the terms of the query highlighted. This generated summary field enables presentation of a document extract in context with the search query.

**HLR58.** Certain categories of documents could be excluded by default from the search results display of a query. (e.g.: corrigenda, previous consolidated versions) - unless the user decides otherwise. The system could provide by default access to the basic act and its most recent consolidated version..

### 4.5.5.2 Facets

**HLR59.** Besides search results, some facets (cf. 7.1.2.1) will be displayed to refine the query: for instance, keyword, EUROVOC category, subject area, author, etc. Facets displayed will vary according to the search performed

The following elements should be displayed:

- The number of results for each facet's item
- Only facet's items with results
- The complete foldable taxonomy tree

### 4.5.5.3 Export results

**HLR60.** Users wanting to export a query result set for external use should specify:

- Fields to be exported (title, date of publication, type of document, etc.)
- The export format (MS Excel, MS Word, XML, etc.)

**HLR61.** Predefined export templates may be available.

## 4.6 Extra features

### 4.6.1 Get started

**HLR62.** This section, available from the home page, is intended for giving to new or occasional users all the tools to use EUR-Lex, such as:

- Site content
- Guides
- FAQ
- Guided tour (online demonstration of how to use EUR-Lex)
- Site map
- User account creation

### 4.6.2 Social network

**HLR63.** EUR-Lex' social network should be a place where users ask, consult and answer questions in a "Yahoo! Answers" way. Though every user could consult questions, only registered ones could ask or answer them. The topics of the social network will be oriented toward EUR-Lex structure, content and use.

The Office is currently studying the possibility of using the same social network tool for all the sites it manages.

**HLR64.** A dedicated search interface may be displayed:

- **Questions type:** all questions, open question or resolved questions.
- **Themes:** questions may be gathered in themes like "EUR-Lex", "Law", etc. Eurovoc's themes might be used.

**HLR65.** The multilingualism issues should be taken into account (automatic translation, language dedicated forums, etc.).

## 4.7 User account

**HLR66.** An signed-on user may save his queries.

## 4.8 Users' feedback

The new EUR-Lex should allow receiving feedback from its users for different elements of the portal, in a form of surveys, polls, suggestions, comments, ratings, etc.

## 4.8.1 Preferences

### 4.8.1.1 Languages

**HLR67.** EUR-Lex is a system that has to be available in all official languages of the EU. All linguistic renditions are treated as equal, but there is a need to handle situations where not all renditions are available (either temporarily or not). Though default languages may be defined in settings, the user can set preferences in his user account.

### 4.8.1.2 Search results

**HLR68.** Search results display preferences could be set in the user account:

- Number of documents per page,
- Default sort criteria,
- Displayed fields.

### 4.8.1.3 Notify me

For signed-in users to be regularly informed of new documents matching a query, the “notify me!” button leads to the personalized notifications creation form filled with the query settings. News

**HLR69.** The idea is to personalize in some way the user’s home page, for example by displaying his notifications’ results in place of the default news.

## 4.8.2 Syndication/notification

**HLR70.** A signed-in user should save queries of any type (simple, advanced, expert), re-play them or use them as notification filters : the user will be notified when a new result appears.

**HLR71.** Current application LexAlert should be replaced by an integrated functionality allowing for:

- Predefined notifications
- Personalized notifications
- Notifications created from search queries

**HLR72.** The notification delivery format might be RSS, XML or e-mail. It is recommended to integrate the OpenSearch standard..

## 4.8.3 Cart

**HLR73.** A personalized cart will enable users to perform further actions on contents they have selected along their navigation. This cart offers the following features:

- A cart is attached to a signed user
- The cart owner may select documents (documents list, search results list) and add them to the cart
- The cart content is persistent and can be retrieved at the next session.
- The cart owner may perform actions on the content of the cart: display, print, send link to, empty, etc.
- The user could receive a notification whenever any modification is introduced in the notice of a particular act (or set of acts).

## 4.8.4 Exports: web services for re-users

EUR-Lex will offer a possibility to export content as a data flow delivered remotely to another application. It will include at least the complete advanced search functionality.

The list of data flow formats will be precised. It may include:

- RSS feeds
- XML
- Other formats (e.g. CSV; Excel, TVS, etc.)

Exports formats should facilitate re-use of documents and metadata.

For performance reasons, the volume of data sent at a time could be limited by administrators.

## 4.9 Portability

In recent years, there has been a proliferation of types of device and access mechanisms using the Web, extending far beyond the conventional personal computer. These access mechanisms range from web tablets, appliances and TVs in the home, to mobile devices including phones and PDAs, and access mechanisms for the physically challenged. Connectivity capabilities have also evolved to include high bandwidth modems, LANs and wireless networks.

Simultaneously, the needs and expectations of the user with regards to access, availability and consumption of web content have also evolved. Users now expect to get to critical information through different access mechanisms from different locations and at different times during their day.

**HLR74.** We should then consider setting up a mobile version of EUR-Lex. It could be a light version of the official EUR-Lex website proposing the top features like search or browsing the HMTL version of the OJ for example. This requirement shall be considered optional.

## 5 Back office

---

The new EUR-Lex will offer access to 2 types of contents:

- Legal content: this content includes document files and metadata produced by PUBLICATIONS OFFICE other applications (e.g. EUDOR, PROCAT) or external sources (e.g.: N-Lex). EUR-Lex must provide access to the documents and data managed by these applications, and must be able to connect to them in order to display the content. These applications and sources are not managed within EUR-Lex Back-Office.
- Editorial content: this content is managed by EUR-Lex administrators in order to explain, enhance, ease and complement the access to the legal content. It is managed within EUR-Lex Back-Office. The content published through the contribution process is currently limited to web oriented content.

In any case, Legal and Editorial content should be distinguished when appearing in the Front Office pages, as directory content or as search results

In the Back Office should be distinguished:

- Features allowing the administrators to organize pages, directories, and associate content under various forms (content layout, lists, etc.) to pages. These actions can be performed either on legal content or on editorial content (“Site structure management”)
- Features allowing administrators to feed EUR-Lex with content and data. These features may be applied to :
  - Content generated through the PUBLICATIONS OFFICE production chains (“Management of content provided by production chains”)
  - Content from external sources (“Management of content provided by external sources”)
  - Editorial content generated by EUR-Lex contributors (“Web oriented contribution”)
- Administration features

**Regardless the requirements described in this chapter, it is obligatory to take into consideration all the principles of the PUBLICATIONS OFFICE's Transformation Program (see chapter 1.)**

### 5.1 Site management

#### 5.1.1 Content elements

The new EUR-Lex Back-Office should allow administrators to define site contents and site structure. The site structure is the result of the design of all content elements.

