# DELIVERABLE

**Project Acronym:** **E-ARK**

**Grant Agreement Number:** **620998**

**Project Title:** **European Archival Records and Knowledge Preservation**

## DELIVERABLE DETAILS

| | |
|---|---|
| **DELIVERABLE REFERENCE NO.** | **D7.3** |
| **DELIVERABLE TITLE** | **Prototype of the Knowledge Centre Service** |
| **REVISION** | **1.0** |

| **AUTHOR(S)** Name(s) | Organisation(s) |
|---|---|
| António Higgs | Instituto Superior Técnico, Universidade de Lisboa (IST) |
| João Cardoso | |
| Diogo Proença | |
| Ricardo Martins | |
| Nuno Reigota | |
| Ricardo Vieira | |
| José Borbinha | |
| **REVIEWER**(S) Name(s) | Organisation(s) |
| David Anderson | University of Brighton (UoB) |
| Kuldar Aas | National Archives of Estonia (NAE) |

| Project co-funded by the European Commission within the ICT Policy Support Programme Dissemination Level | | |
|---|---|---|
| P | Public | X |
| C | Confidential, only for members of the Consortium and the Commission Services | |

# REVISION HISTORY AND STATEMENT OF ORIGINALITY

**Submitted Revisions History**

| Revision No. | Date | Authors(s) | Organisation | Description |
|---|---|---|---|---|
| 0.1 | 11/01/2016 | Ricardo Vieira | IST | First deliverable draft |
| 0.2 | 20/01/2016 | António Higgs | IST | First version with all content created |
| 0.3 | 21/01/2016 | Ricardo Vieira | IST | Overall Revision |
| 0.4 | 26/01/2016 | David Anderson | UoB | Suggested revisions to document following review |
| 0.5 | 28/01/2016 | Kuldar Aas | NAE | Suggested revisions to document following review |
| 1.0 | 29/01/2016 | António Higgs | IST | Overall revision of the document |

**Statement of originality:**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

# Executive Summary

This deliverable details the Knowledge Centre Prototype - an aggregator of services developed under the scope of the E-ARK Project[1] and available at http://kc.dlmforum.eu. These services aim to provide a comprehensive, but simple and easily understood, set of services that allow to understand, contribute and validate Information Governance good practices. The Knowledge Centre is divided into: The Resources Centre and the Assessment Services.

The Resources Centre is where users can learn more about Information Governance terms, requirements and practise. It is composed by the Vocabulary Manager Service (entitled EVOC) and the Reference Requirements Management Service (entitled REQs). EVOC allows users to (1) learn more about Information Governance terms and definitions, (2) contribute to the development of the current Information Governance terminology, and (3) create their own reference vocabularies based on the existing terms and definitions of the service. REQs allows users to learn more about Information Governance requirements and practices by consulting and analysing reference documents.

The assessment services allow users to assess their Information Governance practices. It is composed by the MoReq Export Validator Service and the Maturity Assessment. The former allows users to validate XML data files according to the MoReq2010 Export Schema. The latter allows users to assess their maturity level according to the maturity model defined in deliverable D7.1 of E-ARK.

The deliverable describes the goal and intent of each service and functionality as it explains how those can be used. Additionally, it briefly describes how data is stored in the services, which technology is behind them, and which future developments are planned for the next versions of the Knowledge Centre.

[1] http://www.eark-project.com/

# Table of Contents

Executive Summary ..... 4
1. Introduction ..... 8
2. Knowledge Centre Service ..... 9
3. Resources Centre ..... 11
3.1. Vocabulary Manager (EVOC) ..... 11
3.1.1. User Management ..... 13
3.1.2. Public Spaces View ..... 14
3.1.3. Vocabulary Management View ..... 17
3.1.4. Configuration View ..... 23
3.1.5. Users View ..... 24
3.2. Reference Requirement Manager (REQs) ..... 25
3.2.1. REQs Home Page ..... 26
3.2.2. REQs Overview Page ..... 27
3.3. Data Store ..... 32
3.3.1. EVOC ..... 32
3.3.2. REQs ..... 32
4. Assessment Services ..... 34
4.1. MoReq Export Validator Service ..... 34
4.2. Maturity Assessment ..... 37
5. Future Work ..... 40
5.1. MoReq Assessment ..... 40
5.2. Existing Services Improvements ..... 41
Appendix A – Technological Specifications ..... 43

# Table of Figures

Figure 1. The E-ARK Knowledge Centre Architecture Model. ..... 10
Figure 2. The Knowledge Centre Services Aggregator home page. ..... 10
Figure 3. Resources Centre Page ..... 11
Figure 4. EVOC Main Layout. ..... 12
Figure 5. EVOC Top Toolbar ..... 13
Figure 6. EVOC User Options Menu. ..... 13
Figure 7. EVOC Top Toolbar without a user logged. ..... 14
Figure 8. EVOC User Login Menu. ..... 14
Figure 9. EVOC User Registration Menu. ..... 14
Figure 10. EVOC Public Space when a Group is selected. ..... 15
Figure 11. EVOC Public Space when a Vocabulary is selected. ..... 16
Figure 12. EVOC Public Space when a Term is selected. ..... 16
Figure 13. EVOC Vocabulary Management View layout. ..... 17
Figure 14. Space, Group and Vocabulary management buttons ..... 17
Figure 15. Space, Group and Vocabulary create menu. ..... 18
Figure 16. Form to create a new Space. ..... 18
Figure 17. Form to create a new Group. ..... 19
Figure 18. Form to create a new Space. ..... 20
Figure 19. EVOC Vocabulary Management View right panel layout. ..... 21
Figure 20. EVOC Vocabulary Management View Export options. ..... 21
Figure 21. Term Management buttons. ..... 21
Figure 22. Form to create a new Term. ..... 22
Figure 23. Import terms layout. ..... 22
Figure 24. Import terms management buttons. ..... 23
Figure 25. Terms representations. ..... 23
Figure 26. EVOC Configuration View - Vocabularies External Setup. ..... 24
Figure 27. EVOC User Management menu. ..... 25
Figure 28. EVOC User grouping. ..... 25
Figure 29. REQs Home Page. ..... 26
Figure 30. REQs top toolbar. ..... 27
Figure 31. REQs Overview Page. ..... 27
Figure 32. REQs Navigation Panel. ..... 28
Figure 33. REQs Navigation Panel Header. ..... 28
Figure 34. REQs Navigation Panel – document selection. ..... 29
Figure 35. REQs Content Panel. ..... 29
Figure 36. REQs content panel with multiple sections. ..... 30
Figure 37. REQS Content Panel links ..... 30
Figure 38. REQs Content Panel link target. ..... 31
Figure 39. REQs Requirement layout. ..... 31
Figure 40. EVOC Data Domain Model. ..... 32
Figure 41. REQs Data Domain Model. ..... 33
Figure 42. Assessment Services Page ..... 34

Figure 43. MoReq Export Validator Service main layout ..... 35
Figure 44. Examples panel layout ..... 36
Figure 45. Validation Report layout ..... 37
Figure 46. Example of a list of errors that occurred during XSD validation ..... 37
Figure 47. Maturity Assessment Service ..... 38
Figure 48. Examples of Maturity Assessment Results. ..... 39
Figure 49. MoReq Assessment Use Cases. ..... 41
Figure 50. Base technology stack. ..... 43

# 1. Introduction

The E-ARK project aims to harmonize currently fragmented archival approaches for ingest, preservation and re-use. Despite the lack of standardization in such processes, archival management and more broadly Information Governance (IG) has known extensive research and practice in the past years. In fact, nowadays many business and technical references emerge to guide the processes of ingesting, managing, preserving and accessing information. In terms of designing the processes, standards such as ISO154894 ("Records Management") or ISO30300/15 ("Management systems for records") already exist. For implementing tools and services references such as MoReq2010 and ISO161756 ("Principles and functional requirements for records in electronic office environments") are well known. For assessing organizations and tools one can refer to ISO163637 ("Audit and certification of trustworthy digital repositories") or ISO181288 ("Risk assessment for records processes and systems"). Also it is important to note that the examples above are international references. As stated above, due to lack of standardization, several countries have also been defining national practices and procedures that should not be discarded and constitute relevant knowledge to the field.

Additionally, it is also important that, as a strategy, we admit we live in a world where problems can viewed from different perspectives, and consequently solutions need to consider requirements from different areas. Applying this to IG, it means we must recognize other specific views besides only those of information science, such as also information systems, software engineering, risk management, among others.

The proliferation of standards and references together with the recognition that problems should be analysed from different perspectives, motivations and communities has raised the need for a knowledge system that allows stakeholders to get a consolidated view of existing knowledge. The E-ARK Knowledge Centre is the representation of such a system. This deliverable describes the first fully-functional prototype of the Knowledge Centre. Chapter 2. Knowledge Centre Service starts by introducing the idealized architecture of the Knowledge Centre and the services that characterize and compose the system. Chapter 3. Resources Centre describes the Vocabulary Manager (EVOC) and the Reference Requirement Manager (REQs) respectively in Chapter 3.1 and 3.2. The chapter concludes by describing how data is stored in both EVOC and REQs (Chapter 3.3 Data Store) services. Chapter 4. Assessment Services describe the assessment services namely the MoReq Export Validator Service (Chapter 4.1) and the Maturity Assessment (Chapter 4.2). Chapter 5. Future Work presents the planning for the next versions of the Knowledge Centre where a new service: MoReq Assessment (Chapter 5.1) will be introduced and the existing services will be improved (Chapter 5.2). Finally, Appendix A briefly describes the technology behind the developed services.

# 2. Knowledge Centre Service

The Knowledge Centre (or KC) is an aggregator of services developed under the scope of the E-ARK Project[2]. These services aim to provide a comprehensive, but simple and easily understood, set of requirements for Information Governance. These requirements are elicited from existing reference documents. The main goals of the E-ARK Knowledge Centre are to provide organizations services to:

- Support the definition of Information Governance practices;
- Contribute to the existing body of knowledge of Information Governance;
- Assess existing Information Governance practices;

The Knowledge Centre can be used in many different ways, including:

**By businesses**:

- As a reference source for learning more about Information Governance;
- As a guide to the audit of existing Information Governance practices; and
- As a guide to improve existing Information Governance practices.

**By academia**:

- As a reference glossary for Information Governance terms and their meanings;
- As a teaching resource for academic institutions; and
- To contribute to the existing body of knowledge of Information Governance.

**By industry**:

- To guide the development of Information Governance tools and techniques; and
- As a reference source for testing and compliance of existing Information Governance tools and techniques;

In order to achieve the aforementioned goals, the Knowledge Centre services can be categorized as:

- **Resources Centre** – containing services to store, view and manage vocabulary, requirements and requirement based documents.
- **Assessment Services** – containing services for assessment and validation of existing Information Governance practices and tools.

**Figure 1** presents an idealized architecture for the KC. The services are backed up by a Data Store component, which represents the databases and necessary accesses to keep all the data of the services safe and ready for usage. Finally, a View Manager acts as an entrance door to all the services, providing a gateway user interface and API (programming interface) to access them.

A web interface was developed to serve as an aggregator for all the Knowledge Centre services. This is the user interface part of the View Manager component represented in **Figure 1** and is currently available at http://kc.dlmforum.eu and is

[2] http://www.eark-project.com/

illustrated in **Figure 2**. This user interface represents the Home page of the Knowledge Centre. The webpage shortly describes each one of the existing services and the benefits that users can get from them.

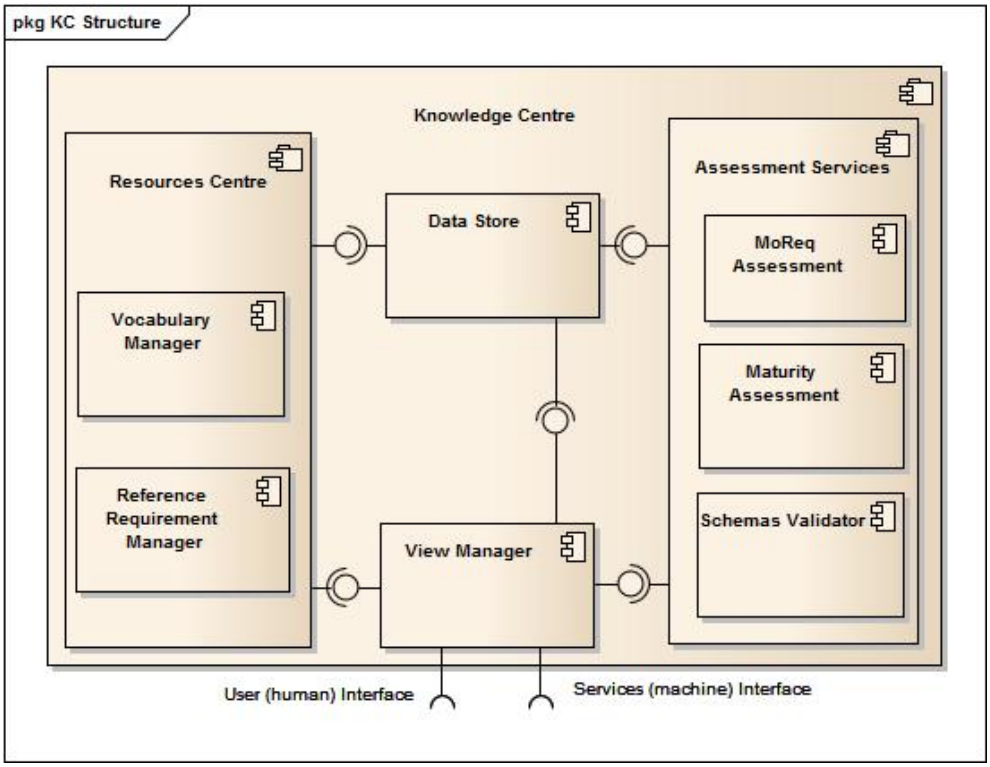


**Figure 1. The E-ARK Knowledge Centre Architecture Model.**

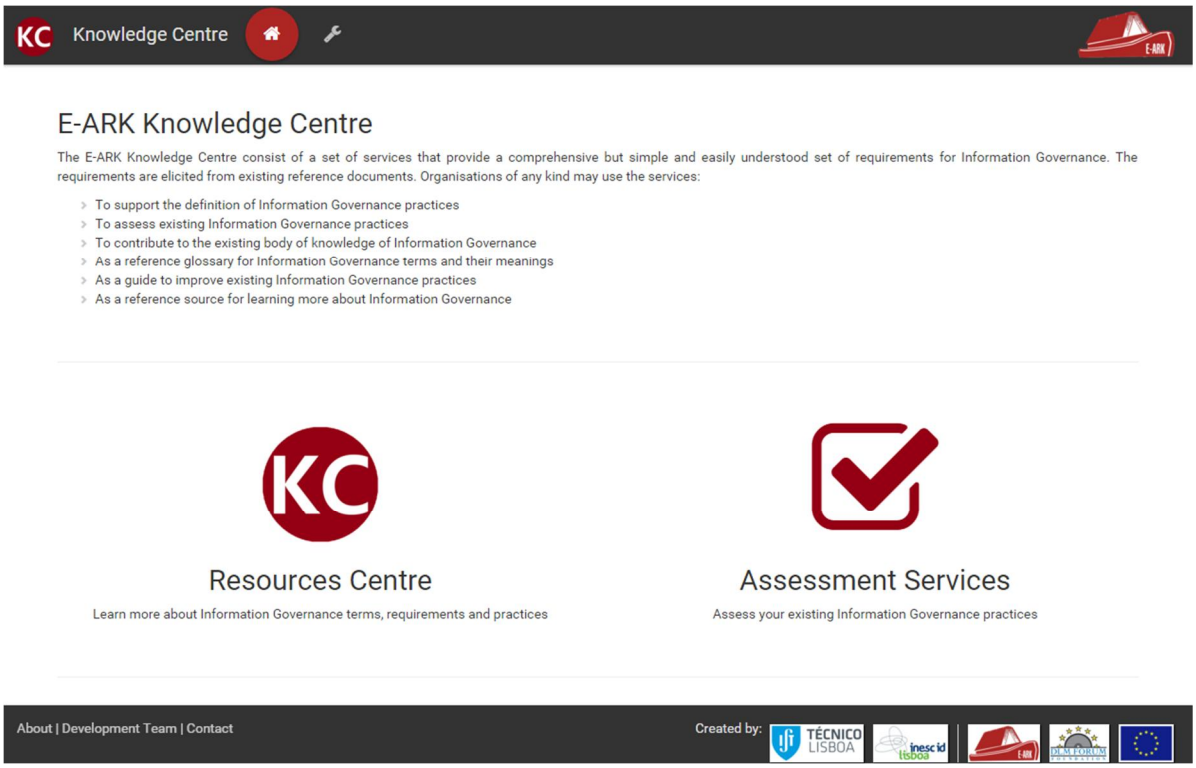


**Figure 2. The Knowledge Centre Services Aggregator home page**.

# 3. Resources Centre

The Resource Centre (**Figure 3**) is the component of the Knowledge Centre containing services to store, view and manage vocabularies, requirements and requirement based documents. In the current prototype, two services are already developed: The Vocabulary Manager, to view and manage vocabularies, and the Reference Requirement Manager to view and edit[3] reference requirements documents. The following sub-sections describe the concepts and functionalities of the current prototypes of EVOC and REQs.

