An Integrated Semantic Approach to Content Management in the Urban Resilience Domain

Ilkka Niskanen, Mervi Murtonen
Technical Research Centre of Finland
Oulu/Tampere, Finland
Ilkka.Niskanen@vtt.fi, Mervi.Murtonen@vtt.fi

Francesco Pantisano
Smart System Infrastructures
Finmeccanica Company
Genova, Italy
francesco.pantisano@finmeccanica.com

Fiona Browne, Peadar Davis
School of Computing and Mathematics
Ulster University
Jordanstown, Northern Ireland, UK
f.browne@ulster.ac.uk, pt.davis@ulster.ac.uk

Ivan Palomares
School of Electronics, Electrical Engineering and Computer Science
Queen's University
Belfast, Northern Ireland, UK
i.palomares@qub.ac.uk

***Abstract*— Content Management refers to the process of gaining control over the creation and distribution of information and functionality. Although there are several content management systems available they often fail in addressing the context specific needs of end-users. To enable more task specific and personalized support we present a content management solution developed for the domain of urban resilience. The introduced content management system is extended with a semantic layer that aims to support the management of heterogeneous and large content repository with domain specific annotation and categorization capabilities. In addition, the applied semantic intelligence allows better understanding of content items, linkages between unstructured information and tools, and provides more sophisticated answers to users' various needs.**

Keywords- content management; semantic technologies; heterogeneous data repository

I. INTRODUCTION

The field of Content Management (CM) refers to the process of gaining control over the creation and distribution of information and functionality. Concisely, an effective Content Management System (CMS) aims at getting the right information to the right people in the right way. Usually, CM is divided into three main phases namely collecting, managing, and publishing of content. The collection phase encompasses the creating or acquiring information from an existing source. This is then aggregated into a CMS by editing it, segmenting it into components, and adding appropriate metadata. The managing phase includes creating a repository that consists of database containing content components and administrative data (data on the system's users, for example). Finally, in the publishing stage the content is made available for the target audience by extracting components out of the repository and releasing the content for use in the most appropriate way [1] [2].

Currently, there are several commercial and open-source technologies available that are applied to address different content management needs across various industries including healthcare [3], and education [4], for example. However, the standard versions of the existing solutions are not always capable of supporting end-users in their specified context to reach their particular goals in an effective, efficient and satisfactory way [5]. For instance, the included content retrieval mechanisms are often implemented using traditional keyword based search engines that are not adapted to serve any task specific needs [6][7][8].

One of the main issues to be resolved is how to convert existing and new content that can be understood by humans into semantically-enriched content that can be understood by machines [9]. The human-readable and unstructured content is usually difficult to automatically process, relate and categorize, which hinders the ability to extract value from it [10]. Additionally, it results in the restriction of development of more intelligent search mechanisms [9]. To address some of the above described deficiencies, semantic technologies are being increasingly used in CM. In particular, the utilization of domain specific vocabularies and taxonomies in content analysis enables accurate extraction of meaningful information, and supports task-specific browsing and retrieval requirements compared to traditional approaches [9]. Furthermore, semantic technologies facilitate creating machine-readable content metadata descriptions, which allows, for example, software agents to automatically accomplish complex tasks using that data. Moreover, semantically enhanced metadata helps search engines to better understand what they are indexing and providing more accurate results to the users [11].

In this paper, we introduce the HARMONISE platform, developed in the FP7 EU HARMONISE [12] project. This paper is an extended version of work published in [1], where a semantic layer implemented on top of the HARMONISE platform was introduced. We extend our previous work by providing more details about the technical realizations of the platform and the semantic layer. Moreover, an evaluation process carried out for the platform and the semantic layer is depicted in this paper.

The HARMONISE platform is a domain specific CMS that provides information and tools for security-driven urban resilience in large-scale infrastructure offering a holistic view to urban resilience. A database contained by the system manages an extensive set of heterogeneous material that comes in different forms including tools, design guidance and specifications. The platform aims at serving as a 'one-stop-shop' for resilience information and guidance and it contains a wealth of information and tools specifically designed to aid built environment professionals. While the platform and the hosted toolkit are aimed to be used by a variety of potential end-users from planners and urban designers to construction teams, building security personnel and service managers, the specialized problem domain and heterogeneous content repository poses significant challenges for users to effectively retrieve information to accomplish their tasks and goals.