**HLR75.** Content elements include:

- A page of content (“Process and Players” for example)
- A section in a page (Introductory Text of a section for example),
- A subset of a page (News on the Home Page, ...),
- A group of internal links to content within EUR-Lex in order to create a « Thematic files »,
- RSS feeds
  - from other sites (www.europa.eu for example)
  - from EUR-Lex to re-users of the EUR-Lex content

**HLR76.** Back office contribution includes management and publishing of links to:

- contents already published on EUR-Lex (or about to be published),
- Other web sites relevant for EUR-Lex.

#### 5.1.2 Site structure management

EUR-Lex Back-Office should allow content administrators to:

**HLR77.** Add a content element at a determined place in the site structure. Available content elements will be part of the Back-Office; Creation of new content elements by administrators should be guided as far as possible.

**HLR78.** Choose a layout for this element, and apply it to a content element; Available layouts for each content element will be part of the Back-Office

**HLR79.** Position a content element in the hierarchy of the site structure. For example: sections in a page, sub-directories in a directory, etc.

**HLR80.** Modify / delete a content element.

### 5.1.3 Multilingualism

This is one of the most important requirements for the future EUR-Lex system.

**HLR81.** EUR-Lex must support all official languages of the UE. This support could be seen through two views:

- The management of the linguistic versions
  - Use of UNICODE in order to handle the different character sets of the official languages and possible further languages.
  - Display of a default linguistic version if the wanted version of a document doesn't exist,
  - Indexation of linguistic versions of an existing document, whenever these versions are produced for the 23 languages of the UE,
  - Similar behaviour in the site for the different linguistic versions of a document if they exist
- The multilingualism of the site itself
  - Equal rendition of the website whatever the language is (if the documents exist). This includes the ability to perform searches in each language and access to return linguistic versions in this language. Extended functionalities may be proposed; even in they are not supported equally for all UE languages.
  - Equal search results<sup>4</sup> when a user performs a query between the different linguistic versions of the site. )

The system should remember (i.e. in a cookie) what languages the users used previously and to propose the same set of languages for the next visit. Actually, this feature could concern all elements of the user interface which can be customised.

**HLR82.** The new EUR-Lex should be capable to handle Digital Object Identifiers (DOI).

## 5.2 Content management

### 5.2.1 Management of content provided by production chains

#### 5.2.1.1 Regular content management

##### 5.2.1.1.1 Store/ Import/Flow management

Legal documents (OJ, COM, SEC, etc.) and associated metadata will be in the future managed through the new EUDOR application, metadata being managed with the new PROCAT. From a

---

<sup>4</sup> the linguistic versions of the same documents must be returned through searches in different languages, provided that the linguistic versions exist, at least if the query is not dependent on natural language (for instance : query on CELEX number

functional point of view, these applications will provide access to all content (documents, metadata) that is used, displayed and indexed by EUR-Lex.

**HLR83.** From an architectural point of view, EUR-Lex should offer a dynamic caching capability in order to improve performance.

**HLR84.** Caching strategy should be configurable (basic parameters: time, frequency of access, etc.)

**HLR85.** The new EUR-Lex should be able to provide access to distributed data and metadata sources. Upload has to be as much automatic as possible, manual intervention should be possible in order to correct problems or for punctual interventions but has to be in the domain of exceptions.

#### **5.2.1.1.2 Updating processes**

Legal content and data is managed, either in PUBLICATIONS OFFICE systems, or in external sources systems.

Concerning legal content management EUR-Lex should allow content administrators to:

- Define settings for selecting categories of metadata from the future PROCAT and the corresponding documents from the future EUDOR.
- Define settings for updating, caching and indexing of contents.
- Perform certain exceptional manual operations.

#### **5.2.1.1.3 Exceptions management**

In addition to automatic updating processes, it must be possible to handle manually some exceptions, and particularly the removal of documents, for instance a primary document, or an OJ edition. Exception handling has always to be logged in order to remain traceable.

#### **5.2.1.2 Management of early publishing**

The future EUR-Lex system will have to provide an early access to a limited volume of data (early publication of the OJ for example without the complete notice).

#### **5.2.1.3 Indexation**

**HLR86.** The indexation process must at least deliver the same level of services as today. It should index the documents (or the text indexes) in full-text but also their metadata. The aim is to facilitate the information retrieval process to the end user.

Indexation of new documents should be incremental, and as close as possible to real time indexation without downgrading performance.

The management of the search engine and indexing, as a highly specialised technical task, could be done in the Back Office at the system administrator level.

#### **5.2.1.4 Management of certified documents and authenticity**

PUBLICATIONS OFFICE will produce in 2009 electronically certified PDF versions of OJs.

The management of PKI and signatures will be implemented as an independent system. The new EUR-Lex will make XML signatures accessible to the user, as well as the relevant metadata specifying the details of the publication.

#### **5.2.1.5 Tracking document feeding**

**HLR87.** Again, management of document content and management of PUBLICATIONS OFFICE applications output (Mainly Eudor and Procat) are not in the scope of this back-office, but:

- Back-office should track and get a picture about the information feeding the portal, in terms of files, pre-notices, notices, TOC (Table Of Content), ...
- Back-office should track and get a picture of the feeding process in terms of file/data transfers, transfer/exchange duration, detection of possible queues problems

## 5.2.2 External content management

External content (e.g. N-Lex) would be accessed to and searched through EUR-Lex, but managed and stored externally. EUR-Lex should be scalable so that it can be connected to future new external content or data sources.

### 5.2.2.1 Deep linking

**HLR88.** This feature will allow creating links to external contents, permanent links, and links to contents different from documents.

It will also provide a way to navigate within EUR-Lex content by displaying the reference links inside a document referring to other documents or to a set of documents via its metadata, and allowing navigation through these links. The system should be able to detect erroneous links.

## 5.2.3 Web Oriented Contribution (editorial content)

Contribution means the ability to easily create, update and publish contents onto the EUR-Lex portal ("a EUR-Lex portal content") for a functional user/contributor without any technical skill.

Back office contribution excludes creation and management of documents (refer to analysis of existing system/document types, analysis of existing system/editorial feeding process) and metadata associated to documents. In fact, all the documents (OJ, Treaties, preparatory acts...) are managed in other systems (PUBLICATIONS OFFICE applications, or other sources) and only published in EUR-Lex.