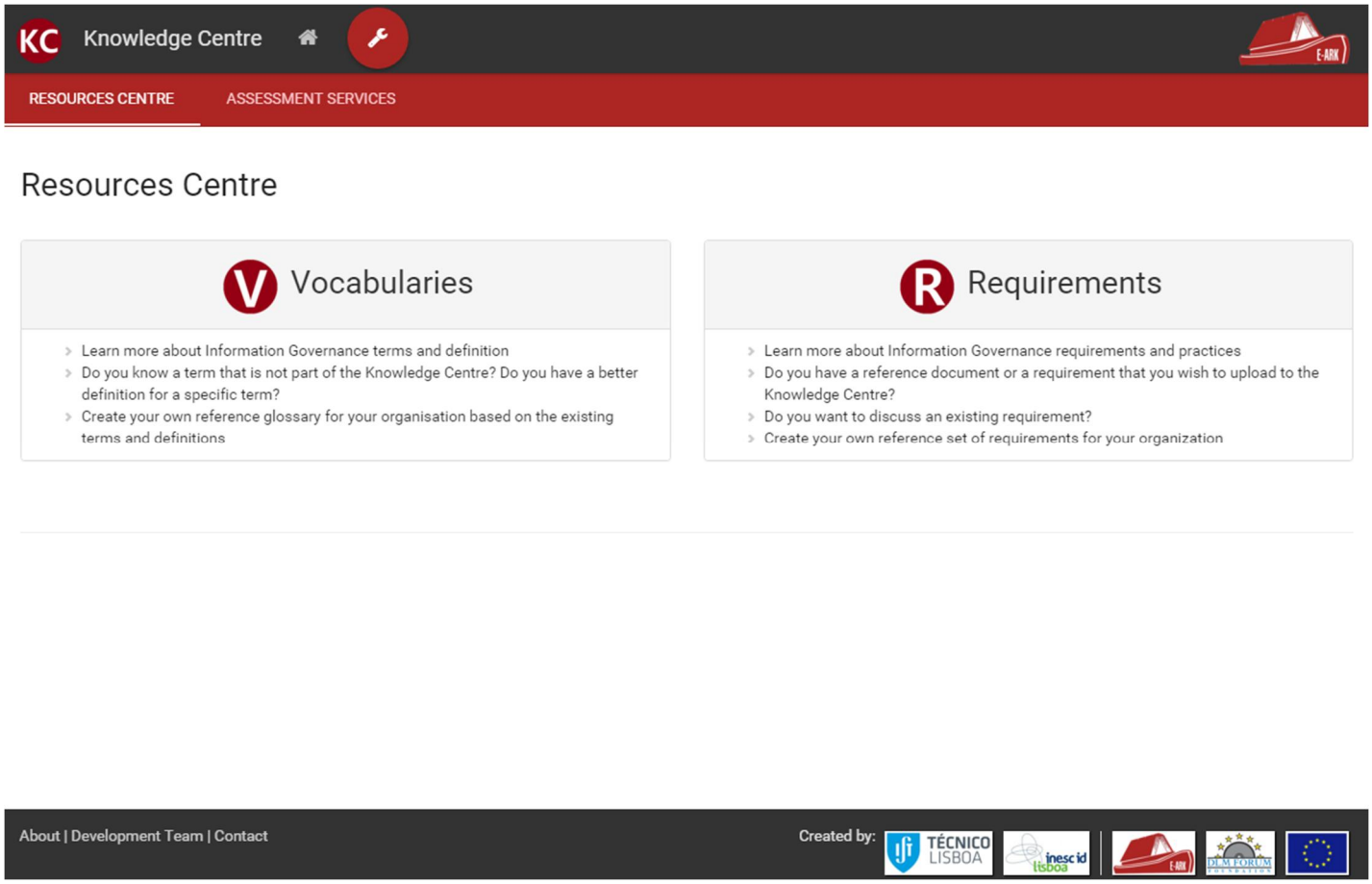


**Figure 3. Resources Centre Page**

## 3.1. Vocabulary Manager (EVOC)

EVOC is a service that allows users to create, manage and view vocabularies. The service allows the definition of different **Vocabulary Types**, i.e., vocabularies characterized by different metadata. A vocabulary node is the generic name given to each entry of the vocabulary. A **Vocabulary Type** defines the specific name for the nodes of that type, and the metadata fields that compose each node. A **Node Attribute Type** represents the metadata fields that can be used in one or more **Vocabulary Types**. They define the name of each metadata field, the type of field (e.g. text area, checkbox, etc.) and the mandatory status of the field (defines if the field is mandatory or not). Using the mentioned concepts a User can create a **Vocabulary Type** entitled "Glossary" with nodes named "Terms", and "Name", "Definition" and "Language"

[3] The existing service at the moment only has the capabilities of visualization of the reference requirements documents. Editing capabilities will be added in future versions of the service. More information about future developments can be found in chapter 5. Future Work.

as fields defined by **Node Attribute Types**. Additionally, **Vocabularies** are organized by **Spaces** and **Groups**. A **Space** represents a context and it can be, for example, an organization, a project, a person, etc. **Spaces** can be divided into **Groups** for better organization of the existing vocabularies. For example, the E-ARK Glossaries are under the E-ARK Space and divided into two Groups "Deliverables", where all vocabularies from the deliverables are inserted, and "Others", where reference vocabularies that were used in the project are inserted.

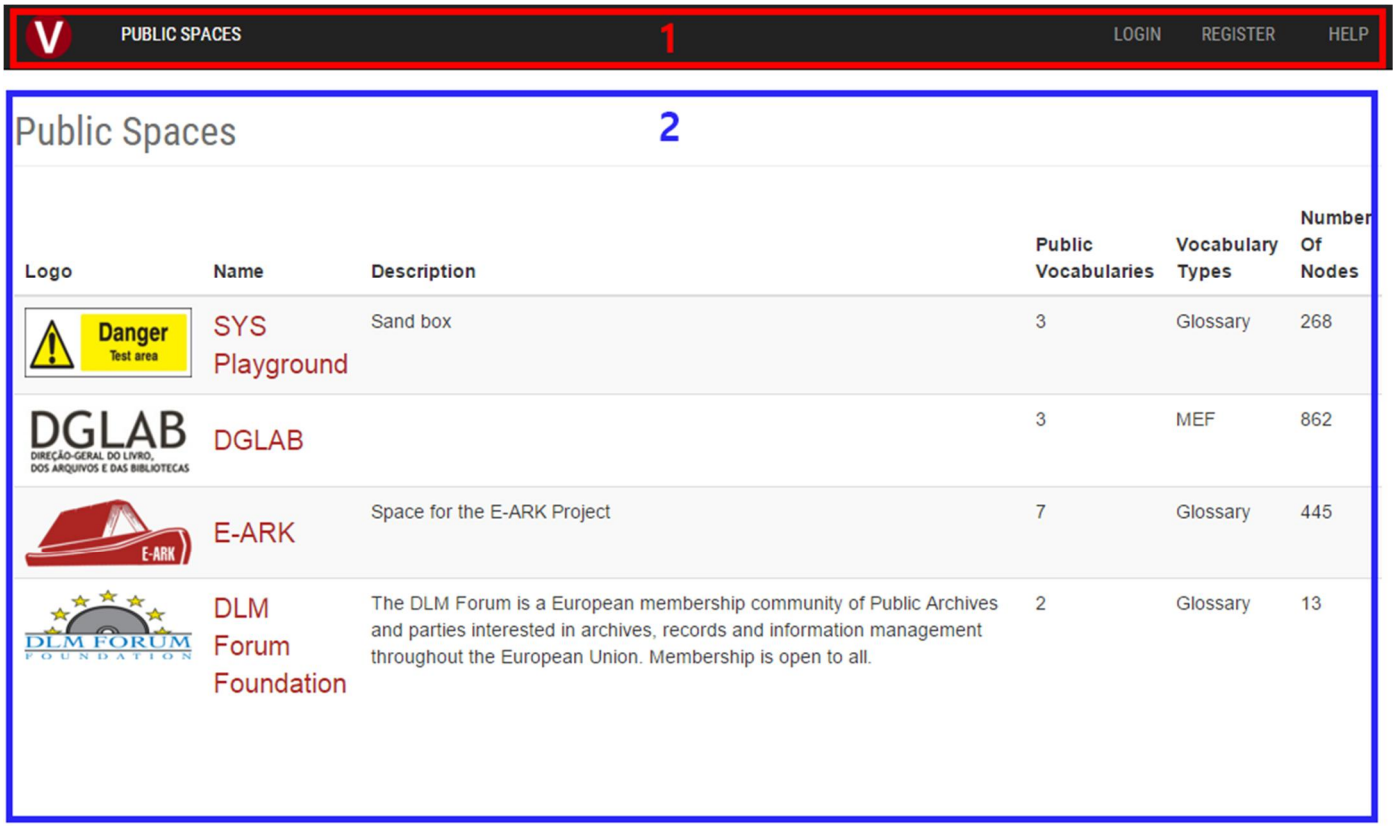


**Figure 4. EVOC Main Layout.**

The service is composed by a set of views: Public Spaces View, Vocabulary Management View, Configuration View, and Users' View. Each view allows access to a set of functionalities. Users of the service have different roles depending on the views they have access to. The main layout of the application is composed of two main sections (Figure 4):

1. The **top toolbar**, common across all the application, providing easy access to its features and views available to the user (which depend on the associated role); and
2. The v**iew display**, where the selected view is presented.

The **top toolbar** can be used to access the various features and views of the application (**Figure 5**):

1. The EVOC logo that allows the user to instantly go to the Public Spaces View;

2 to 5. The various views of the application, which availability depends on the current user's role.

6. The username, when a user is logged in. Can be used to access the User Settings and Logout (**Figure 6**).
7. Help button, allowing the users to send an email to the EVOC support team.

**Figure 5. EVOC Top Toolbar**

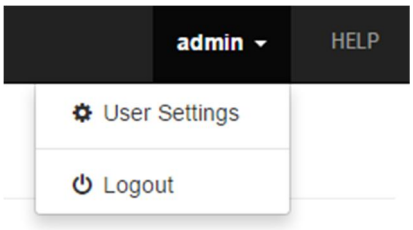


**Figure 6. EVOC User Options Menu.**

### 3.1.1. User Management

There are currently three user roles available in the EVOC application, each with access to different views and their functionalities:

- **Public** – unregistered users who have only access to Public Spaces View and consequently can only consult the vocabularies that are public available (see Chapter 3.1.2 Public Spaces);
- **Registered** – registered users who have access to the Public Spaces and Vocabulary Management Views. The latter allows the user to manage their (public and private) vocabularies. Additionally, Registered Users can create new Spaces, Groups and Vocabularies (see Chapter 3.1.3 Vocabulary Management); and
- **Admin** – registered users with administrator privileges who have access to all views including the Configuration and Users Views. The former allows the definition and management of Vocabulary Types. The latter allows to define and management User Accounts (see Chapter 3.1.4 Configuration and 3.1.5 Users). The administrators are the maintenance team of EVOC. This is the IST development team for the E-ARK project and, when de project concludes, it will pass to DLM Forum administration.

When no user is logged in the top toolbar presents two other options rather than the username (**Figure 7**):

1. Login button, granting access login menu (**Figure 8**). Here it's also possible to retrieve lost passwords.
2. Register button, granting access to the registration menu (**Figure 9**).

User logout is available by clicking the username in the toolbar (6 from **Figure 5**) and selecting 'Logout' (**Figure 6**).

**Figure 7. EVOC Top Toolbar without a user logged.**

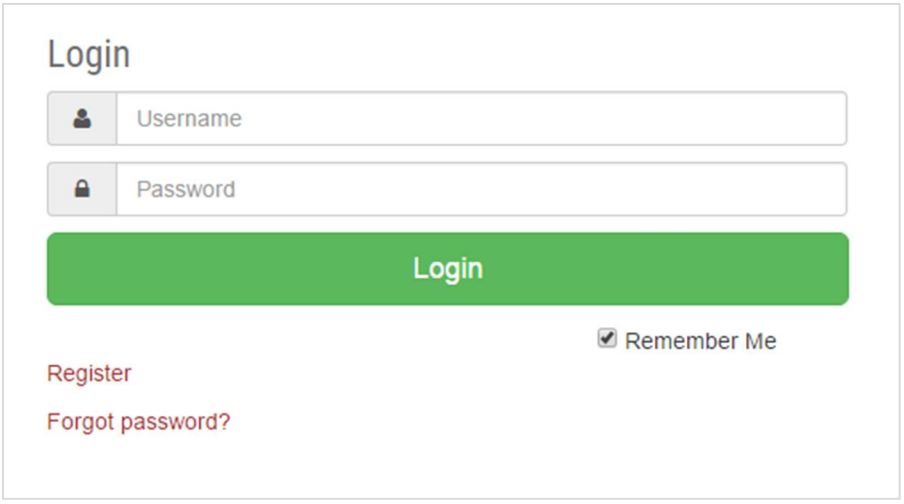


**Figure 8. EVOC User Login Menu.**

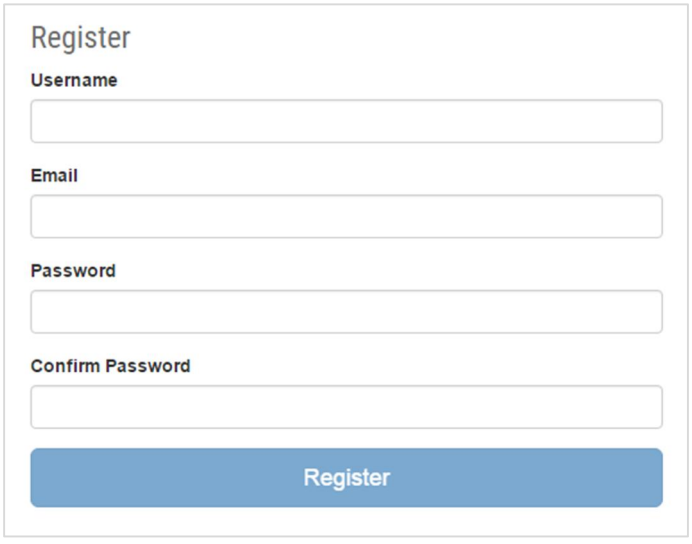


**Figure 9. EVOC User Registration Menu.**

### 3.1.2. Public Spaces View

The Public Spaces view (2 from **Figure 5**) is available to all users. It presents all the public vocabularies available in EVOC. The view starts by listing all the spaces with public vocabularies (2 from **Figure 4**), detailing their logo, name, description, number of public vocabularies, the vocabulary type and the number of vocabularies nodes in that space.

Clicking on a space name will open its contents view (**Figure 10**). On the left side panel of this view (2 from **Figure 10**) is displayed the list of public vocabularies of the selected space, which may be organized by groups. These are indicated by icons:

- : represents a Group;
- : represents a Public Vocabulary.

**Figure 10** show the view presented when a vocabulary group is selected, divided in parts:

1. **Public Space** name, logo and group name;
2. The list of **Public Vocabularies** and their **Groups**;
3. The list of **Terms** of the **Vocabularies** in the selected **Group**;
4. Allows the user to filter the list of **Terms** (3) by **Vocabulary Type**. Please note that this option exists because a **Group** can contain different **Vocabulary Types**;
5. Allows the user to search a selected **Vocabulary Node** for a specific text.