As earlier discussed, the HARMONISE platform is extended with a novel Semantic Layer for the HARMONISE (SLH) approach. The SLH is a semantic content management solution developed to address many of the above discussed challenges related to domain specific content management. It is implemented on top of the HARMONISE platform and it aims at offering more task specific and personalized content management support for end-users. Additionally, by utilizing domain specific annotation and categorization of content the SLH facilitates the management of heterogeneous and large content repository hosted by the HARMONISE platform.

The semantic information modelling allows better understanding of platform content, linkages between unstructured information and tools, and more sophisticated answers to users' various needs. Moreover, the semantic knowledge representations created by the layer help end-users to combine different data fragments and produce new implicit knowledge from existing data sets. Finally, by utilizing Linked Data [13] technologies the SLH fosters interoperability and improves shared understanding of key information elements. The utilization of interconnected and multidisciplinary knowledge bases of the Linked Data cloud also enables applying the solution in other problem areas such as health care or education.

The rest of paper is organized as follows. Section II provides a through description of the HARMONISE platform and its application area. In Section III the architecture and different components of the SLH are described. Section IV provides a Use Case example demonstrating the functionality of the SLH. Finally, Section V concludes the paper.

II. THE HARMONISE CONTENT MANAGEMENT PLATFORM

At present, there exist a number of content management systems that enable publishing, managing and organizing electronic documents. For example, Drupal [1] [14] and WordPress [2] [15] are well-known, general-purpose CM solutions providing such basic CM features such as user profile management, database administration, metadata management, and content search and navigation functionalities [5]. These tools provide functionality to create and edit a website's content often with easy-to-use templates for digital media content publishing.

The HARMONISE platform is a web platform that provides information and tools specifically designed to aid urban decision makers in enhancing the resilience of large scale urban built infrastructure. The platform includes an innovative search process designed to promote holistic decision making at each stage of the resilience cycle, an automatic content recommendation mechanism to suggest most relevant contents for a user, educational elements that provide content and self-assessment tools that help end-users to assess the general resilience and security level of an existing or proposed large scale infrastructure. Moreover, the semantic layer developed within the platform enables better understanding of data, linkages between unstructured information and tools, and more sophisticated answers to users' various needs.

The platform is mainly composed by two macro-components, the HARMONISE web site which represents the front-end of the platform to the user and the semantic layer that includes services for enhancing the overall functionality allowing more personalized user experience for stakeholders who utilize the platform their daily work. In Figure 1 the main elements of the HARMONISE platform are shown.

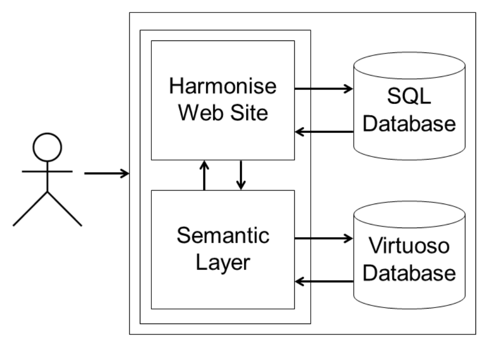

Figure 1. The key elements of the HARMONISE platform

As stated above, the HARMONISE platform is a CMS specifically tailored for the domain of urban resilience. The system provides information and tools for security-driven urban resilience in large-scale infrastructure and contains a variety of interactive elements allowing users to both import and export data to and from the platform and personalize the platform to their own needs. The core functionalities of the HARMONISE platform are implemented using ASP.NET

[1] www.drupal.org
[2] https://wordpress.org/

web application framework and it utilizes Microsoft SQL 2012 database to store content items.

The main features and functionalities are divided between three user profile categories defined for the HARMONISE platform. To begin with, a standard registered user can navigate through the different sections of the platform excluding the upload section, use the search functionalities and view all the platform contents. However, he cannot upload or edit any content. In contrast, an uploader who has been granted permission by the administrator can access the upload section and upload contents in addition to the functionalities available for the standard user. Moreover, the uploader can edit/delete the content he has uploaded. In order to become an uploader user has to insert a special password (i.e., uploader password) provided by the platform administrator. Finally, the administrator of the platform can edit/delete the content uploaded by an uploader. Moreover, he can manage the user assignment to a specific group and generate the uploader passwords. The main functionalities enabled for each user category are depicted in Figure 2.