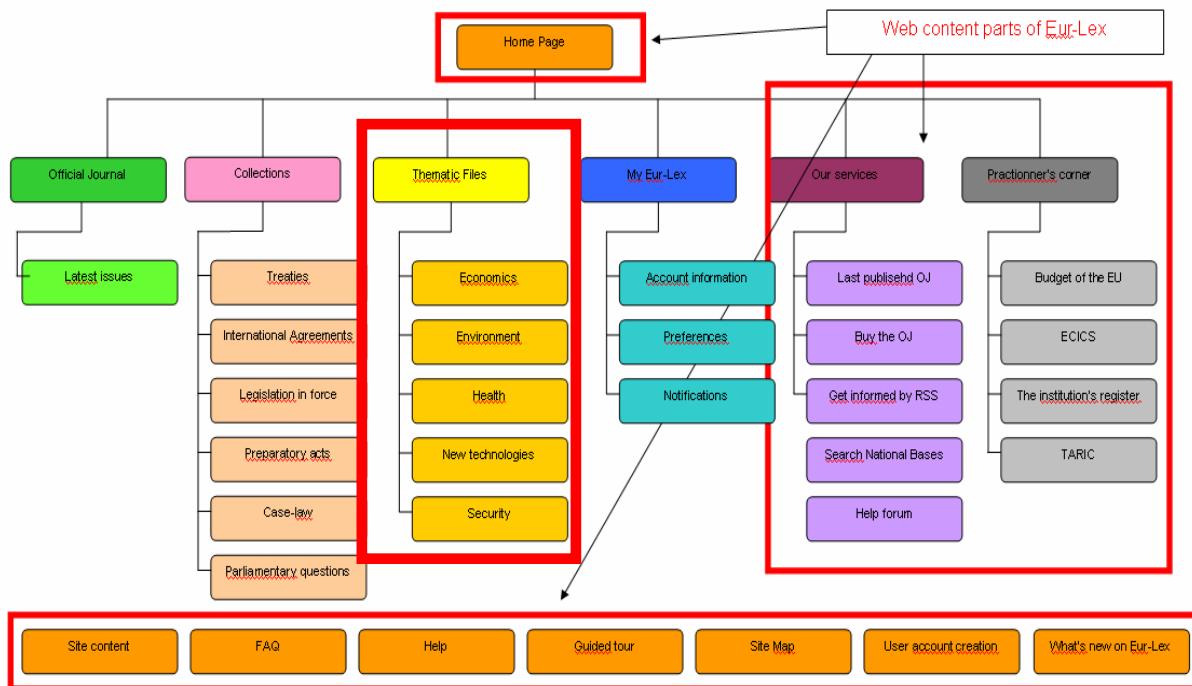
Back office contribution may also take into account the management of specific metadata associated to a EUR-Lex portal content<sup>5</sup>. Those metadata shall be treated by search functionality like any metadata associated to documents regardless to the source of information (Eudor or EUR-Lex).

N.B.: For the sake of rationality and efficiency, the Publications Office intends to adopt the same web content management system for all its sites.

---

<sup>5</sup> Specific key-words associated to a FAQ item, publication date of a static page, ...

### 5.2.3.1 Possible EUR-Lex Site Map and contents



The paths to the editorial contents are:

- Home Page
  - Site Content
  - Guided Tour
  - Help
  - FAQ
  - What's New on EUR-Lex
  - Site Map
  - User account creation
- Our Services
  - Last published OJ
  - Buy the OJ
  - Get informed by RSS
  - Search National Bases
  - Social network
- Practitioner's corner
  - Budget of the EU
  - ECICS
  - The institutions' register
  - TARIC
- About EU law
  - SCAD Plus
  - Process and Players
  - Legislative drafting

For each section of editorial content, the type of contents can be different (News, links, etc.). The next section identifies all this type of contents.

### 5.2.3.2 Content Models

What is interesting with a content management system (CMS: refer to chapter 7.2) is that the content model and the content itself are completely separated. You can fill a simple form in your back-office and have a completely different rendition on your front-office site.

Some metadata can be used to define the position of content on the site (heading, subheading, etc.) or to provide information to the user when accessing the content. Others are used to manage its life cycle (creation date, status, author, validator, etc.)

The content can consist of simple text or shaped (rich text), files, links, images, videos, etc.)

Thus, we have to think about the different content models concerning each type of contents. Here are two examples of what the content models could be:

- The content model of a “News” could be:
  - 1 Title (text) → mandatory
  - 1 Introduction text (rich text) → mandatory
  - 1 Main text (rich text ) → mandatory
  - 1 Link → optional
  - 1 image/video/flash content → optional
  - 1 Publication date (by default the current day) → mandatory
  - 1 Withdraw date → optional
  - Tags (text) → mandatory
- The content model of a “thematic file” could be:
  - 1 Title (text) → mandatory
  - 1 Introduction text (rich text) → mandatory
  - Link 1 (to OJ, treaty...) → mandatory
  - Link 2 → optional
  - Link ... → optional
  - 1 Publication date (by default the current day) → mandatory
  - 1 Withdraw date → optional
  - Eurovoc Tags (text) → mandatory

### 5.2.3.3 Back-Office Structure concerning the editorial parts

The administrators/contributors of the site will need a simple structure and simple access to the contents. Thus we can propose to align the back-office and the front-office architectures in order to have a clear view of all the contents and the sections they belong to.

Here is an example of architecture focusing on the “Practitioner's corner” section:

1. At first, the user sees all the sections he can modify in regard to his access rights

#	<input type="checkbox"/> Nom de la Section	Publiée	Réorganiser	Trier	Accès	ID Section	# Catégories	# Actives	# Corbeille
1	<input type="checkbox"/> OJ ( OJ )			1	Public	5	0	0	0
2	<input type="checkbox"/> Collections ( Collections )			2	Public	1	1	6	0
3	<input type="checkbox"/> Thematic Files ( Thematic Files )			3	Public	2	1	4	0
4	<input type="checkbox"/> My Eur-Lex ( My Eur-Lex )			4	Public	3	1	2	0
5	<input type="checkbox"/> Our services ( Our services )			5	Public	12	0	0	0
6	<input type="checkbox"/> Practionner's Corner ( Practionner's Corner )			6	Public	13	0	0	0
7	<input type="checkbox"/> Site Content ( Site Content )			7	Public	14	0	0	0

<< Première < Précédente 1 Suivante > Dernière >>  
 Éléments par page 30 Résultats 1 - 7 de 7