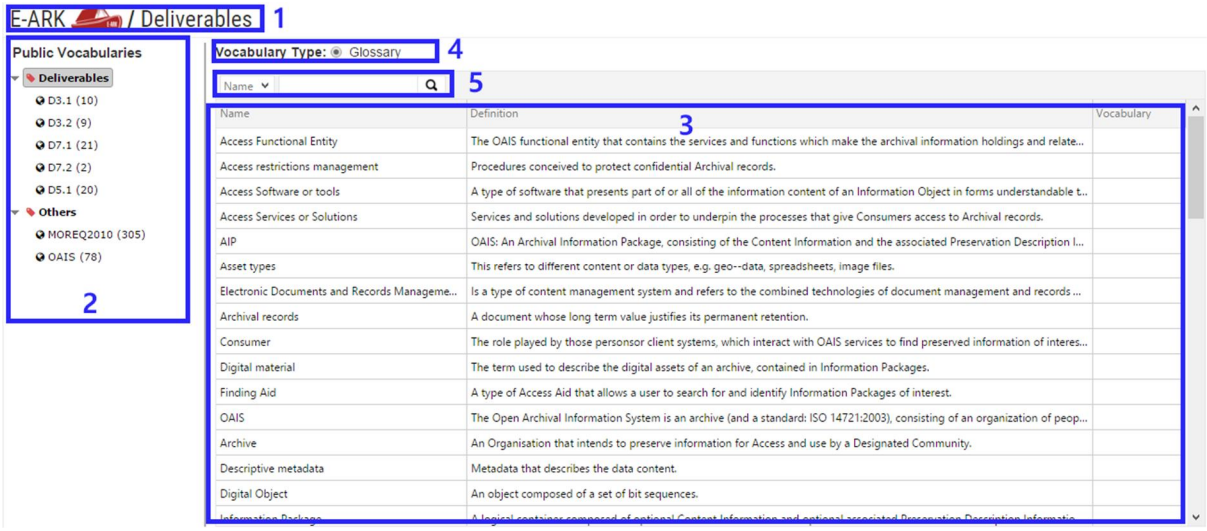


**Figure 10. EVOC Public Space when a Group is selected.**

On the other hand, when a **Vocabulary** is selected, some other options are available (**Figure 11**):

1. **Public Space** name, logo and vocabulary name;
2. The list of **Terms** in the selected **Vocabulary**;
3. Allows the user to hide or show **Vocabulary Nodes** from the **Terms** list (2).
4. Allows the user to filter the list of **Terms** (2) based on specific text. Different text can be specified for different **Vocabulary Nodes**.
5. **Detailed View** and **Export**. By default, when a **Vocabulary** is selected, the list of **Terms** (2) is displayed in a compact format with only one line per term and hiding overflowing text in each cell. The '**Detailed View**' button expands each cell to show all the text of each field. The '**Export**' button exports a vocabulary to XSLX format (Excel file).

**Figure 11. EVOC Public Space when a Vocabulary is selected.**

Any term in the **Terms** list of a **Vocabulary** can be selected, which will prompt the appearance of a right panel with further details about that term (**Figure 12**). This panel can be collapsed by pressing the arrow button (1 of **Figure 12**). Also note that, the URL of the EVOC application changes depending on the selected vocabulary and term, allowing the sharing of links that point directly to a term or vocabulary.

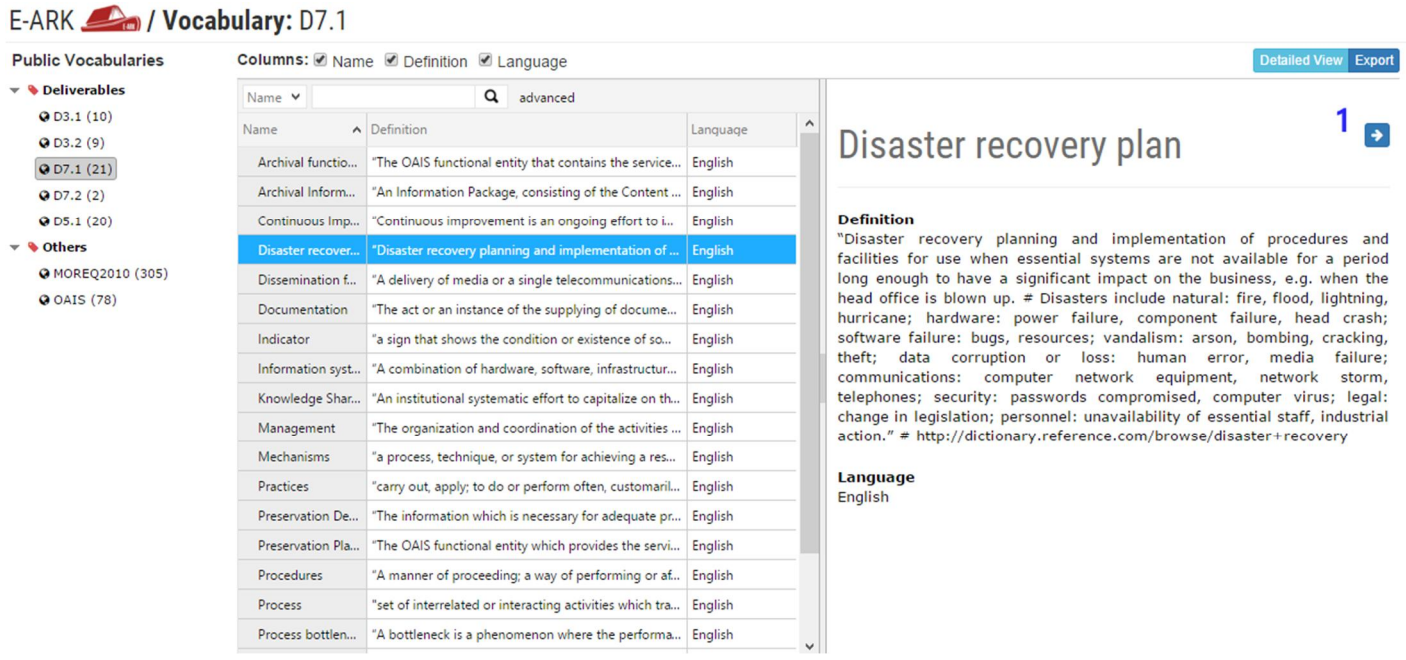


**Figure 12. EVOC Public Space when a Term is selected.**

### 3.1.3. Vocabulary Management View

Registered users in the EVOC system (**Administrators** or **Registered**) have access to the **Vocabulary Management View**, where they can explore and edit vocabularies. Note that users only have access to vocabularies they have created, or those which the owner (a registered user) of the vocabulary has granted them access.

The layout of the **Vocabulary Management View** is divided in two main areas (**Figure 13**):

1. The left panel, listing all the **Spaces**, **Groups** and **Vocabularies** available to the user, as well as the buttons to manage them;
2. The right panel, showing all the **Terms** of a **Vocabulary**, **Group,** or **Space** depending on which is selected in 1, as well as the buttons to manage them.

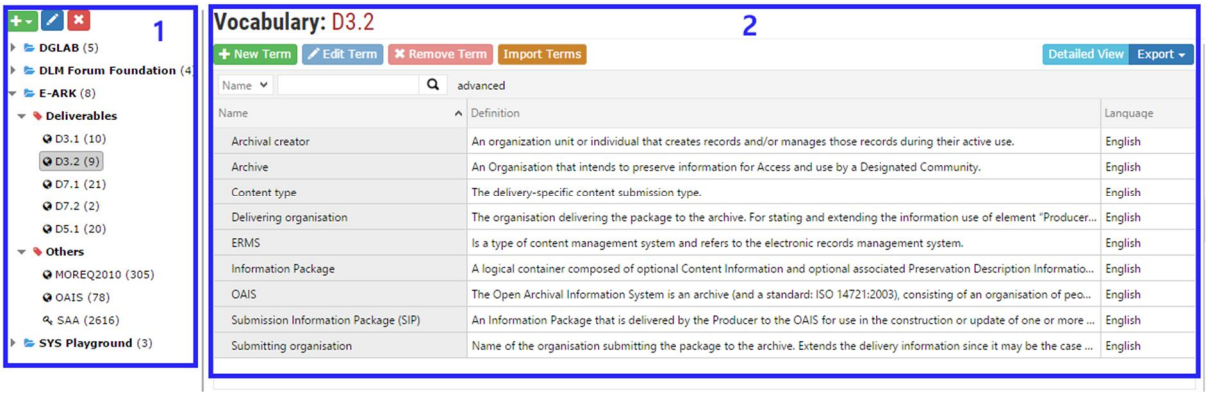


**Figure 13. EVOC Vocabulary Management View layout.**

A user can manage **Spaces**, **Vocabularies** and **Groups** in his Vocabulary Management view, through the tool buttons on the top side of the left panel of the screen (**Figure 14**):

1. The create menu (**Figure 15**), with options to create new **Spaces** (**Figure 16**), **Vocabularies** (**Figure 18**) and **Groups** (**Figure 17**);
2. The edit button, available when a **Space**, **Vocabulary** or **Group** is selected. allows the editing of the selected item proprieties;
3. The delete button, available when a **Space, Vocabulary** or **Group** is selected, deletes the selected item.

**Figure 14. Space, Group and Vocabulary management buttons.**

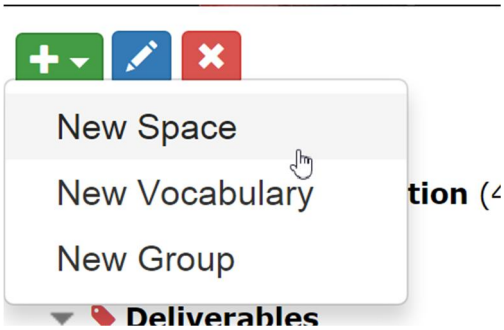


**Figure 15. Space, Group and Vocabulary create menu.**

To create a new **Space**, the following metadata needs to be defined (**Figure 16**): 'Name', 'Description', 'Users' who will have access to the space, the confidentiality of the **Space** ('Public' or Private), and an image to serve as the **Space** 'Logo'.

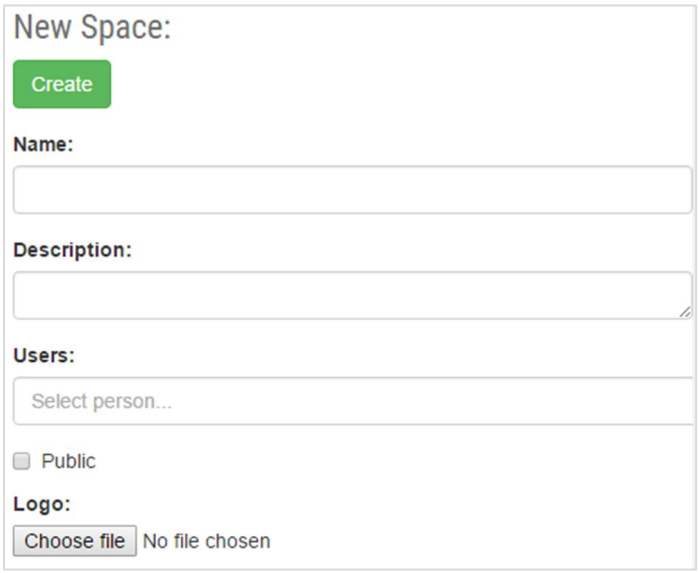


**Figure 16. Form to create a new Space.**

To create a new **Group** (**Figure 17**), users need to define a name, the **Space** of the **Group** being created, and, if relevant, another **Parent Group**. If a parent group is defined, the new Group will be classified under the parent group. This functionality allows users to define a hierarchy of groups if they wish so.

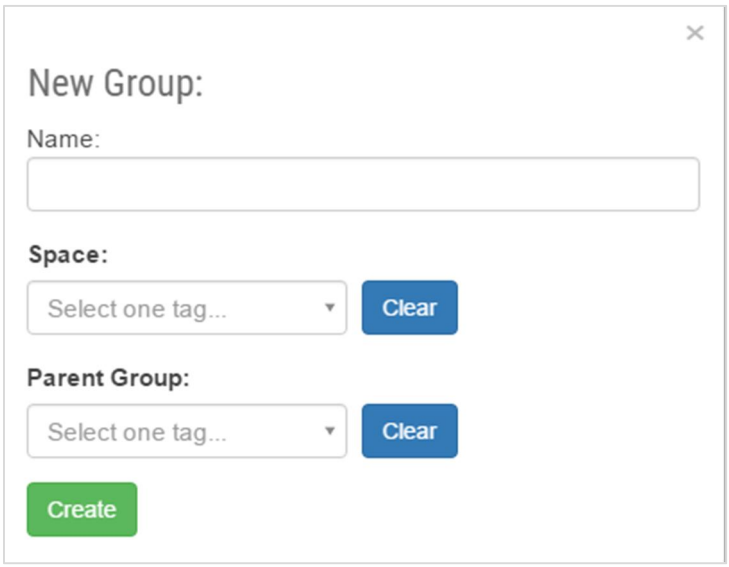


**Figure 17. Form to create a new Group.**

To create a new Vocabulary, EVOC allows the following metadata (**Figure 18**):

- *Acronym*, a small, unique name that will serve to identify the **Vocabulary** through EVOC;
- *Name* of the **Vocabulary**;
- The confidentiality of the **Vocabulary** – *Public* or *Private* (not public);
- *Description* of the **Vocabulary;**
- *External References* of the **Vocabulary**. To be used when the Vocabulary is based on a website, document, or other reference source. *External References* are shown next to the **Vocabulary** name. Multiple *External References* should be separated by a new line (EVOC reads each line as a reference). If Users wish to provide an URL to the *External Reference* the following syntax can be used <name>___<hyperlink> where <name> should be the name of the *External Reference*, '___' should be three underscores in succession, and <hyperlink> should be the URL you wish Users get redirected when clicking the *External Reference*. **Figure 19** is an example of a **Vocabulary** where an *External Reference* with an URL was defined;
- The **Vocabulary** *Type* to be chosen from the existing **Vocabulary Types** defined in the Configuration View (see Chapter 3.1.4 Configuration View for more details);
- The ***Space*** and the *Parent* ***Group*** of the **Vocabulary** being created; and
- The *Import Vocabulary.* Instead of creating a **Vocabulary** from scratch, Users can import a vocabulary from an Excel spreadsheet. In order to do so, the columns of the spreadsheet need to match the **Vocabulary Nodes** of the **Vocabulary Type** selected, and the language of the imported terms needs to be selected.

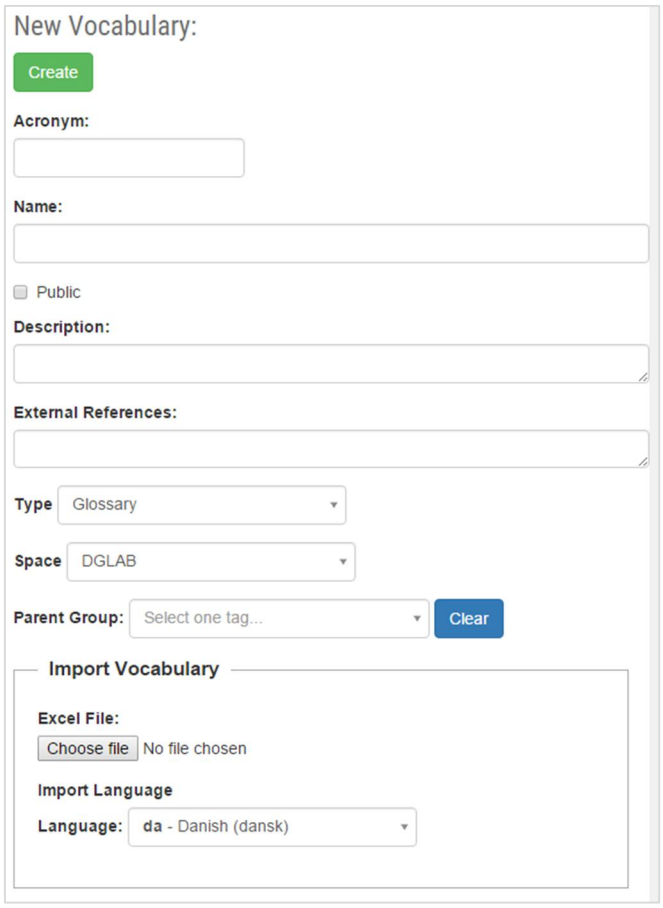


**Figure 18. Form to create a new Space.**

When a **Vocabulary** is selected in left panel of the Vocabulary Management View, all the **Terms** of that **Vocabulary** are shown in the right panel. The viewing options are similar to those of the **Public Spaces View** (see Chapter 3.1.2 Public Spaces View for more information). **Figure 19** presents the general layout of the right panel when a vocabulary is loaded. Please note:

1. The buttons to manage the **Terms** in a **Vocabulary**;
2. The **Terms** list. In the Vocabulary Management View selecting a term will give access to different options in the management controls (1 in **Figure 19**); and
3. The Export options (**Figure 20**). Apart from the export functionality already described in the Public Spaces View (Chapter 3.1.2 Public Spaces View), the Vocabulary Management View allows Users to export the Excel Template, i.e. the spreadsheet already matching the Vocabulary Nodes of the Vocabulary Type being edited. This functionality allows to have a spreadsheet file that can be used for future imports.