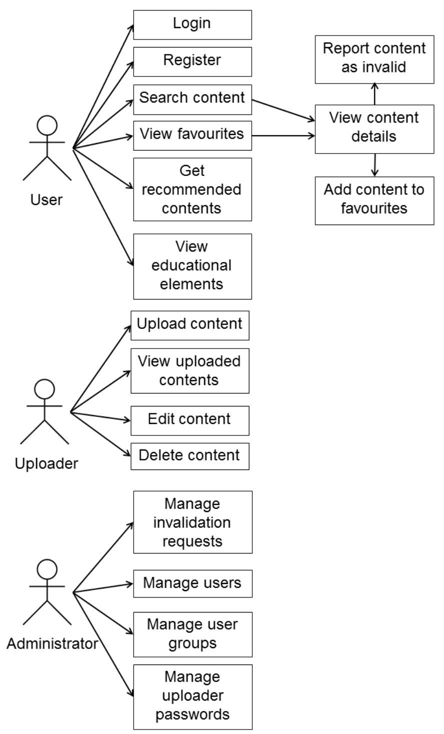

Figure 2. The user profiles and main features of the platform

An important part of the HARMONISE content management platform is the Thematic Framework [16] that was created to structure information within the platform and to guide end-users through an innovative step-by-step search process. The Thematic Framework is set out in Figure 3.

By unpacking resilience into a number of key layers the Thematic Framework provides the necessary taxonomy needed for realizing effective domain-specific content annotation and categorizing functionalities, as later discussed. The objective of the domain-specific annotation is to allow users to easily identify and access information and tools within the platform, and to search the platform according to their unique needs or interests.

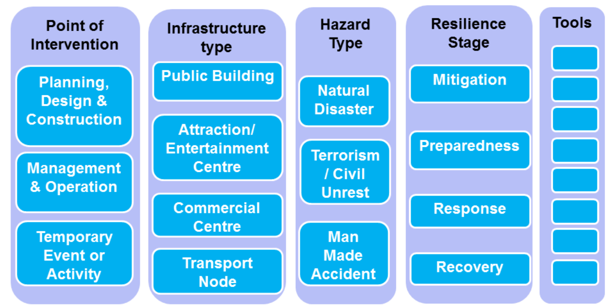

Figure 3. The Thematic Framework (adopted from [16])

III. The Semantic Layer

The HARMONISE content management platform hosts a large portfolio of urban resilience related content. However, finding relevant information and tools from such a knowledge base with conventional information retrieval methods is usually both tedious and time consuming, and tends to become a challenge as the amount of content increases [9]. Often users have difficulties in grouping together related material or finding the content that best serve their information needs, especially when content is stored in multiple formats [17].

In general, the existing CMSs usually lack consistent and scalable content annotation mechanisms that allow them to deal with the highly heterogeneous domains that information architectures for the modern knowledge society demand [18]. The semantic layer described in this study aims at addressing the above mentioned challenges by integrating semantic data modelling and processing mechanisms to the core HARMONISE platform functionalities. For example, the application of semantic mark-up based tagging of web content enables expressively describing entities found in the content, and relations between them [9]. Moreover, by utilizing the Linked Data Cloud links can be set between different and heterogeneous content elements and therefore connect these elements into a single global data space, which further facilitates interoperability and machine-readable understanding of content [19].

The main features of the SLH are divided to four parts. First, the metadata enrichment part produces information-rich metadata descriptions of the content by enhancing content with relevant semantic metadata. Second, the semantic metadata repository implements the necessary means for storing and accessing the created metadata. The third component of the SLH realizes a semantic search feature. In more detail the search service aims at returning more meaningful search results to the user by utilizing both keyword-based semantic search and “Search by theme” filtering algorithm that restricts the searchable space by enabling users to select certain categories from the Thematic Framework. The final part, content recommendation, combines information about users’ preferences and profile to find a target user neighborhood, and proactively recommends new urban resilience tools/resources that might be of potential interest to him/her. In Figure 4 the logical architecture of the SLH is represented.