Figure 3. Main sections

2. When the user clicks on a section, for example “the practitioner's corner”, he sees all the subsections:

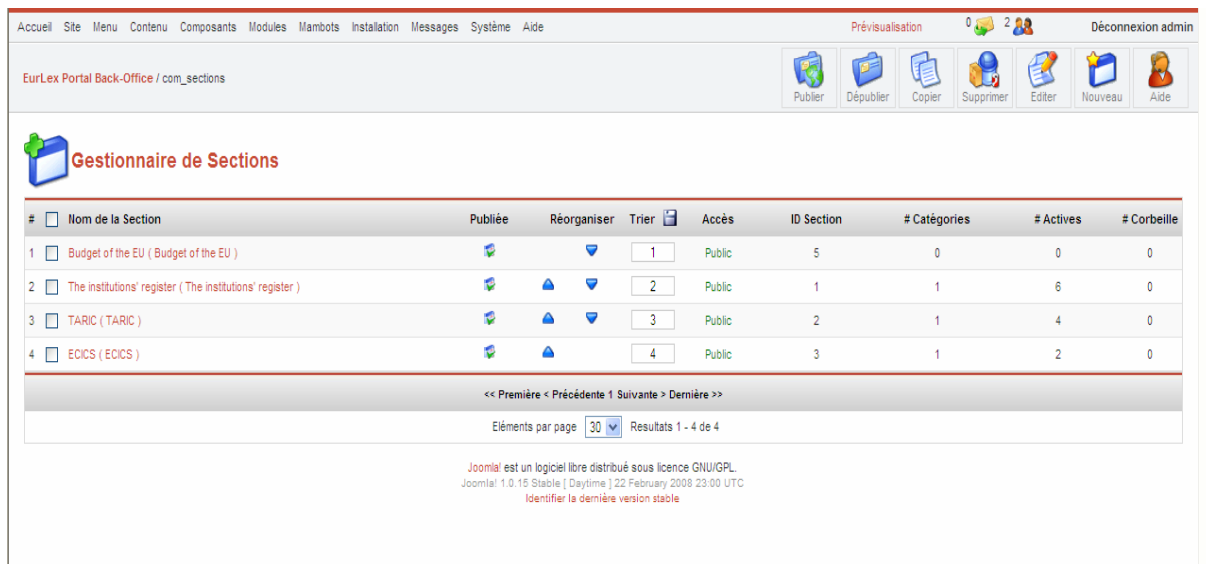


Figure 4. Sub sections of the “Practitioner’s corner”

- Finally, when he clicks on the subsection, he can see all the contents within this subsection and then create, modify or delete them.

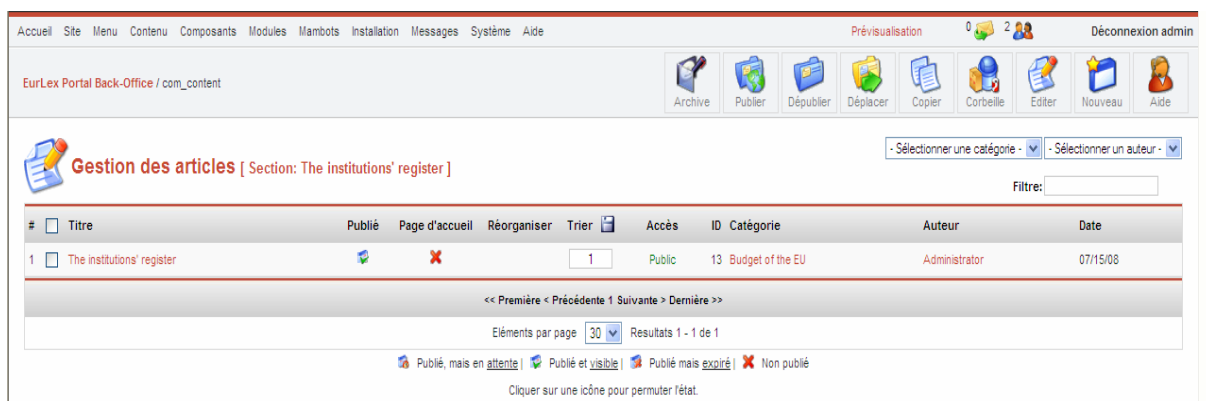


Figure 5. Subsection’s content display

#### 5.2.3.4 Contribution process

**HLR89.** The contribution process is composed of :

- The creation and management of the EUR-Lex portal content,
- The edition of this content,
- The insertion/relocation of this content within the portal,
- The publishing services, which enable to push one or multiple contents onto the diffusion site. This service will propagate changes onto the EUR-Lex portal in use.

**HLR90.** The contribution process shall be completed by:

- Preview facilities in order to get the exact picture of the modified portal (**WYSIWYG** : What You See Is What You Get),
- Workflow involving translation into different languages,
- Synoptic validation,
- Optional control in order to easily insert a validation workflow before publishing,
- Undo facilities,
- Publication schedule, programming the duration of a publication,

**HLR91.** Contribution shall be performed very easily and quickly, merely by the team actually in charge of the EUR-Lex editorial process. Specific management rights for the contribution process shall be available with specific rights and habilitations for the different services (publication, creation, validation, etc.).

**HLR92.** The publishing services will include a staging area where validated contents can be thoroughly tested before publishing.

### 5.2.3.5 Creation/edition

**HLR93.** As stated previously, creation and edition of contents shall not require technical skills: forms could be filled in order to write data and metadata. It shall also be possible to reuse files produced by office tools.

**HLR94.** During the creation process, it shall be possible to browse content about to be published (or already published) in order to collect content identifiers/links.

**HLR95.** When the user browses the content of a section or a subsection, he can:

- Create a content,
- Modify a content,
- Move a content,
- Publish/Remove a content from Front Office publication. Removed contents will be kept in the Back Office as long as they are not deleted.
- Copy a content,
- Delete a content from the Back Office..

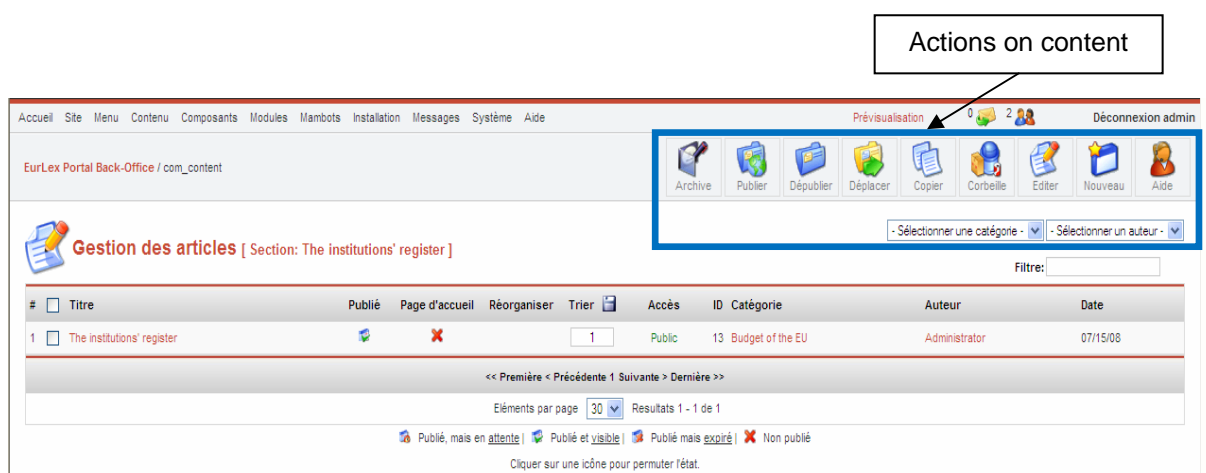


Figure 6. Subsection's content display

In order to edit content, the user just has to click on the selected content. As explained before, the content is defined by its metadata and by the content itself. The example below shows the content model for a News article with the metadata defined on chapter 5.2.3.2.