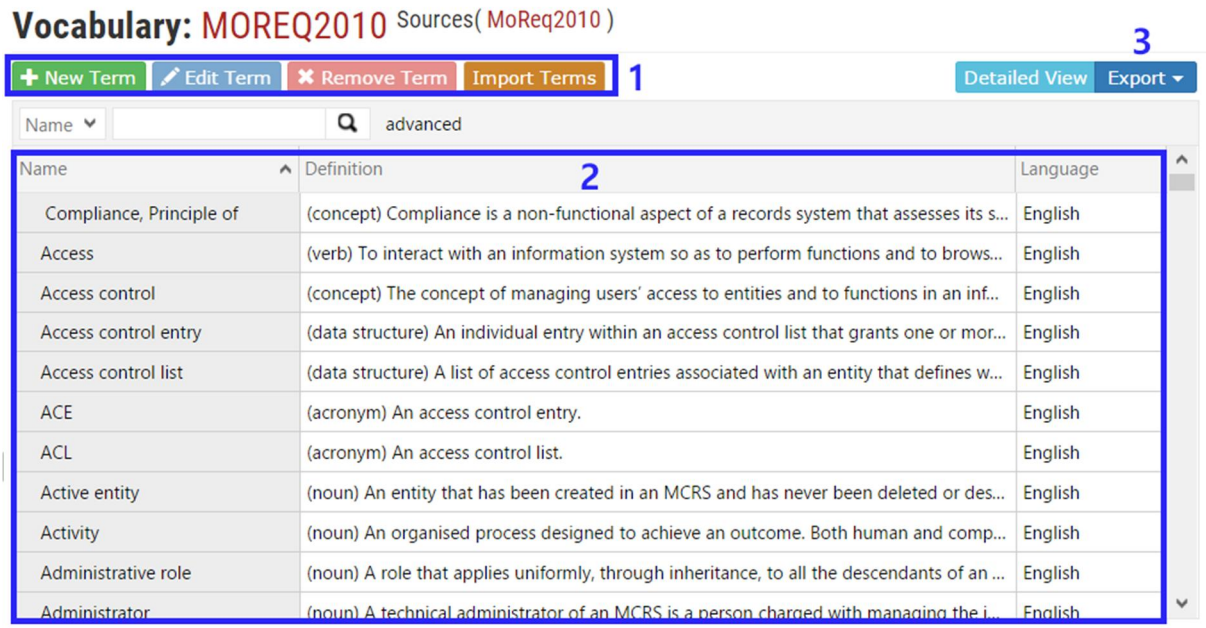


**Figure 19. EVOC Vocabulary Management View right panel layout.**

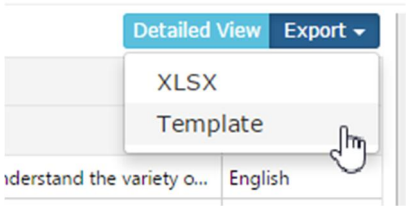


**Figure 20. EVOC Vocabulary Management View Export options.**

There are four buttons to manage terms in a vocabulary (**Figure 21**):

1. The 'New Term' button gives access to the new Term (node) form (**Figure 22**) to create a new **Term**. The required metadata to create a Term will vary depending on the **Vocabulary Type**.
2. 'Edit Term' button, available when a **Term** is selected.
3. 'Remove Term' button, available when a **Term** is selected. If the **Term** is imported from another **Vocabulary,** then the **Term** is only removed in the selected **Vocabulary**.
4. 'Import Terms' button that allows the import of **Terms** from other **Vocabularies** (1 from **Figure 23**).

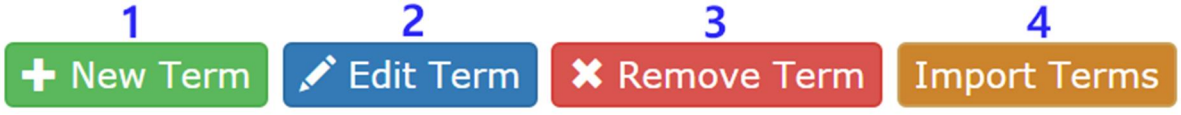


**Figure 21. Term Management buttons.**

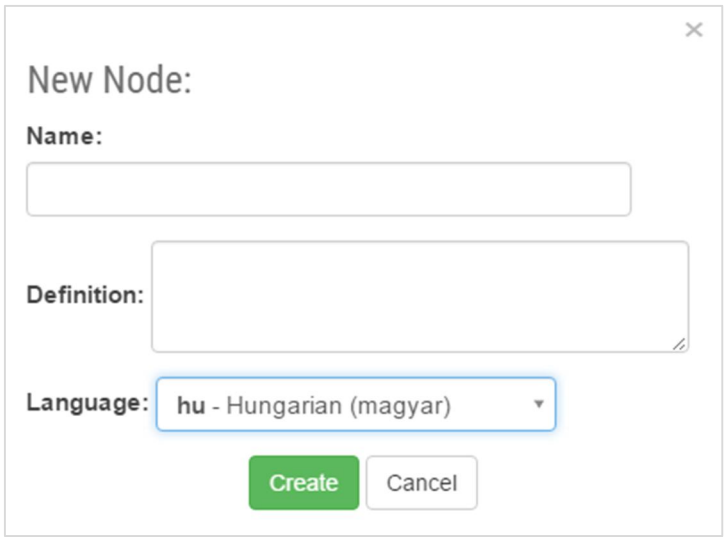


**Figure 22. Form to create a new Term.**

Selecting the 'Import Terms' button will open the import terms panel (**Figure 23**). **Figure 24** shows a close-up of the controls available in this panel, which are:

1. The 'Copy' button, to copy the selected **Term** from another **Vocabulary**;
2. The 'Copy All' button, to copy all **Terms** from another **Vocabulary**;
3. The 'Source Vocabulary' combo box, where the **Vocabulary** to import terms from is selected. It includes all the Vocabularies of the current space;
4. 'Hide' button, to hide the import terms panel and return to the default right panel layout (**Figure 19**).

Please note, that you can only import terms from **Vocabularies** in the same **Space** and with the same **Vocabulary Type**.

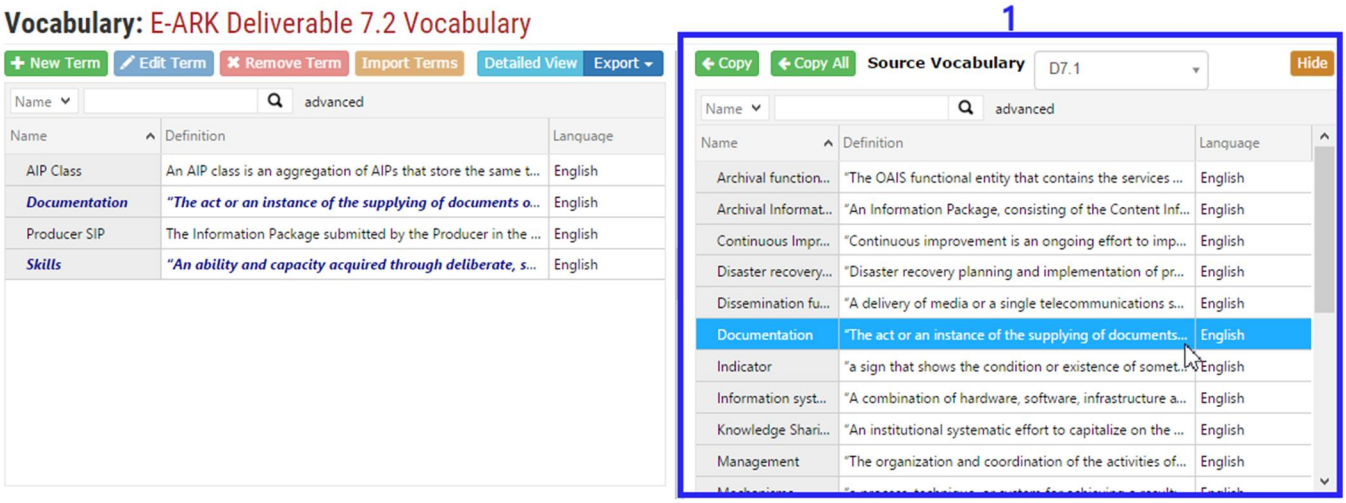


**Figure 23. Import terms layout.**

**Figure 24. Import terms management buttons.**

When a **Vocabulary** has an imported **Term**, that **Term** will appear in blue text on the **Terms** list (**Figure 25**). Editing an imported **Term** will unlink it from the original **Vocabulary's Term**, making them effectively two separate terms. However, changes on the origin **Vocabulary** will be reflected in all the **Vocabularies** that **Term** is imported to. Similarly, deleting an imported **Term** from a **Vocabulary** will have no effect on the original **Vocabulary's Term**. However, deleting a **Term** in the origin **Vocabulary** will remove it from all the **Vocabularies** that term is imported to.

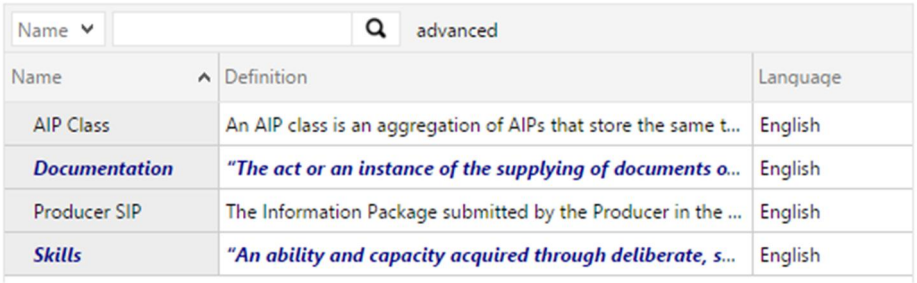


**Figure 25. Terms representations.**

### 3.1.4. Configuration View

As described in Chapter 3.1 Vocabulary Manager (EVOC), **Vocabularies** in EVOC can have different **Vocabulary Types**, i.e., vocabularies characterized by different metadata. . A vocabulary node is the generic name given to each entry of the vocabulary. A **Vocabulary Type** defines the specific name for the nodes of that type, and the metadata fields that compose each node. A **Node Attribute Type** represents the metadata fields that can be used in one or more **Vocabulary Types**. They define the name of each metadata field, the type of field (e.g. text area, checkbox, etc.) and the mandatory status of the field (defines if the field is mandatory or not).

The Configuration View is available only to 'Admin' Users and allows the definition of **Vocabulary Types** through the Vocabularies External Setup menu[4] (**Figure 26**):

1. It allows the definition of a new **Node Attribute Type** by defining its 'Name', 'Type', and whether the node attribute type is mandatory or not ('Required'). If a **Node Attribute Type** is selected in 2, then you can use this menu to edit the selected **Node Attribute Type**. The 'Clear' button will deselect the currently selected **Node Attribute Type**;
2. List of defined **Node Attribute Types**. The list can be order by 'Name', 'Type', or 'Required'. **Node Attributes Types** can be deleted using the cross in the last column of the list;

[4] For understandability and due to the dependencies between concepts, we describe the definition of a **Vocabulary Type** from the **Node Attribute Type** to the **Vocabulary Type** itself.

3. Similarly to 1, it allows the definition of a new **Vocabulary Type** by defining its 'Name', 'Node Name' and 'Node Attribute Types'. **Node Attribute Types** are selected by dragging from menu 2 to 3. If a **Vocabulary Type** is selected in 4, then you can use this menu to edit the selected **Vocabulary Type** (as illustrated in **Figure 26**). The "Clear" button will deselect the current selected Vocabulary Type;
4. Similarly to 2, it presents the list of defined **Vocabulary Types**. The list can be order by 'Name', or 'Node Name'. **Vocabulary Type**s can be deleted using the cross in the last column of the list.

Please note that menu 3 also contains a 'Has Concepts?' option that if selected allows multiple Terms to be grouped into a Concept. This functionally allows to define that multiple terms are used to describe the same concept.

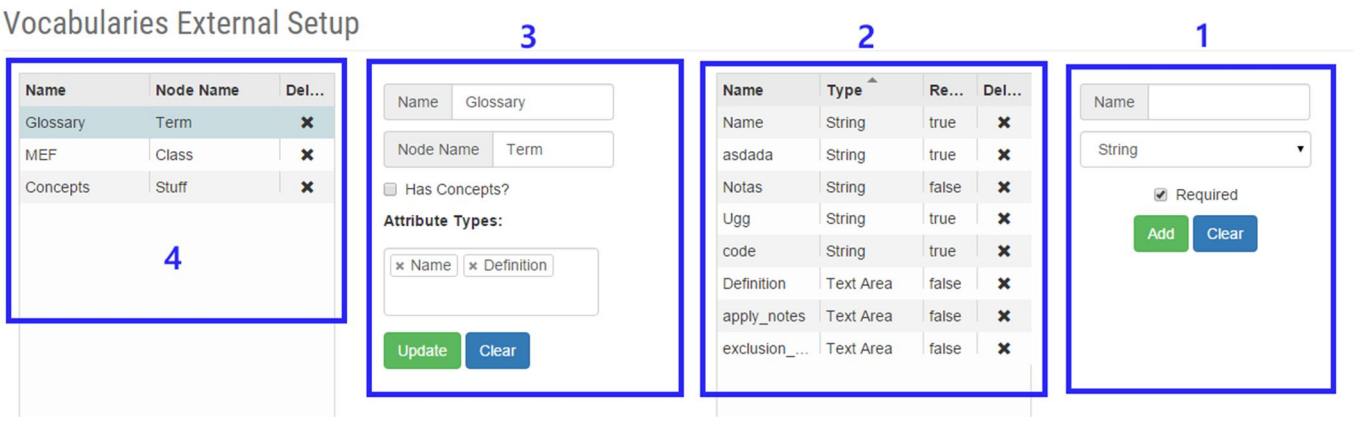


**Figure 26. EVOC Configuration View - Vocabularies External Setup.**

### 3.1.5. Users View

The Users View is available only to 'Admin' Users and allows access to the User Management menu (**Figure 27**):

1. List of registered **Users** in EVOC. The list can be order by 'Username', 'Roles', or 'Email'. **Users** can be deleted using the cross in the 'Delete' column. Users can be validated using the letter button in the 'Validate' column. Validation sends a User a notification email with is username and instructions on how to recover its password if needed;
2. It allows the definition of a new **User** by defining its 'username', 'email', and 'Role' (for more information on user roles please refer to Chapter 3.1.1 User Management). If a **User** is selected in 1, then you can use this menu to edit the selected **User**. The 'Clear' button will deselect the current selected **User**. Please also note that the 'Update' button will change to 'Add' if no User is selected in 1.

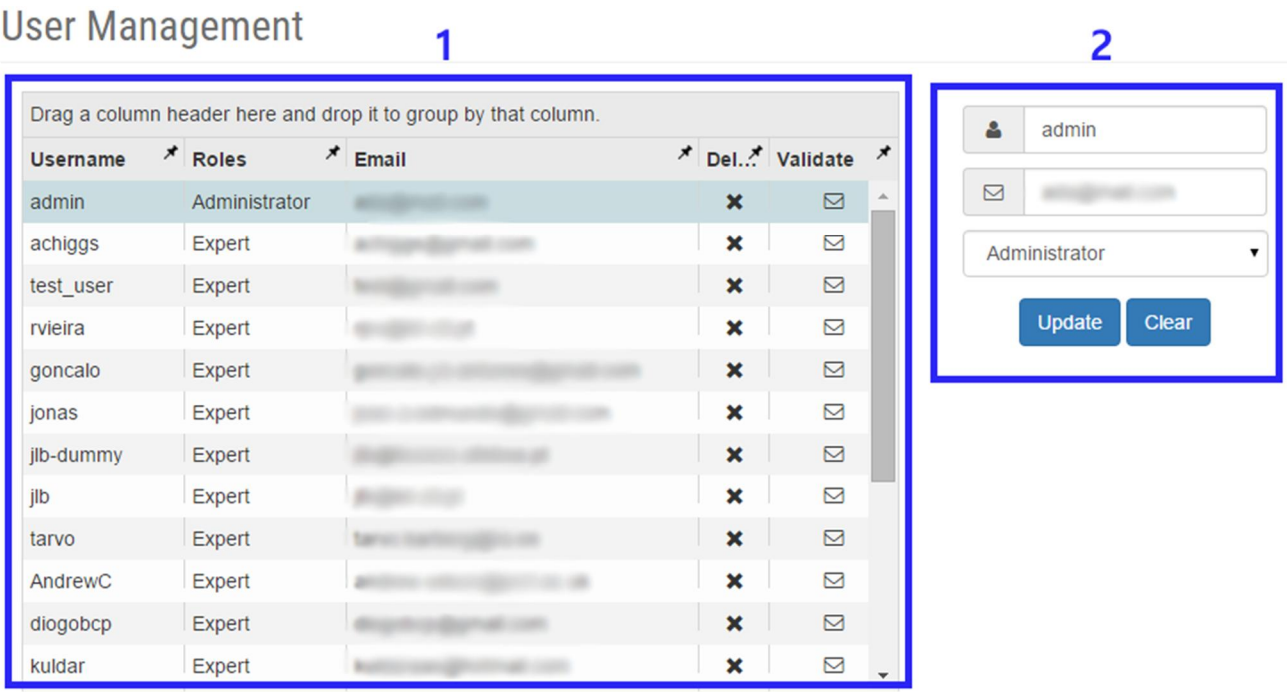


**Figure 27. EVOC User Management menu.**

Users may also be grouped by a common attribute for convenience, which is achieved by dragging a column header to the designated area in the top of the table. For instance, this allows grouping the users by 'Role' (**Figure 28**) to easily find out which users have administrator permissions.

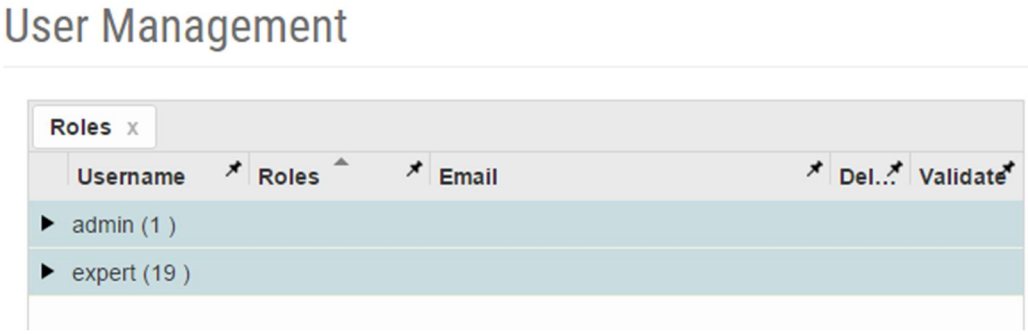


**Figure 28. EVOC User grouping.**

## 3.2. Reference Requirement Manager (REQs)

The Reference Requirement Manager (or REQs) is a service that allow **Users** to view and management reference documents – in this case Information Governance reference documents. The service intends to allow an easy and understandable navigation to the reference documents. Additionally, the service will allow the identification of requirements and the documents so that Users can manage and relate Information Governance requirements. The goal is that users can consult and contribute to the body of knowledge of Information Governance.

The prototype version of REQs, here being described, is mainly composed of two pages:

- The **Home Page**, from where Users can consult the available reference documents in REQs and learn more about the service; and
- The **Overview Page**, where Users can navigate and consult the available reference documents in REQs;

### 3.2.1. REQs Home Page

The layout of the home page of REQs (**Figure 29**) can be described as having three main sections:

1. The top toolbar providing easy access to the features and views of REQs;
2. The list of available reference documents in REQs; and
3. The bottom bar, available only in the home page, providing useful information about the service, its development team and the entities that sponsored and contributed to the development.

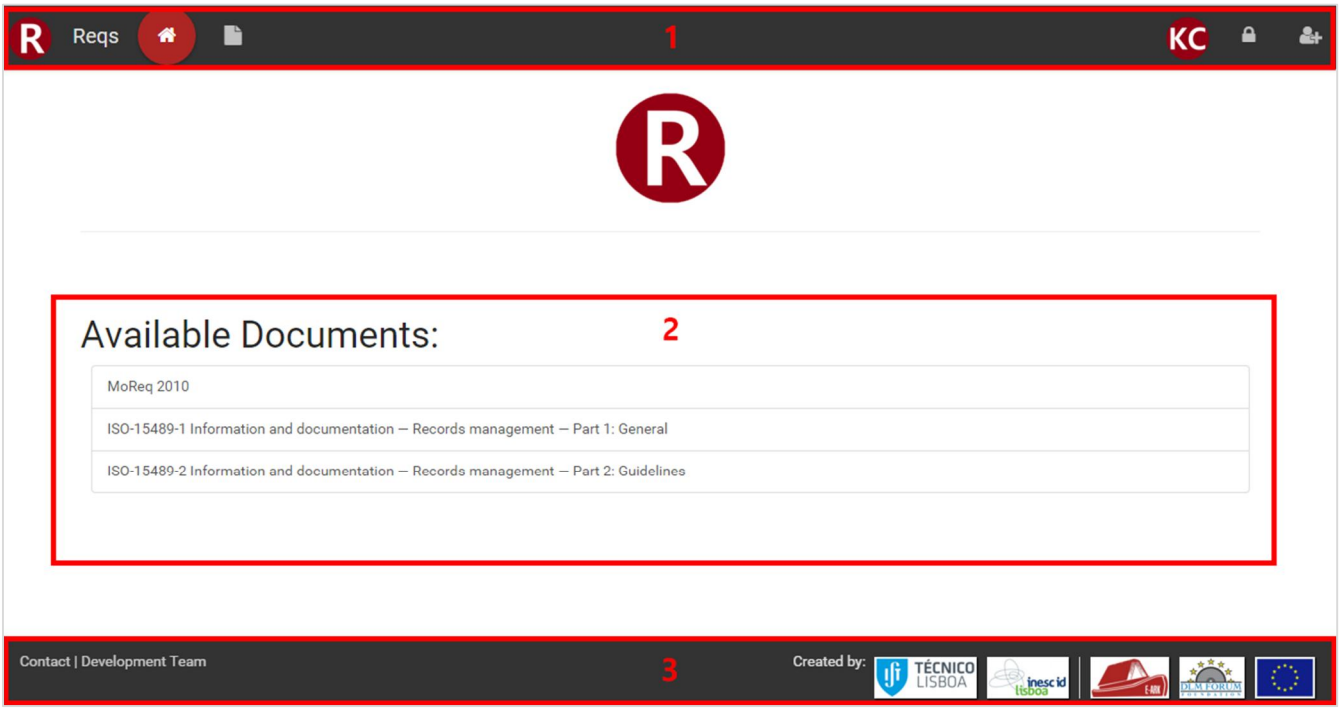


**Figure 29. REQs Home Page.**

While 2 and 3 are self-explanatory, the top toolbar (**Figure 30**) requires additional explanation:

1. The REQs logo and name, allowing the User to quickly go back to the home page;
2. The home page button;
3. The overview page where the available documents can be consulted;
4. The knowledge centre logo, allowing the User to quickly access the Knowledge Centre main page (**Figure 2** in Chapter 2 Knowledge Centre Service);
5. The login button that allows the User to login giving access to management functionalities (still under development in the prototype being described); and

6. The registration button that allows the User to register in REQs (still under development in the prototype being described).

**Figure 30. REQs top toolbar.**

### 3.2.2. REQs Overview Page

The Overview Page (**Figure 31**) is accessible from the application top toolbar (**Figure 30**) or by clicking one of the available documents in the home page (2 in **Figure 29**). It provides means to explore the various reference documents in the database. It is composed by two main panels:

1. The navigation panel from where you can navigate through the sections of the selected reference document; and
2. The content panel where the content of the selected section is displayed.

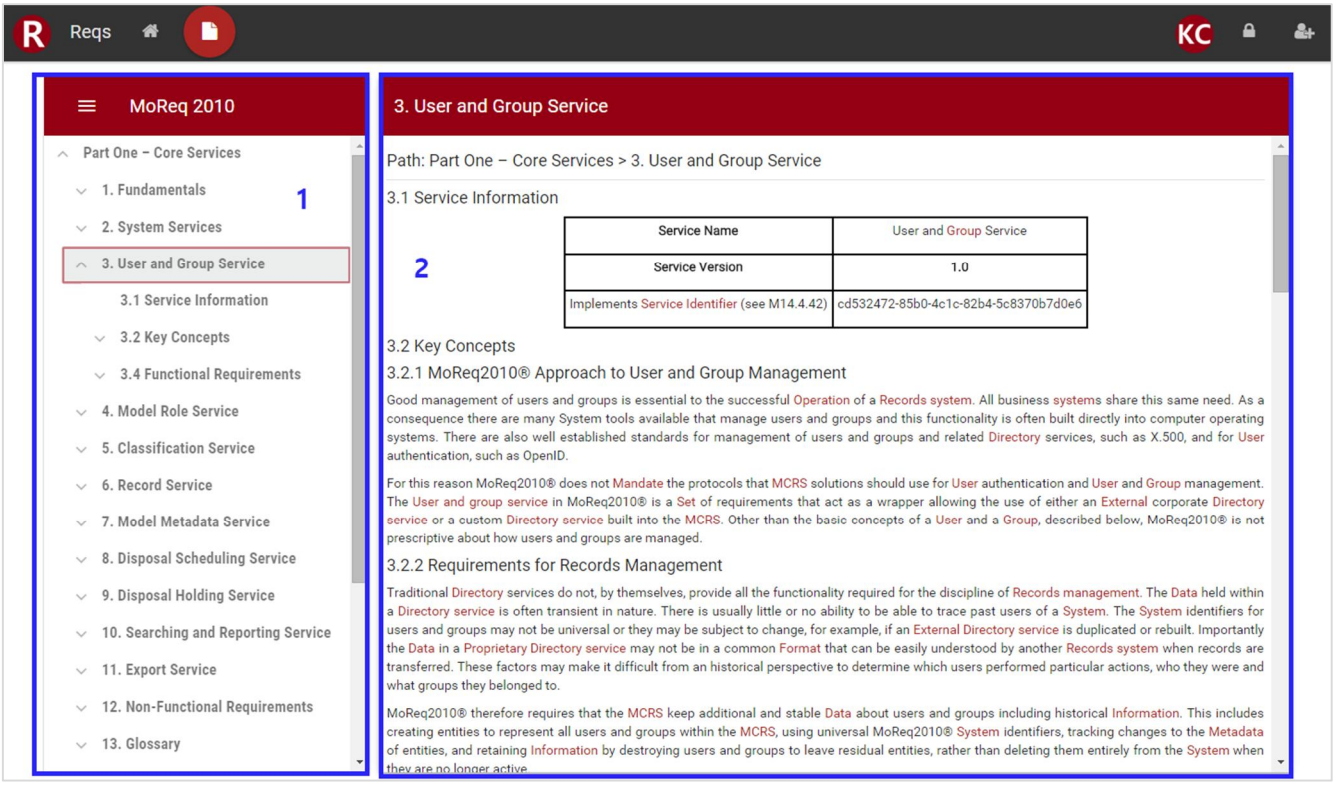


**Figure 31. REQs Overview Page.**

In the navigation panel of the REQs Overview Page (**Figure 32**) the selected documents' sections are shown on a tree-list representation, replicating the selected document index. The sections can be expanded and collapsed, showing and hiding their subsections. Clicking on a section in the navigation panel will cause it to expand or collapse, and it will also select the section consequently loading its contents into the content panel.

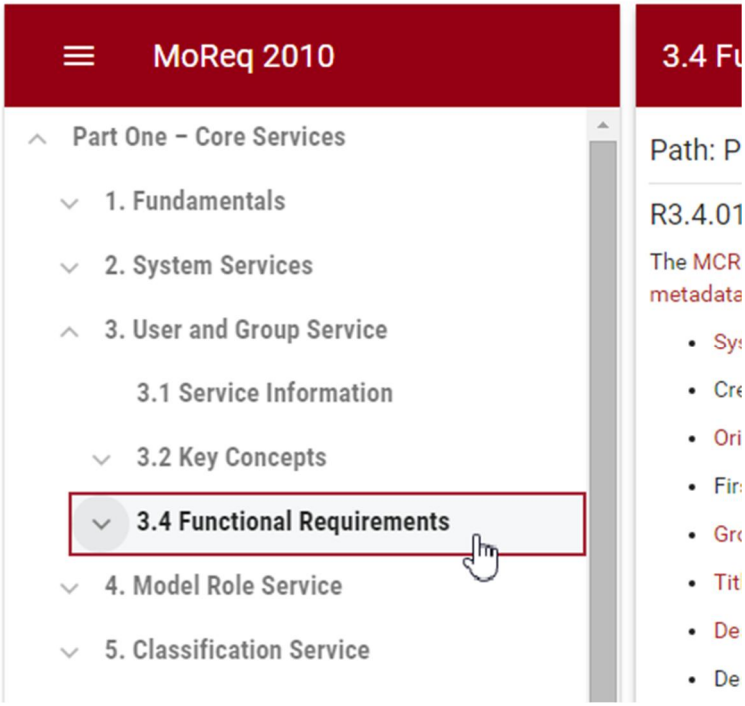


**Figure 32. REQs Navigation Panel.**

The header at the top of the navigation panel has two main functionalities (**Figure 33**):

1. Allow users to quickly switch between the available documents. Selecting 1 will present a menu (**Figure 34**) with all reference document available in REQs; and
2. Present the name of the reference document currently loaded in the REQs Overview Page.

**Figure 33. REQs Navigation Panel Header.**

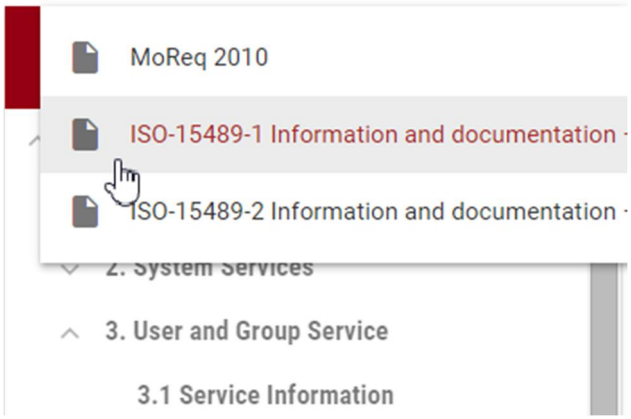


**Figure 34. REQs Navigation Panel – document selection.**

In the content panel the selected section of the reference document is displayed (**Figure 35**). The panel presents at all times:

1. The selected section title;
2. The path (composed of sections and subsections) to the selected sections; and
3. The selected section textual content.

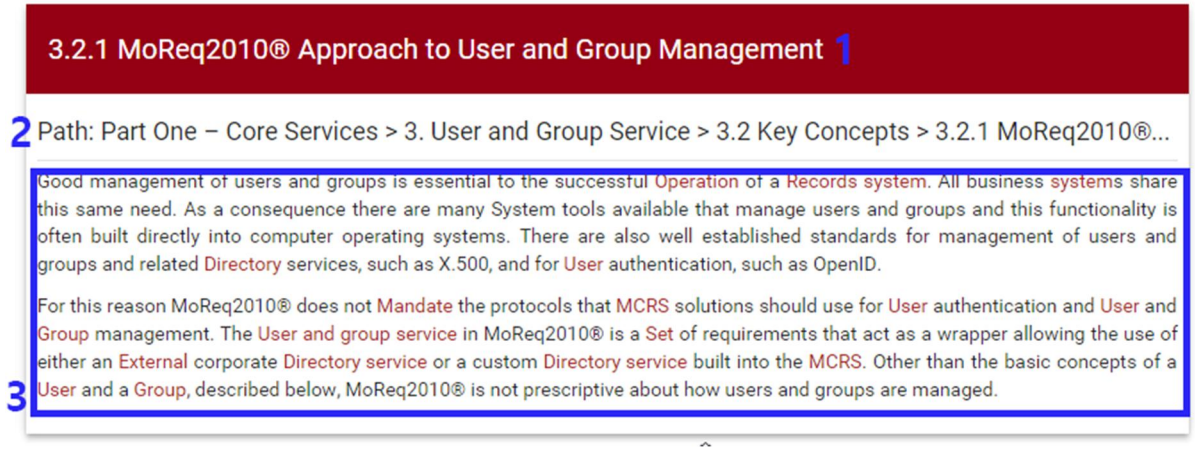


**Figure 35. REQs Content Panel.**

Note that if a User selects a high-level section in the navigation panel, the content panel will display the content of the selected section and its respective sub-sections (**Figure 36**). To keep the loading time of large high-level sections short, a dynamic loading technique was employed. Therefore, when a section is selected, only a predetermined size of its subsections' content is loaded to the content panel. As the User scrolls down in the content panel the remaining content will be loaded.

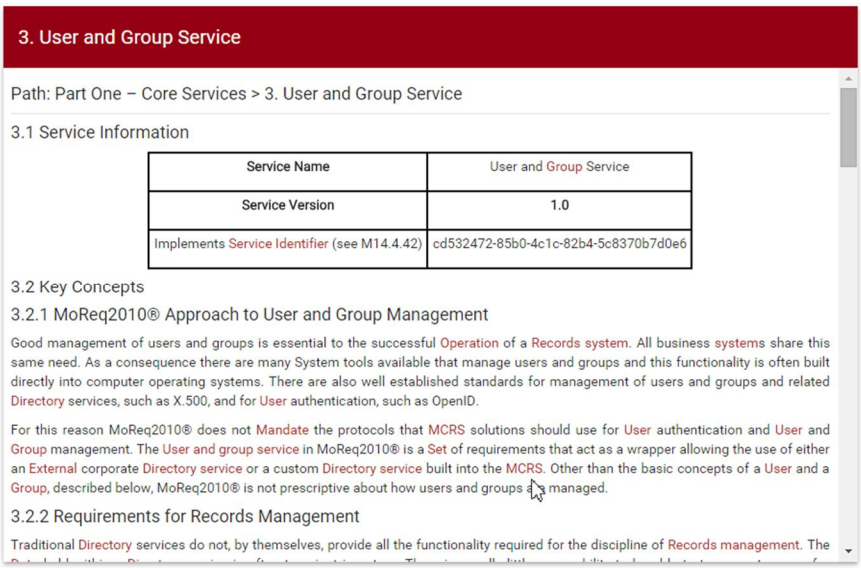


**Figure 36. REQs content panel with multiple sections.**

The selected reference document can also be navigated through the existing links the content panel. The content of a document section can have linking information attached to any part of its text, which allows the user to swiftly navigate between linked sections. As an example, if the user clicked in the link highlighted in **Figure 37** the content panel would reload to show the contents of the link's target (**Figure 38**). Also note that the current document and section selected are also reflected at all times in the URL of service.