- The user has to fill the form with data and then publish the article.
- The content can be published directly or require validation (it depends of the category of the content). When validation is needed, a pool of moderators could be alerted (for example, by e-mail or through a specific dashboard).
- One of the moderators validates the contents after possibly complemented.
- The content is then either released or published directly to a default date when created.
- At the end of life, the content can be unpublished, either automatically or manually by the author of the content or a moderator which have rights on this content, before being deleted or archived.

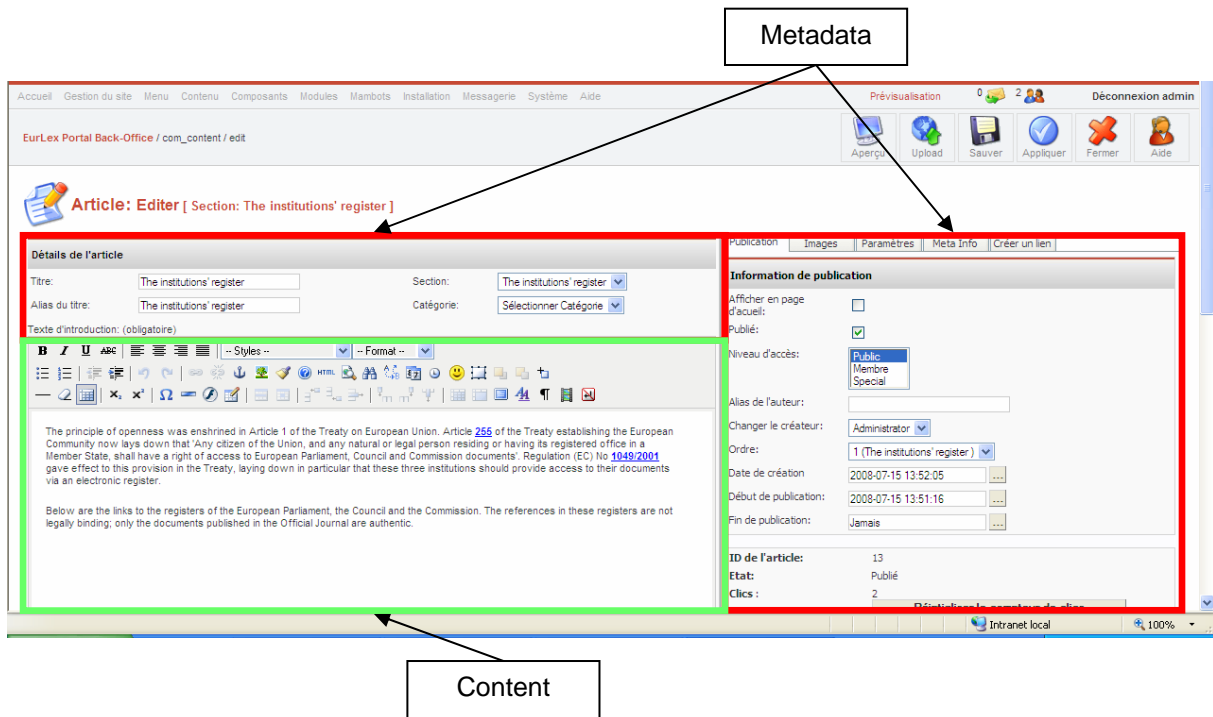


Figure 7. Edition of a content article

As stated in the chapter 5.2.3.4, the user has preview facilities called WYSIWYG and can directly see from the Back-Office the final rendition of the content on the Front-Office.

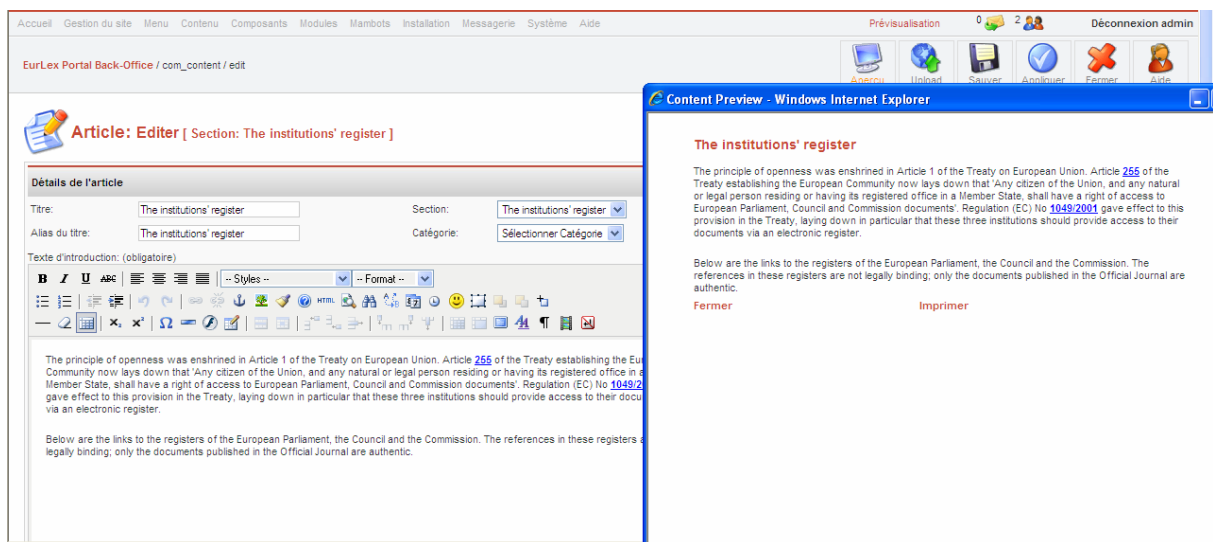


Figure 8. Preview of an article

## 5.2.4 Migration of present EUR-Lex content and data

The future system should provide tools helping PUBLICATIONS OFFICE to import the existing EUR-Lex contents, files and data into the future system.

It is foreseen that the current accounts and setting will be migrated to new EUR-Lex. In particular the LexAlert notifications and saved Expert search queries and settings should be migrated.

For compatibility until some time, the new EUR-Lex should provide a mechanism to map old links into new ones, whenever possible. In case the present EUR-Lex and the future EUR-Lex should co-exist during a period of time, it must be noticed that there is a need to ensure the creation of permanent links.

## 5.3 Administration

### 5.3.1 Users management

User's management includes:

- Management of Front Office users,
- Management of Back office users.

#### 5.3.1.1 Front Office users

There are 2 kinds of Front Office users:

- Non signed users : they may access all unrestricted content and features, they do not need to be identified in EUR-Lex,
- Signed users: they access all unrestricted content and features, and also user account content and features. These users will have to be identified and authenticated to access these services, EUR-Lex should propose a system for managing and authenticating Front Office users, as well as groups of users and associated rights.

The user registration and authentication system will be implemented in a way that it could be re-used for other portals managed by the PUBLICATIONS OFFICE, as a common SSO platform.

#### 5.3.1.2 Back-Office /editorial users

Back office users belong mainly to the editorial team. Back Office users' management addresses the management of access rights for back office services with different roles.

Certain roles can be foreseen (non exhaustive list):

- Administrator : Management of the users and Access Control List
- Author : Profile that creates contents (Textual part)
- Editor : Profile that validates contents (Textual part)
- Publisher : Profile that publishes contents (Textual + Visual front-end part )
- Translator
- Etc.

Note that, for example, a user can be author and publisher in the same time.

Hereunder the dashboard of the administrator profile, he can add/edit/create profiles:

Accueil Site Menu Contenu Composants Modules Mambots Installation Messages Système Aide Prévisualisation 0 1 Déconnexion admin

EurLex Portal Back-Office / com\_users

Annuler Supprimer Editer Nouveau Aide

Modifications enregistrées pour: PascaleDELL

Gestion des utilisateurs

Filter: - Sélectionner Groupe - - Sélectionner par Statut de connexion -

#	<input type="checkbox"/>	Nom	Identifiant	Connecté	Activé	Groupe	E-Mail	Dernière visite	ID
1	<input type="checkbox"/>	Administrator	admin	✓	✓	Super Administrator	belhiti@yahoo.com	1999-11-29 23:00:00	62
2	<input type="checkbox"/>	TomaszSLAWICZEK	TSLAWICZEK	✓	✓	Super Administrator	Tomasz.SLAWICZEK@europa.eu	1999-11-29 23:00:00	63
3	<input type="checkbox"/>	MariaManuelaCRUZ	MMCRUZ	✓	✓	Administrator	MariaManuelaCRUZ@europa.eu	1999-11-29 23:00:00	64
4	<input type="checkbox"/>	AndreaBARTOLINI	ABARTOLINI	✓	✓	Author	AndreaBARTOLINI@europa.eu	1999-11-29 23:00:00	65
5	<input type="checkbox"/>	GeorgianaTIVDA	GTIVDA	✓	✓	Editeur	GeorgianaTIVDA@europa.eu	1999-11-29 23:00:00	66
6	<input type="checkbox"/>	PascaleDELL	PDELL	✓	✓	Publisher	PascaleDELL@europa.eu	1999-11-29 23:00:00	67

<< Première < Précédente 1 Suivante > Dernière >>

Éléments par page 30 Résultats 1 - 6 de 6

Figure 9. Management of the users

When editing or creating a user, the administrator will have to set different information such as:

- The name of the user,
- The e-mail of the user,
- The login and password of the user,
- The Group and habilitations (Author, Editor...)
- Etc.

Accueil Gestion du site Menu Contenu Composants Modules Mambots Installation Messagerie Système Aide Prévisualisation 0 1 Déconnexion admin

EurLex Portal Back-Office / com\_users / new

Sauver Appliquer Annuler Aide

Utilisateur: Ajouter

Détails utilisateur

Nom: Tomasz SLAWICZEK

Identifiant: TSLAWICZEK

Email: Tomasz.SLAWICZEK@publications.europa.e

Nouveau mot de passe: \*\*\*\*\*

Confirmer le mot de passe: \*\*\*\*\*

Groupe:

- Public Frontend
- Registered
- Author
- Editor
- Publisher
- Public Backend Manager
- Administrator
- Super Administrator

Bloquer utilisateur  Non  Oui

Paramètres

Editeur Utilisateur - Sélectionner un éditeur -

Information liée au contact

Aucun contact lié à cet utilisateur.  
Voir 'Composants -> Contact -> Gérer les Contacts' pour les détails.

Figure 10. Creating a user

For the back office users the method of authentication will be ECAS (European Commission).

### 5.3.2 Web services for re-users

**HLR96.** The administrators of EUR-Lex need specific features in the Back-Office to manage the “export” Web Services dedicated to re-users. They will be able to add or edit predefined flows for example, letting also the possibility to the user to generate its personal flow. Flows configuration and control should be managed through a dedicated user interface

## 5.3.3 Statistics

### 5.3.3.1 Statistics of use

**HLR97.** It shall be possible to generate statistics of use concerning content (a EUR-Lex portal content, other contents) or a collection with a single criteria or a combination of criteria: The web usage statistics might be implemented with an external tool, common to other PUBLICATIONS OFFICE's applications.

- A period
- Number of users with indication on their localisation (domain), with indication of used services,
- Metadata, used terms for search,
- Statistics on usage of notifications (RSS), and web services.

The Back Office should allow collecting of statistics on the web usage generated by external web analytics tools and store and manage these reports.

N. B.: For the sake of rationality and efficiency, the Publications Office intends to adopt the same system for producing statistics for all its sites.

### 5.3.3.2 Production statistics

**HLR98.** It shall be possible for PUBLICATIONS OFFICE's staff, to generate production statistics concerning every content reachable in the EUR-Lex portal, with one criteria or a combination of different criteria:

- A period,
- Number of new documents concerning a specific type of document, a set of types, all types,
- Number of new notices,
- Number of contents referenced by a specific metadata,
- Content exceeding a volume threshold,
- A contribution phase (Creation, stacking before publishing, publishing, ...)
- A level of completion (pre-notice, complete notice, available with all the required formats, ...)
- ...

### 5.3.3.1 Statistics availability

**HLR99.** Statistics of use and production statistics shall be computed with the data and logs coming from the hosting/operating environment.

**HLR100.** All these statistics shall be available with:

- Predefined dashboards,
- Customized dashboards where back-office users select criteria and formats to be used.