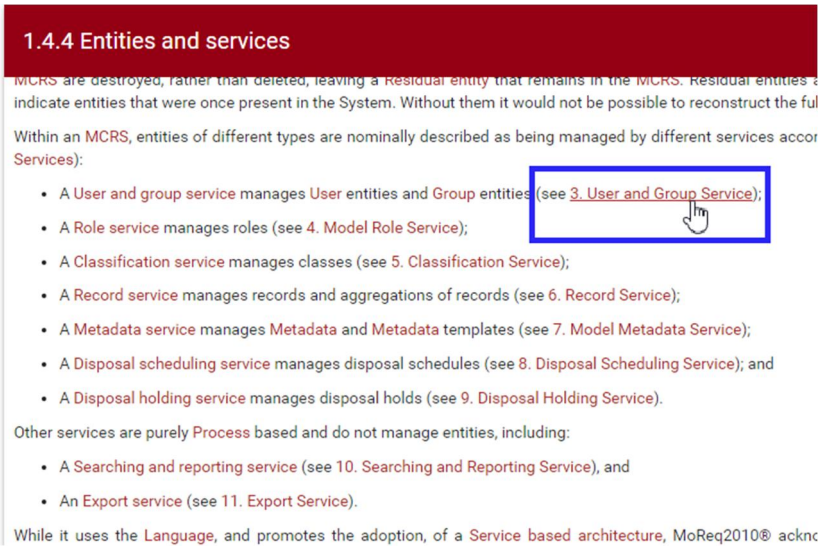


**Figure 37. REQS Content Panel links.**

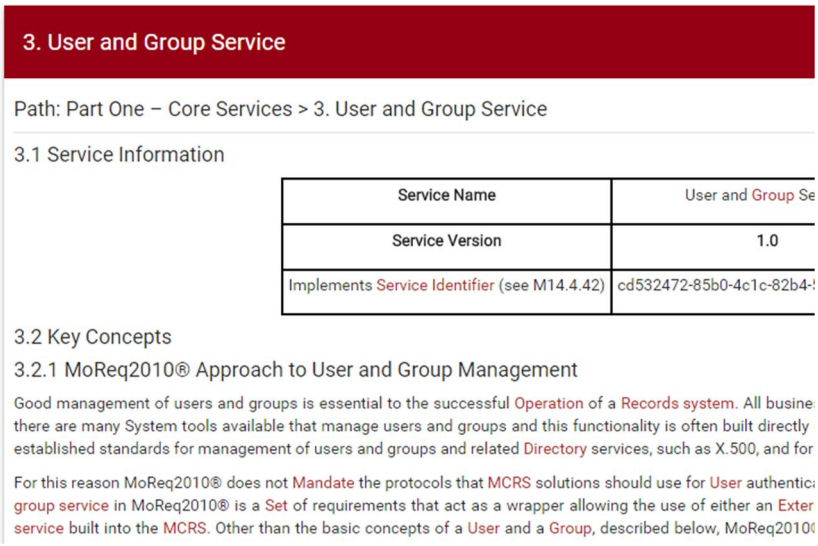


**Figure 38. REQs Content Panel link target.**

It is important to understand that a section in REQs does not necessary imply a chapter in the reference document. In fact, it is possible to also define and represent requirements as sections as illustrated in **Figure 39.** When as section is classified as a requirement the service also displays the relation of the selected requirement to other requirements (3 in **Figure 39**) - apart from the already mentioned requirement title or identifier, path (1 in **Figure 39**) and content (2 in **Figure 39**).

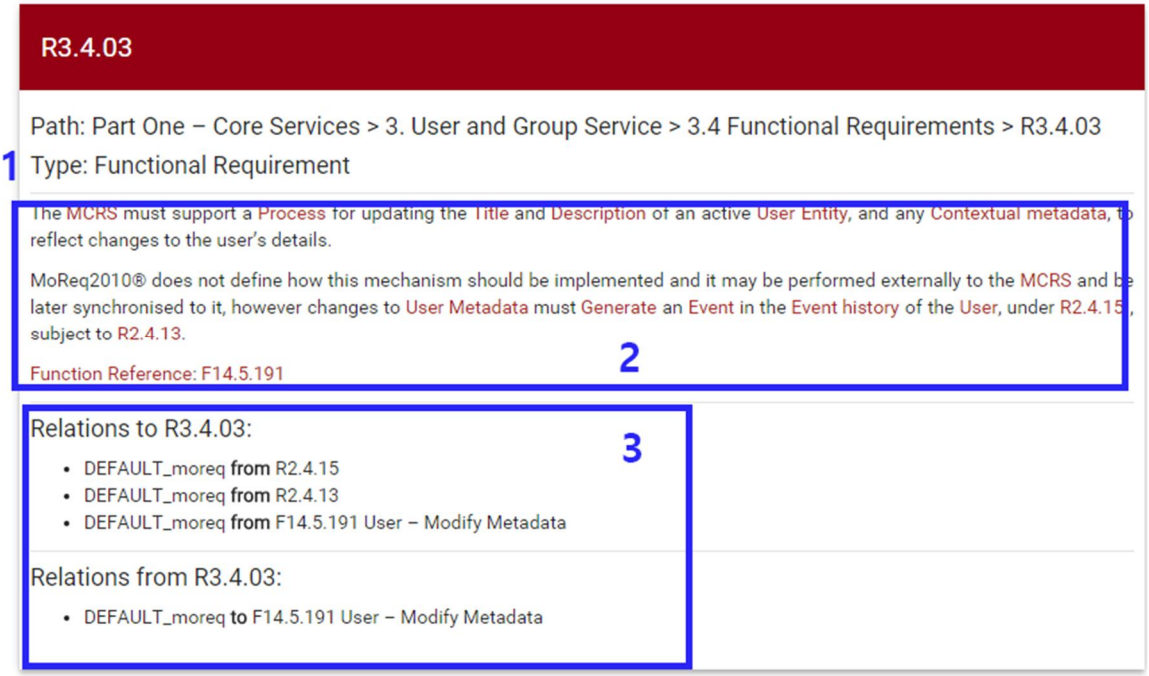


**Figure 39. REQs Requirement layout.**

## 3.3. Data Store

This section describes how data is stored in both Resources Centre services: EVOC and REQs. Both services data are stored in MongoDB databases (more technology details in the **Appendix A – Technological Specifications**).

### 3.3.1. EVOC

The data domain model of EVOC is presented in **Figure 40**. The Vocabulary Class is central to the service data' structure as it connects the remaining Classes. A Vocabulary needs to be classified under a Space. Additionally, it can also be classified under a Group. Space have a one or more Users associated representing those who can access and edit that space and the vocabularies within.

Vocabularies have associated Nodes that define the contents of each vocabulary. A node can belong to more than one vocabulary. Vocabularies have a Vocabulary Type associated which determines the name of the nodes associated with the vocabulary and what Attributes those nodes can have. This, and the other characteristics of a node attribute can have, is given by the associated Attribute Type. Nodes also have a language associated with them.

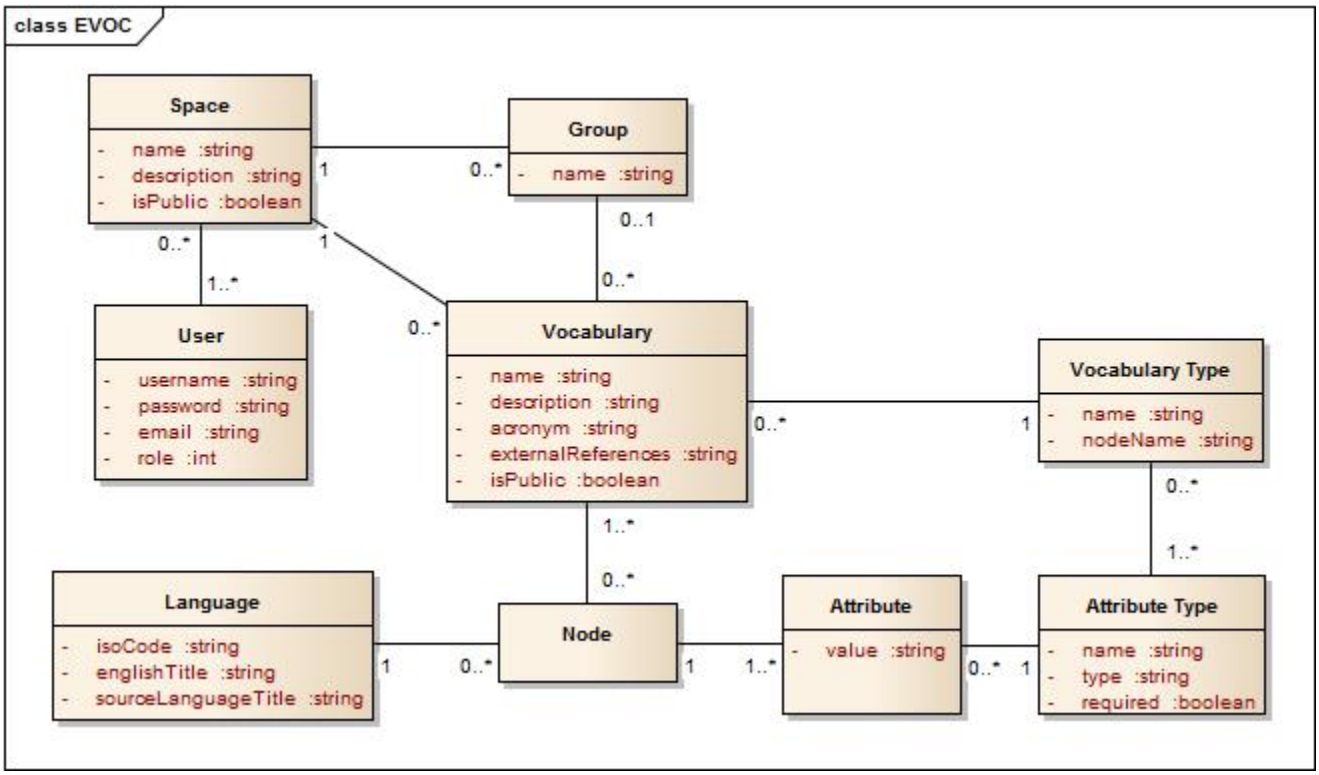


**Figure 40. EVOC Data Domain Model.**

### 3.3.2. REQs

The data domain model of REQs is presented in **Figure 41**. All Documents have an associated Structure. To better capture all the details inherent to a document, this structure is broken down, up to small text parts. First there are the Sections, which can have other sections inside and contain Blocks. These blocks can be Images, Paragraphs, Bullet Lists (which are

made of paragraphs) or Tables (which are made of Table Elements, or cells, that can contain other blocks). Paragraphs are also divided into Sentences and those into Text parts, which can have different formatting attributes associated with them, like being in bold or having a Link associated with them.

Also, note that a section may have a *specialType*, which indicates that section has additional meaning to the application (for instance, a functional requirement of the document). The application can then take this information and act accordingly, displaying or offering special interactions with the section.

Finally, from the sections to the texts, all parts of a document's structure are called Components. The generic property of a component, is the fact that it knows who its children components are and the order in which they are supposed to appear. This is essential to show the components in the correct order and also, it allows to form a unique URI path to any component of the application (starting with a document's *uri_base* attribute).

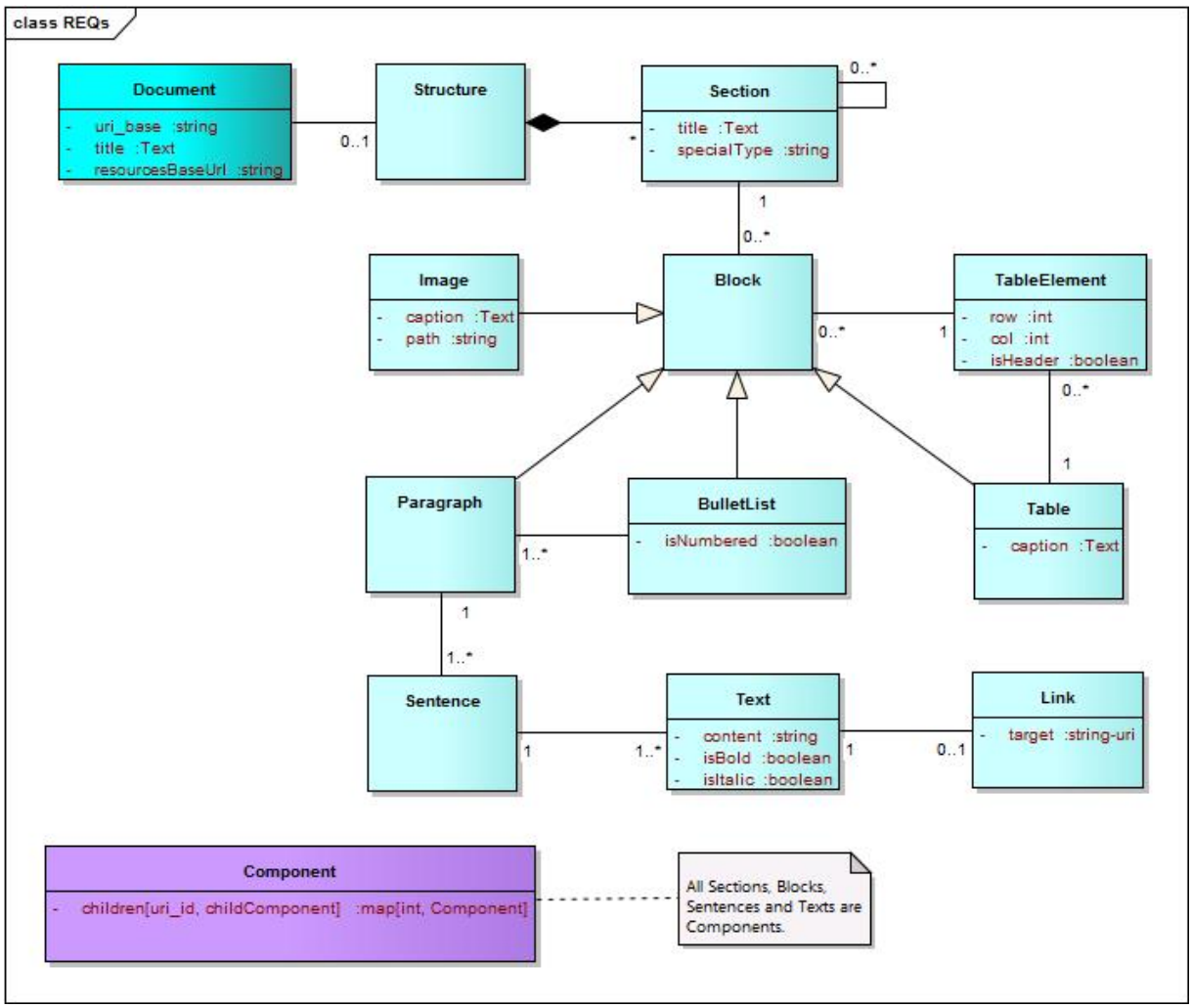


**Figure 41. REQs Data Domain Model.**

# 4. Assessment Services

The Assessment Services (**Figure 42**) is the component of the Knowledge Centre containing the services for assessment and validation of existing systems and businesses. Currently, two prototype services have been developed to achieve these objectives: The MoReq[5] Schemas Validator, which implements a specific case of schema validation, and the Maturity Assessment questionnaire. The following sub sections describe both assessment services.

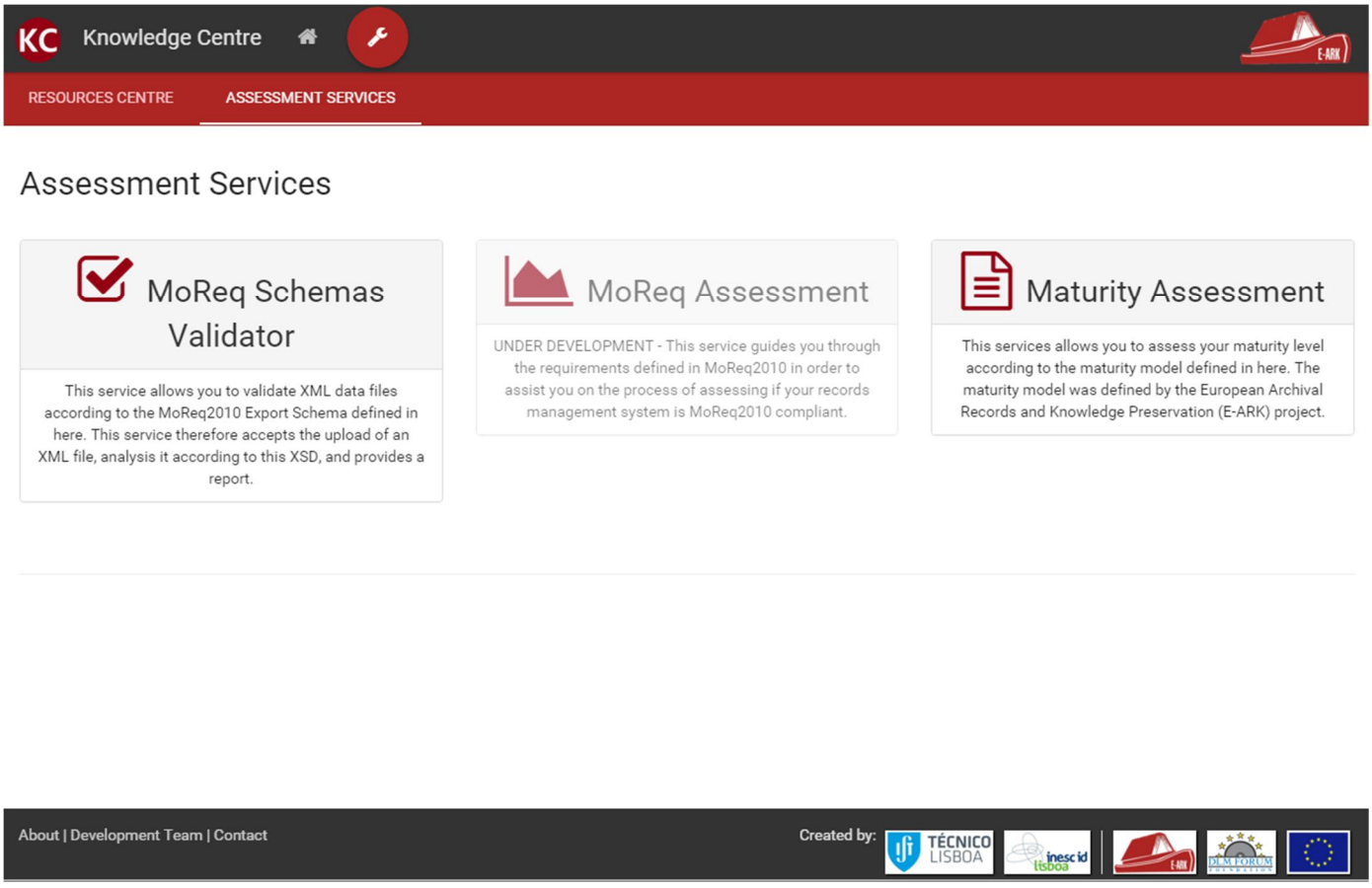


**Figure 42. Assessment Services Page**

## 4.1. MoReq Export Validator Service

The MoReq Export Validator Service was developed with the purpose of assisting records management developers to validate their MoReq2010 Compliant Export Services. According to MoReq2010, an authorised user can export entities in a MoReq Compliant Records Systems (MCRS) to an XML[6] data file that can be validated using the MoReq2010 XML Schema[7]. Therefore, the MoReq Export Validator Services implements an automatic process that validates data files

[5] MoReq2010, Modular Requirements for Records Systems, Volume 1, Core Services & Plug-in Modules, Version 1.0. More information at http://moreq.info/.
[6] The Extensive Markup Language (XML) is a markup language which establishes a set of rules for encoding documents in a format that is both human and machine readable.
[7] Requirement R11.4.1 from Section 11 – Export Service in MoReq2010.

according to the MoReq2010 XML Shema[8]. The mentioned schema is defined as an XSD[9] that defines the structure of how entities in a MCRS must be exported.

The main layout of the MoReq Export Validator Service can be seen in **Figure 43**, and is described in the following list:

1. A set of MCRS exports examples that can either be run or download for further analysis;
2. The validation form that allows the user to perform the validation through the controls describe below;
3. The 'Browse" button, which triggers a dialog box that allow users to select an XML file to be submitted for validation;
4. A text box where the selected XML File' name will be displayed;
5. The "Submit" button starts the validation of the selected XML File;
6. A checkbox which allows the user to select if the results of the validation are to be displayed in the same page or in a new window.

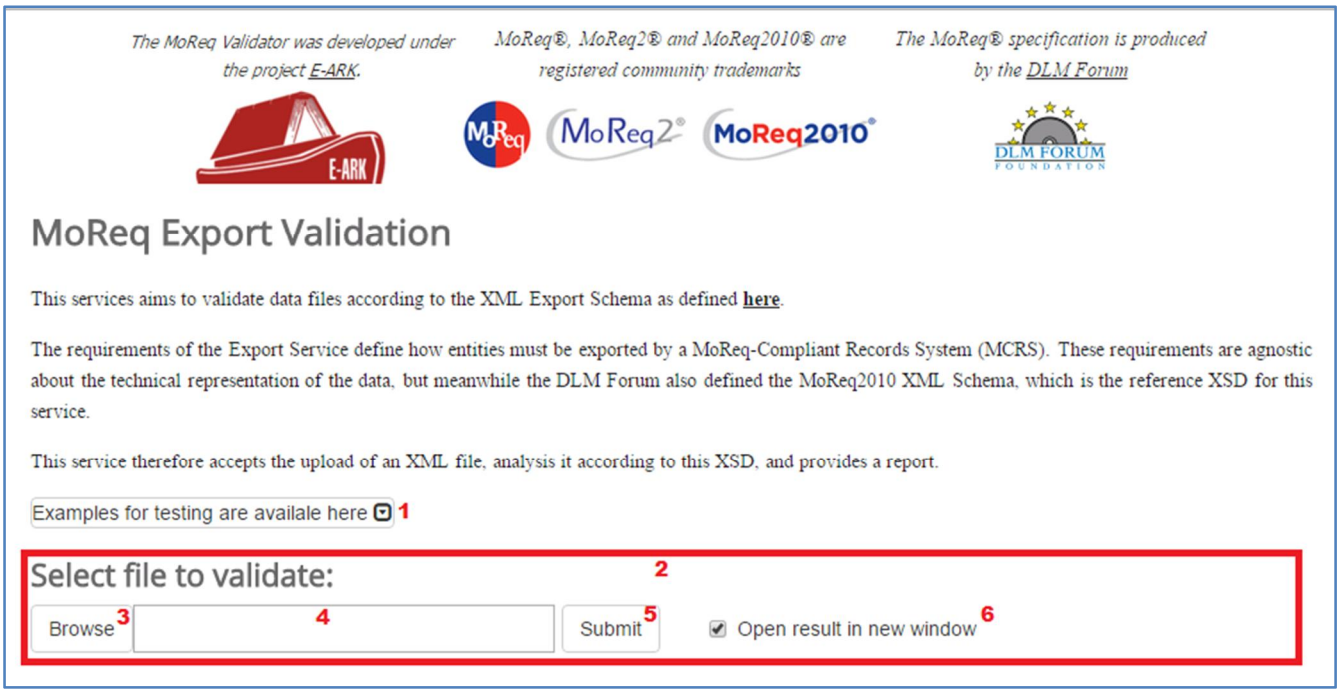


**Figure 43. MoReq Export Validator Service main layout**

To facilitate understanding of what constitutes a MoReq2010 Compliant Export XML File the service provides a set of valid and invalid examples. Those examples can be accessed by selecting 1 in **Figure 43** resulting in the panel presented in **Figure 44**, where the user is presented with:

1. A short description of the example case;
2. The possibility of downloading the selected example in order to analyse it; and

[8] The MoReq2010 XML Schema is available at http://moreq.info/files/export/export.xsd.
[9] The XML Schema Definition (XSD) is used to express rules to which XML document must comply so as to be considered valid according to the schema which is being used to perform the validation.

3. The possibility of performing a validation of the selected example.

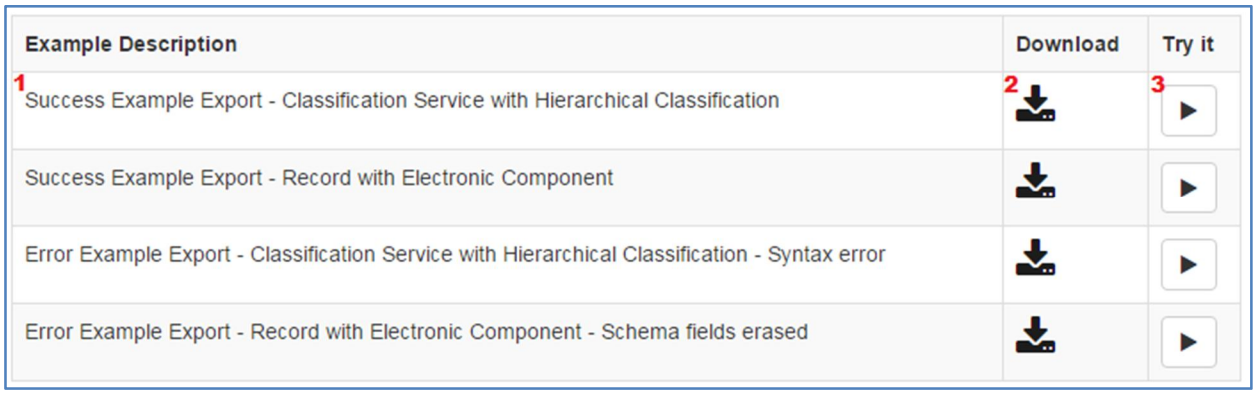


**Figure 44. Examples panel layout**

Once the system has performed a validation on an XML file, a report is automatically generated (**Figure 45**) with the following structure:

1. The Print Report button allows users to get a printable version of the validation report;
2. The Get as CSV button allows users to download a version of the validation report in .csv format;
3. The Close Window button closes the window in which the validation report is being displayed;
4. The File Properties section focuses on the imported file properties, specifically in its file name, number of lines and size;
5. The Errors Report section lists any XSD validation errors that may have occurred during validation, providing details regarding the line and column number and message of each error (see **Figure 46**). An error can be:
    - Fatal Error: if the XML contains entities that do not belong and consequently disrupt the mandatory structure required;
    - Error: if a mandatory element is missing in the XML;
    - Warning: if optional elements are missing (for example, no "<!DOCTYPE …>" defined or references to an undefined parameter entity);
6. The Structure Analysis Report section displays a simple structure analysis based on the MoReq2010 Export Schema. It simply signals if a given element is present on the file that was subjected to validation.

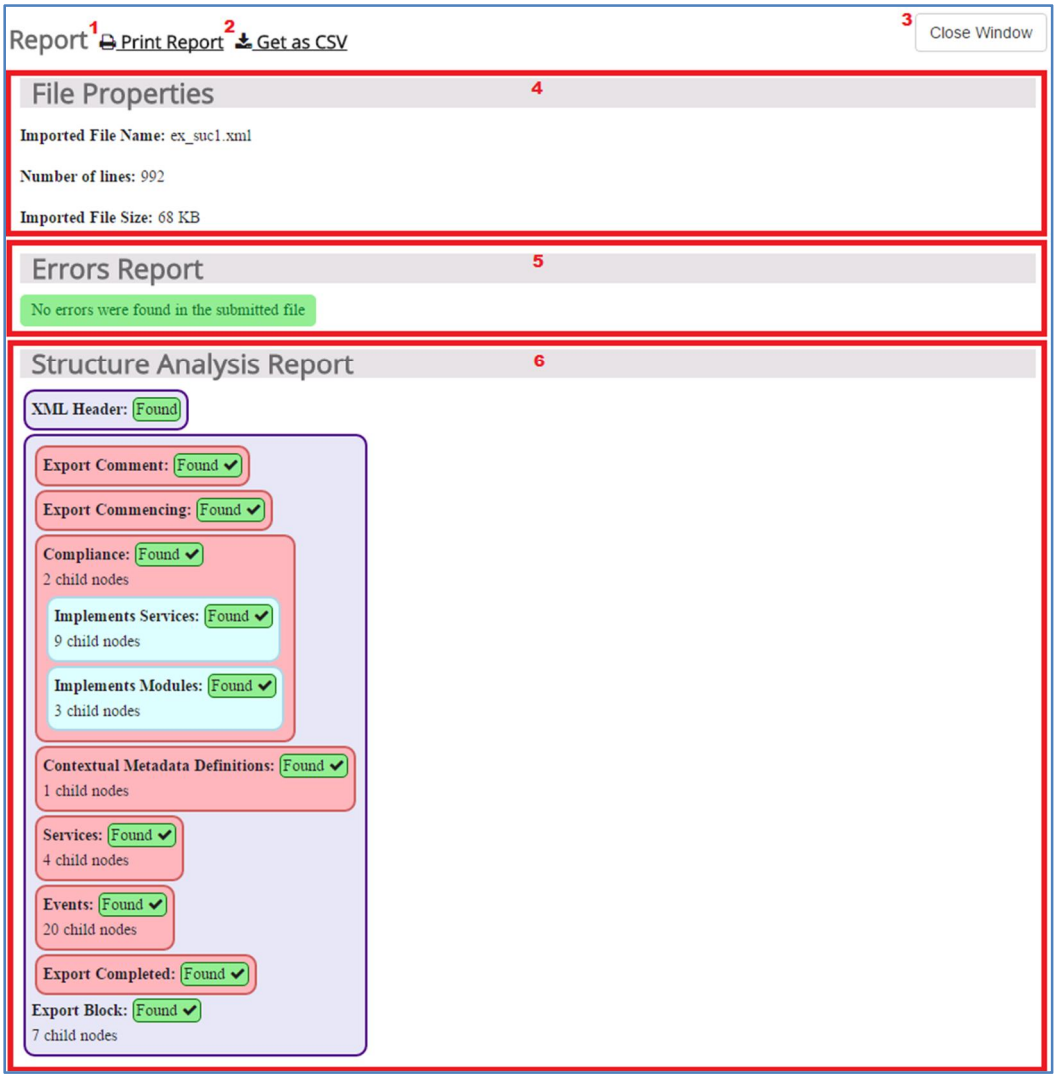


**Figure 45. Validation Report layout**

| Line Number | Column Number | Error Message |
|---|---|---|
| 20 | 14 | Error: cvc-complex-type.2.4.a: Invalid content was found starting with element 'Compliance'. One of '{"http://moreq.info/files/export":ExportCommencing}' is expected. |
| 643 | 4 | FATAL ERROR: The element type "Export" must be terminated by the matching end-tag "". |

**Figure 46. Example of a list of errors that occurred during XSD validation**

## 4.2. Maturity Assessment

The maturity assessment was developed within the Knowledge Centre as a questionnaire which consists of a set of questions that were used to determine the maturity level of the E-ARK pilots for each of the five capabilities of the E-ARK General Model detailed in D7.2. All questions are mandatory.

The questionnaire is presented in a set of five tabs, one for each of the capabilities being assessed (shown in **Figure 47**). In each tab a short description of the capability is presented followed by the:

1. **Question:** Which details the question itself;
2. **Objective:** Which details the objective of that question, i.e. what knowledge the question intends to capture;
3. **Notes:** Which either clarifies some aspects and/or terms of the question or details examples of evidence to substantiate the answer for the question;
4. **Terms:** Which identifies the terms that are detailed in EVOC;
5. **Answers:** Which depicts the five possible answers to the question;
6. **Comment:** Which allows the respondents to provide additional comments, clarifications or a justification to the answer. These comments are considered by the assessment team when evaluating the answers.