## 6 Operations

---

The contractor will have to propose the most reliable environment, and one that will satisfy new EUR-Lex specific hosting requirements.

### **HLR101. Service Level Agreements (SLA) requirements**

For each independent module of the system the contractor will propose the necessary measures and requirements to be included in the Service Level Agreement for hosting and operating, to assure that the new EUR-Lex as a whole will meet the very high requirements for its availability to the end users. This means for example very limited unscheduled downtime and short periods of degraded performance. The severity of problems should be measured statistically, according to the number of unsatisfied users depending on the day of the week, hour and length of the incident. All these conditions and levels of availability have to be proposed for the SLA to be signed subsequently with the site host.

### **HLR102. Redundant Infrastructure**

The contractor has to propose the hardware infrastructure or capacity, taking into account the massive audience (more than 350 million page views/year on EUR-Lex, including the web crawlers), in a redundant failover configuration in order to deliver the service even if one or more elements fail or become unavailable for any other reason.

# 7 Annexes

---

## 7.1 Search state of the art

### 7.1.1 Advanced Linguistic Processing

This chapter introduces the basic concepts of advanced linguistic processing on full text and metadata including:

- Linguistics and Relevancy
- Lemmatization
- Synonyms and Spell Variations
- Spell Checking
- Anti-Phrasing
- Wildcard Search
- Handling Special Characters and Accents

#### 7.1.1.1 Linguistics and Relevancy

In search, linguistics is defined as the use of information about the structure and variation of languages so that users can more easily find relevant information. The document's relevancy with respect to a query is not necessarily decided on the basis of words common to both query and document, but rather the extent that its content satisfies the user's need for information. Linguistics tools determine the intent behind keywords. In order to achieve relevancy, linguistic processing is performed at the document level – during document processing – and at the query level – during query and result processing.

On the query side linguistic processing results in a query transformation; on the document side, linguistic processing results in document enrichment prior to indexing in order to cover grammatical forms and synonyms.

#### 7.1.1.2 Linguistic Concepts

There are a number of basic linguistic concepts that are used throughout the documentation. Understanding these concepts makes it easier to understand how relevancy with respect to linguistic works in search engines. These concepts include: entity extraction, lemmatization, tokenization, normalization, synonym expansion, and spell checking.

- **Entity extraction** is isolating known linguistic constructs.
- **Lemmatization** is expansion of all known terms in the document to their inflectional forms (such as give to gave), or reduction to base forms. Lemmatization enables searches to match documents with similar meaning, but different key words.
- **Synonym expansion** can be performed at query time or index time. Synonym expansion at query time lets specific user (like the search administrator) modify thesauri (or similar concept) when necessary without the need to re-index.
- **Tokenization (also called segmentation)** is the detection of white space characters and symbols that separate words from each other that are not relevant to the matching process.
- **Character normalization** is the replacement of characters or character sequences to improve search results. This helps improve recall, but may, in some cases, have a negative impact on precision. Character normalization is an option and can be beneficial in languages that have specific characters that are used inconsistently or in different variations.
- **Phonetic normalization** is normalization using phonetic matching rules and is performed on the query and document side. Terms that are written differently but sound the same can give

the same result. For example, if searching for the country Poland, a user could type in Poland or Pauland and get the same result.

### **7.1.1.2.1 Lemmatization**

#### **7.1.1.2.1.1 What is Lemmatization?**

Generally speaking, lemmatization means the mapping of a word to its base form and / or all its other inflectional forms, such as

- Singular or plural for nouns,
- Tense and person for verbs,
- Positive, comparative, or superlative forms for adjectives.

The purpose of lemmatization is to enable that a query for one of the possible forms of a word will also match documents that contain a different form of the word.

This allows a user to search for a term like “car” and get both documents that contain the word “car” and documents that contain the word “cars”.

In contrast to stemming or wildcard search, which would match all documents containing words starting with “car”, such as “cared” or “career”, lemmatization allows for recognizing words as matching terms on basis of their being inflectional variations of the query word.

With this, lemmatization also takes irregular inflections such as “tooth” and “teeth” into account.

#### **7.1.1.2.1.2 Lemmatization by Expansion and Reduction**

A search engine should provide support for the following types of lemmatization:

- Lemmatization by document expansion: The words in a document are expanded to all full forms. During indexing, all produced full forms for the document are put into a separate part of the index.
- Lemmatization by query expansion: Words in a query are expanded to all full forms. If no language is set on the query side, the default language is assumed. Nothing is done on the document side.
- Lemmatization by reduction: The words in documents and in the query are reduced to canonical forms (lemmas). Lemmatized query terms are sent to the index of lemmatized words.

#### **7.1.1.2.1.3 Lemmatization and its Influence on the Search Experience**

Enabling lemmatization has the following effects:

- Recall is improved: Documents that would not match since they contain a word form that is different from what the end-user sent to the system will now be retrieved as well. For example, an end-user looking for "laws" on EUR-Lex will also get documents that only contain the word "law".
- The improved recall will also influence the ranking: Documents will be ranked regardless of whether they contain the word the end-user sent to the system or one of its inflected forms. In either way, the documents ranked at the top of the result list will be the most relevant documents.
- These effects are multiplied for multi-words queries where every component might occur in different forms.

### **7.1.1.2.2 Synonyms and Spell Variations**

Synonyms are words that have the same or identical meaning, for example, “live” and “dwell”.

Spelling variations can be viewed as a special case of synonyms.

There are two available options for synonym handling:

- Query-side synonym expansion. This enables dictionary-based synonym expansion on the query side.
- Index-side synonym expansion. This feature enables synonym expansion similar to applying lemmatization – a document to be indexed is expanded with a defined list of synonyms or spell variations to the words it originally contains.

As with lemmatization by expansion on the document side, the original document is indexed in the original search index, whereas the expanded document is indexed in a separate expanded index. This allows administrator to control enabling synonym expansion on a per query basis. One can decide whether a query is to be executed with synonym expansion, in which case the query is sent to the synonym index, or without synonym expansion, in which case the query is sent only to the original index.

#### **7.1.1.2.3 Spell Checking**

The purpose of spell checking is to improve the quality of the queries by comparing the query terms against dictionaries and identifying misspelled query terms. As a result of the spell checking process, a search engine could either replace the query terms automatically with the correct terms, or it could suggest modifications to the query terms to the end-user.

The spell checking algorithm operates on individual query segments. A query segment is a portion of the query that forms a syntactical entity of some kind. For example, if something within the query is put in quotes, that quoted part forms a query segment. Additionally, the “-“ and “+” modifiers in the simple query language also introduce segment breaks, as do explicit field specifications. This means that implicit phrases will not be detected across query segments.