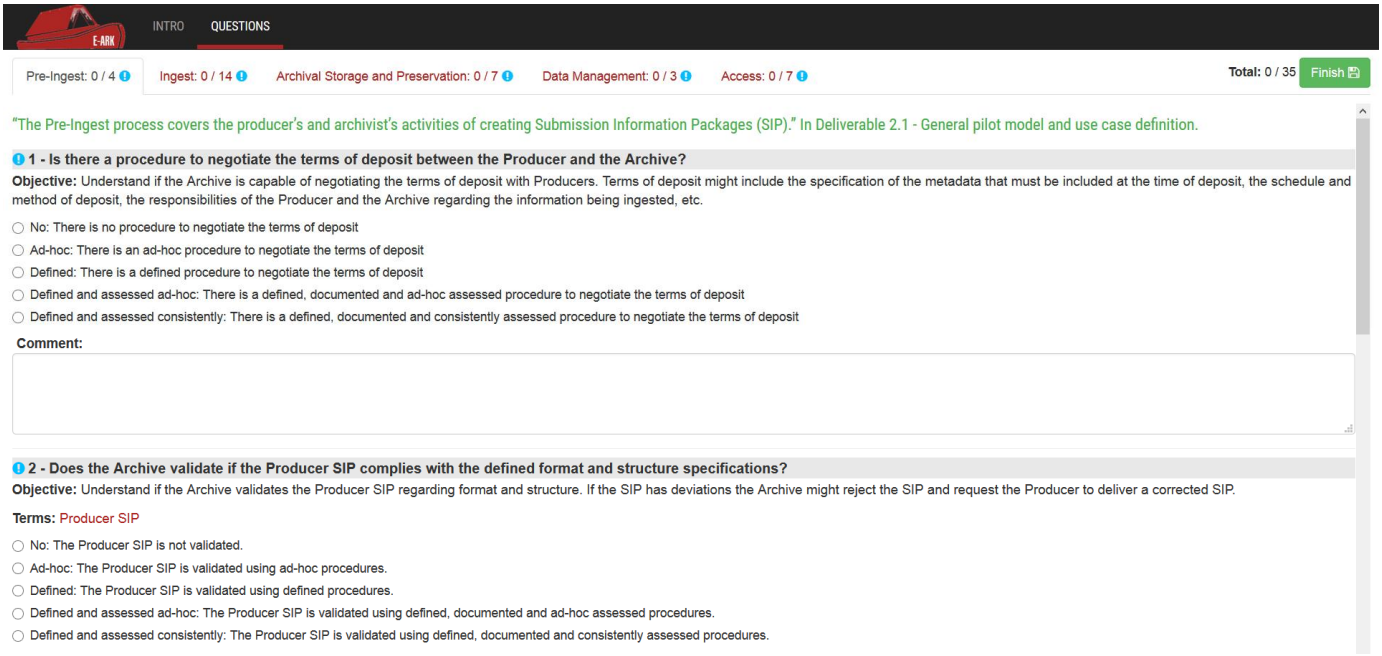


**Figure 47. Maturity Assessment Service**

The questionnaire uses the following definitions of measurement in the answers:

- **No** indicates that there is no procedure or mechanism in place;
- **Ad-hoc** refers to actions performed but not being repeatable in the future, which can be due to the lack, outdate or no use of proper documentation, procedures or mechanisms, and thus leading to different people performing different tasks to achieve the same outcome;
- **Defined** refers the ways to achieve an outcome are supported by defined procedures or mechanisms, and thus leading to the actions performed being capable of being repeated in the future. This level does not give an assurance that the defined procedures or mechanisms are being consistently complied with or assessed;
- **Ad-hoc assessed** means that there is a concern with the assessment of some specific aspects, but that is not performed under a defined process but ad-hoc and when the need arises**;**

- **Consistently assessed** means that there is a concern with the assessment of some specific aspects, and that such is performed continuously, under a defined process, with alerts triggered by a defined set of indicators considering these dimensions, for example:
    - **Completeness**, which focuses on assessing if a procedure performs all relevant steps, aligned with the most recent documented requirements for that;
    - **Effectiveness**, which focus on assessing if the results of a procedure are free of errors and do not require further handling;
    - **Efficiency**, which focus on assessing if a procedure executes with the optimal efforts (for example, if automation is used instead of human effort), in an agreed time period as to avoid bottlenecks on the infrastructure and to minimize the time spent on executing it;
    - **Relevance**, which focus on assessing if the implemented requirements are still relevant for the intended purpose (as legislation change, for example, there is the need to assess if implemented requirements are still relevant).

These are just examples of aspects that need to be measured at higher levels of maturity, there might be further aspects to measure depending on the specific requirements of the pilot.

In the initial assessment and evaluation, the answers provided by the pilots were analysed by the Information Governance Maturity Model development team and a report was issued detailing all the findings of the assessment. The set of assessment reports is part of deliverable D7.2. For the final assessment and evaluation (D7.6) there will be a new process for calculating the final maturity level for the pilots. In the initial assessment the process was performed manually by gathering all the results from the pilots and analysing them manually through a spreadsheet. In the final assessment the Knowledge centre will automatically generate the final maturity levels in order to automate the assessment process, providing detailed graphs of the results as shown in the examples provided by **Figure 48**.

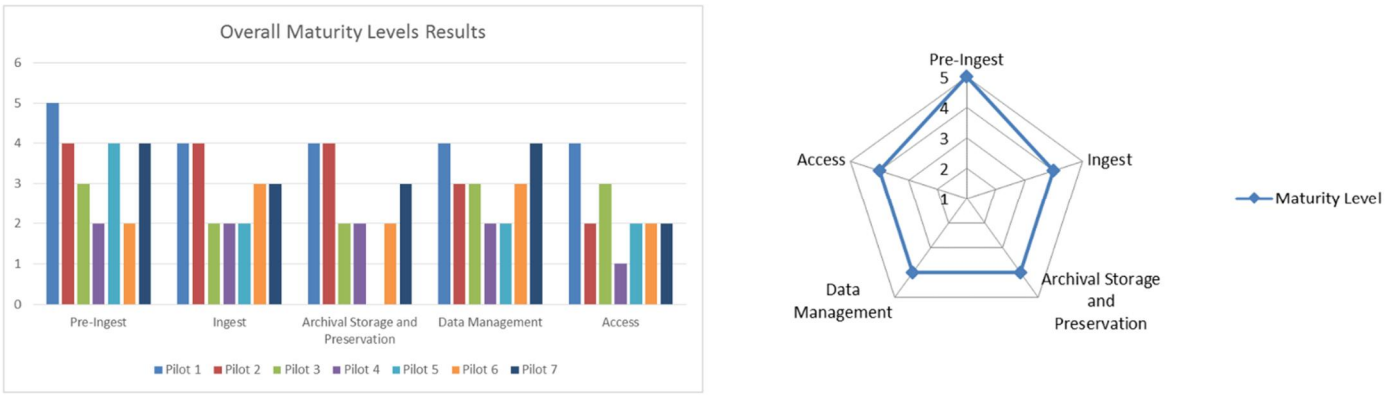


**Figure 48. Examples of Maturity Assessment Results.**

One other aspect to take into consideration is that although the initial assessment and evaluation only focused on assessing the E-ARK pilots' maturity, the final version will be publicly available and will allow the assessment of the Information Governance Maturity Model by any organization through the Knowledge Centre.

# 5. Future Work

The current version of the Knowledge Centre contains the essential services required to have a fully-functional release of the Knowledge Centre Service. This prototype serves two main purposes: (1) to allow Users to get a look and feeling of the service, and (2) to test the quality of current services particularly regarding aspects such as usability and efficiency that require user feedback. In order to achieve the former, E-ARK Project will promote the Knowledge Centre through newsletters, conferences and publications. Additionally, members of the E-ARK's advisory board will be invited to test it and give feedback.

Nonetheless, the Knowledge Centre still has functionalities and services under development and planned for the final year of the project that are described in the next sections.

## 5.1. MoReq Assessment

The main issue regarding the assessment of a certain system's capabilities is the way the system itself is built. Even though there are standards and guidelines that mitigate this issue, if two entities attempt a solution using the same functional requirements, there are business and logic decisions during the planning process that might alter the final product and result in two completely different systems with completely different capabilities.

The main idea behind the service here described is to provide an expert user – one with a background of certain experience in a particular system – with a portal that contains a set of questions that require evidence base answers in order to establish a level of compliance between the standard and the system. In this particular case of the MoReq Assessment service, the system under assessment should be a records management system and the standard (de facto) is MoReq2010.

In more detail, the service requires at least two actors: The Vendor Representative and the Expert. The former will represent the system under assessment and will be responsible to browse the reference document under assessment and submit evidence of compliance with the requirements of said document. The latter will represent the standard (reference document) expert that will be responsible for analysing the evidence and create a report detailing the level of compliance of the system with the reference document. **Figure 49** illustrates the use cases of the proposed services.

The "browse standard" functionality will be fulfilled through the REQs service that will provide the necessary navigation capabilities to the reference document being used for compliance. The "submit proof of compliance" functionality will allow Users to upload any type of files that they wish so such as images, documents and videos, as well as textual comments to backup or further explain the uploaded evidence. The user might repeat this process as many times as necessary. The service will allow the user to notify when the process of submitting evidence is complete. Only then, the expert user will be notified that there are evidence ready for analysis.

To the expert user the service will need to provide necessary functionality to (1) analyse, comment and provide feedback based on the submitted evidence, and (2) summarize all expert comments and feedback in an easy and understandable compliance report.

As mentioned, the first version of this service will used MoReq2010 as the reference document from which records management system vendors can submit compliance evidence. The reason behind this choice is that (1) it allows us to

re-use the work already done around MoReq2010 in REQs, and (2) the service will be able to replace the current MoReq2010 certification process of the DLM-Forum[10].

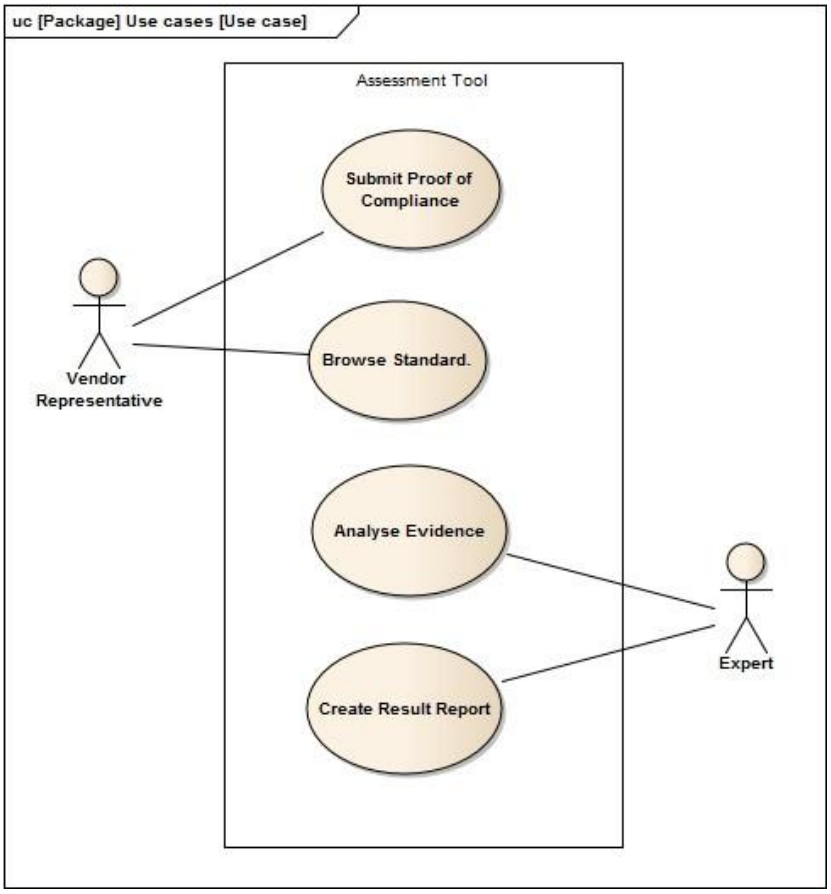


**Figure 49. MoReq Assessment Use Cases**.

## 5.2. Existing Services Improvements

The release of a fully-functional prototype of the Knowledge Centre will allow, among the aforementioned benefits, to start the testing of the services and discover potential bugs or faults missed during the development process. Those will be corrected as discovered. In fact, since the services here described were released one by one at different times they have different levels of maturity. For example, the EVOC service was released at the end of the first year of the E-ARK project, and therefore it is already the result of feedback from organizations that tested it and it is already currently used by, for example, the National Archives of Hungary and the National Archives of Portugal.

The following list represents the main functionalities under development or planned to be released in the third year of the E-ARK project:

[10] The current MoReq2010 certification process is described at http://moreq.info/index.php/testing-and-certification

- The functionality of relating terms between them is going to be added to the EVOC service. This will allow us, for example, to define that a specific term is a specialization, generalization, similar, or equivalent to another. Additionally, a functionality to merge terms in different vocabularies is also planned;
- As mentioned before, the REQs service currently only has viewing functionalities. These will be expanded with new views for different document types, and a new set of features for creating and editing new documents and requirements (or other special entities) has been planned;
- An API to access and use the EVOC and REQs services in other systems is partially complete, lacking some functions and generally needing a detailed documentation to specify them;
- The MoReq Export Validator Service will be expanded to incorporate other schema validations namely the validation of XML files according to the E-ARK SIP, AIP and DIP schemas;
- As mentioned, the Maturity Assessment is currently lacking the functionality of automatically generating reports based on the answers of the questionnaire;
- Finally, a centralized authentication service is under development since in the current version each service has an independent login. The authentication service will allow to define different types of users with different permissions and will be integrated with the DLM-Forum webpage authentication service preparing already the deployment of the Knowledge Centre to the DLM-Forum infrastructure[11].

[11] As planned in the DoW, after the end of the E-ARK project, the DLM-Forum will assume control and responsibility over the Knowledge Centre and will continue to maintain and develop the existing services.

# Appendix A – Technological Specifications

The Knowledge Centre was developed using the JavaScript programming language[12]. The main technology stack, and the current versions, used across the services is (**Figure 50**):

- Node.js[13] (v0.10.33) – the execution environment for the server-side application.
- MongoDB[14] (v2.6.10) – open-source document database. Interfaced with using the Mongoose ODM[15] (v4.3.4).
- Express.js[16] (v4.5.1) – the web application framework that runs on Node.js.
- Angular.js[17] (v1.4.8) – the JavaScript framework for the client-side application.

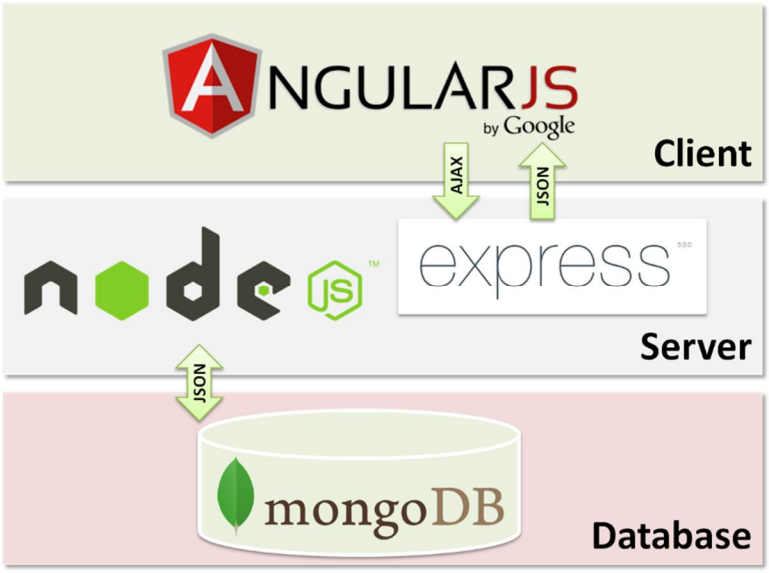


**Figure 50. Base technology stack.**

This technology stack follows the MVC[18] pattern. Mongoose is used to build the data model to be stored in the MongoDB (server-side), while Angular.js takes care of the views and controlling them (client-side). The bridge between the server (model) and client (view) sides is controlled by Express.js, while Node.js gives the foundations to build the application upon. The project libraries are managed using NPM[19] (server side) and Bower[20] (client side) package managers.

Using Angular.js as the base framework, some important libraries were used to develop the user interface for the front-end of the Knowledge Centre services:

[12] https://www.javascript.com/
[13] https://nodejs.org/en/
[14] https://www.mongodb.org/
[15] http://mongoosejs.com/
[16] http://expressjs.com/
[17] https://angularjs.org/
[18] https://en.wikipedia.org/wiki/Model%E2%80%93view%E2%80%93controller
[19] https://www.npmjs.com/
[20] http://bower.io/

- Bootstrap[21] (v3.3.6) – styling framework for responsive applications.
- Font Awesome[22] (v4.5.0) - application icons toolkit.

Additionally, the following notable libraries were used in specific services of the Knowledge Centre.

**EVOC and Maturity Assessment**:

- js-xlsx[23] (v0.8.0)– Excel spreadsheet parser and writer.
- socket.io[24] (v1.4.4) – real time engine.

**EVOC:**

- node-jsonwebtoken[25] (v5.5.4) – JsonWebToken[26] implementation for node.js.

**REQs:**

- mongoose-schema-extend[27] (v0.2.0) – mongoose inheritance extension.
- mongoose-deep-populate[28] (v2.0.3) – mongoose plugin to enable deep population of nested models.
- angular-material[29] (v0.11.2) – (front-end) material design implementation for angular.

[21] http://getbootstrap.com/
[22] https://fortawesome.github.io/Font-Awesome/
[23] https://github.com/SheetJS/js-xlsx
[24] http://socket.io/
[25] https://github.com/auth0/node-jsonwebtoken
[26] https://jwt.io/
[27] https://github.com/briankircho/mongoose-schema-extend
[28] mongoose-deep-populate
[29] https://material.angularjs.org/latest/