#### **7.1.1.2.4 Anti-Phrasing**

Anti-phrasing is closely related to the concept of stop-words (or noise-words). In contrast to stop-words, however, the anti-phrasing feature does not remove single words, but entire phrases only. Removing single words implies the risk of removing important words that happen to be identical with stop-words. Phrases, in contrast, are more unambiguous and can therefore be removed from the query more safely.

#### **7.1.1.2.5 Wildcard Search**

With full wildcard support it is possible to use '\*' and '?' when specifying a query-term, where '\*' indicates any number of wildcard characters and '?' indicates a single wildcard character. The wildcard characters may be anywhere in the query term.

## **7.1.2 Facetted search and taxonomies**

What appears in the result set of a search, and how it is displayed, depends on many factors. How data is structured, and how the system is set up to display results affect what the user sees after submitting a search. Navigation, taxonomy, and unsupervised clustering make it possible for users to have different views on the result sets.

This chapter includes information about:

- Navigators, which allow users to view a list of values or ranges.
- Taxonomy, which is an organized classification structure that groups documents by category.
- Unsupervised clustering, which allows for automatic grouping of similar documents in the result set and suggested naming of these groups or clusters.

### **7.1.2.1 Facets**

Facets (or Navigators) provide functionality for drilling down into the query results based on value distribution of one or more individual fields. It is possible to apply facets to all fields (or just some

fields) from, for example, a database or document catalog. A search engine should provide both textual, like document type, and numeric, like grant amount attributes.

Different types of facets can be applied depending on the field types.

Document type
OJ-L (20) Treaties (15) Legislation (14) Preparatory acts (12) <i>More...</i>

Figure 11. Example of textual navigator

Date of publication
Before 2000 (20) Between 2000-2004 (15) After 2004 (14)

Figure 12. Example of non-textual navigator

In some search engines, it is possible to perform multi-dimensional navigation in structured data based on facets of the content such as document types, etc.

Thus, relevant results can be found faster using a combination of searching and browsing by parametric value and range.

Facets can also be used on taxonomy fields to apply deep navigation (meaning the entire result set) into categories that occur within the results. When used with taxonomy, each taxonomy node that appears within the result set appears as a navigation entry. The feature is dynamic.

The range for numeric values is computed on-the-fly, trying to give a mean distribution of values to displayed range categories. It is also possible to manually specify internal boundaries.

For each field range drill-down links are provided in order to navigate within the displayed value range, for example, "Publication date 01/10/2006-01/10/2007".

In some search engines, the navigation parameters are sometimes denoted faceted metadata, and may apply for applications such as:

- Documents may have attributes such as document type, author, document source, country of origin...

An indexed field or attribute can be seen as a dimension in which the query can be refined.

The search results are examined on the fly, and data is produced that can be rendered in the form of hyperlinks. This will help the user navigate to find what he or she is looking for by modifying the query.

This is especially relevant in the context of legislation search, where the searchable index is a database or acts catalog. The fields indexed for each document may vary according to the type of the document. By supplying a navigational aid on top of the search engine that is adaptive to the search results for the user's query, relevant results can be found faster.

### 7.1.2.2 Taxonomy

Taxonomy is an organized classification structure that groups documents by category. A document could, for example, belong to the category legislative organizations, or laws. Categorization is the process of mapping documents to specific categories. Some search engine let administrators configure and maintain taxonomies and the mapping of categories. When a set of results is returned, a taxonomy tree is created, which lets users browse information by category.

The categorization process inserts category tags into the documents prior to indexing. When the documents in an index have been categorized, end users can restrict a query to a specific category in that index.



Figure 13. Example of navigation using taxonomy

**Applying taxonomy gives a “category” view of the result set.**

### 7.1.2.3 Unsupervised Clustering

If there is no taxonomy information associated with documents it is possible to setup the system to automatically suggest a category for a document in the result set. This is referred to as unsupervised clustering. Unsupervised clustering is a kind of automatically-generated taxonomy.

### 7.1.2.4 Creating Taxonomy on the Fly

Some search engines provide unsupervised clustering which means that documents are clustered (“grouped” or “categorized”) based on how similar they are rather than using static taxonomy information. Similarity is calculated by comparing document vectors, which are lists of prominent words in the document.

Document vectors are representations of the unstructured textual content that is associated with a document. Vectorization is the process of computing document vectors and is performed as part of document processing using the Vectorizer document processor.

When a set of documents have been put into a cluster, appropriate name(s) or label(s) for the group are calculated based on the terms in the document vectors.

## 7.2 About Content Management System

Content management systems (CMS) are software products enable to fulfill most of the requirements identified for the back office.

In few words, why using a CMS?

- A CMS makes it easier for people to create, edit and publish content on a website. Historically, website publishing has required significant technical skills (HTML, programming). A good CMS allows non-technical authors and editors to easily and quickly publish their content.
- A CMS makes it easier for you to manage who creates, edits and publishes content. Because it establishes defined publishing processes, administrators can allocate specific publishing rights to various individuals.
- By easing technical hurdles in the publication of content, a CMS can reduce the need for training, while facilitating more people to publish. At the same time, it reduces the daily stream of calls to the department in charge of the changes to the website.
- A CMS reduces time-to-publish, allowing you to get content published faster
- A CMS allows for the design of common and consistent information architecture (metadata, classification, navigation, search, layout and design).
- A CMS will allow for the consistent management of metadata through content template structures. Of the many benefits this delivers is a significantly improved search process. Basically, if the appropriate metadata is captured on all documents, then people can find the right content a lot more quickly. Of course this could be limited to EUR-Lex portal content
- A CMS can facilitate better content security. It can control who is allowed to publish to the website, and who is allowed to see what content.
- A CMS can facilitate the measure of the success of the publishing efforts. It is possible to track who is publishing what, how quickly content is getting published, whether the publication schedule is being adhered to, whether out-of-date content is being removed quickly enough, etc.
- A CMS really comes into its own when you have a lot of authors and editors, based at multiple locations, publishing substantial quantities of content on an ongoing basis.

## 7.3 Table of figures

Figure 1. General overview of the Transformation Program .....	4
Figure 2. Transformation focus : from individual pillars to common access layers.....	5
Figure 3. Main sections .....	27
Figure 4. Sub sections of the “Practitioner’s corner” .....	28
Figure 5. Subsection’s content display .....	28
Figure 6. Subsection’s content display .....	29
Figure 7. Edition of a content article.....	30
Figure 8. Preview of an article .....	30
Figure 9. Management of the users .....	32
Figure 10. Creating a user.....	32
Figure 11. Example of textual navigator .....	38
Figure 12. Example of non-textual navigator.....	38
Figure 13. Example of navigation using taxonomy.....	39