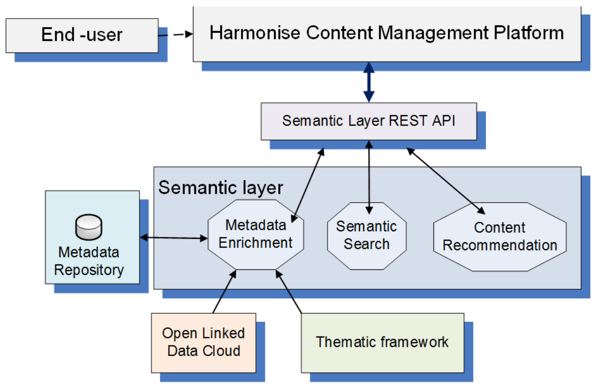

Figure 4. The logical architecture of the SLH

The following sections describe the logical architecture in more detail.

A. *Semantic Layer REST API*

The Semantic Layer REST API provides the necessary interface for the HARMONISE Platform to interact with the SLH. It enables, for example, to transmit query requests from the platform to the SLH or retrieve content recommendations personalized for a particular user.

B. *Metadata Enrichment*

The purpose of the Metadata Enrichment service is to produce information-rich metadata descriptions of the content that is uploaded to the HARMONISE platform. Enhancing content with relevant semantic metadata can be very useful for handling large content databases [2]. A key issue in this context is improving the "findability" of content elements (e.g., documents, tools).

The enrichment process is based on tagging. A tag associates semantics to a content item, usually helping the user searching or browsing through content. These tags can be used in order to identify the most important topics, entities, events and other information relevant to that content item. The tagging data is created by analyzing the uploaded content and the metadata manually entered by the user. This information consist e.g., title, keywords, Thematic Framework categories, topics, content types and phrases of natural language text.

In the metadata analysis the following three technologies that provide tagging services are utilized: ONKI [3] [20], DBPedia [4] [21] and OpenCalais [5] [22]. The ONKI and DBPedia knowledge bases provide enrichment of the human defined keywords by utilizing Linked Data reference vocabularies and datasets. The Metadata Enrichment service utilizes the APIs of the above mentioned technologies to search terms that are somehow associated to the entities defined by a user. The relationships between the enriched terms and the original entity are illustrated in Figure 5, in which examples of enriched concepts for the term 'Building' are represented.

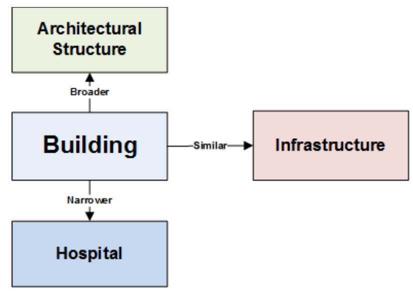

Figure 5. The enrichment of the human defined keywords

As shown in Figure 5, the enriched terms fall into three categories: similar, broader and narrower. The similar terms are synonyms to the original entities whereas broader terms can be considered as more general concepts. The narrower terms represent examples of more specific concepts compared to the original entity. Each of the acquired terms contains a Linked Data URI that can be accessed to get more extensive description of that term. By enriching the human defined keywords with additional concepts and Linked Data URIs more comprehensive and machine-readable information about uploaded content items can be generated.

The uploaded content items are also examined using the OpenCalais text analyzer tool. Using such mechanisms as natural language processing and machine learning the tool allows analyzing different text fragments contained by the uploaded content item. As a result, OpenCalais discovers entities (Company, Person etc.), events or facts that are related to the uploaded content element.

In the final part of the metadata enrichment process the metadata elements created by different tools are merged as a single RDF (Resource Description Framework) metadata description and stored to the metadata database.

C. *Semantic Metadata Repository*

The database technology used for storing the semantic metadata of content is OpenLink Virtuoso [23]. Virtuoso is a relational database solution that is optimized to store RDF data. It provides good performance and extensive query interfaces [24] and was thus selected as the metadata storage to be used in the SLH.

D. *Semantic Search*

The Semantic Search service aims at producing relevant search results for the user by effectively utilizing the machine-readable RDF metadata descriptions created by the Metadata Enrichment service. Unlike traditional search engines that return a large set of results that may or may not be relevant to the context of the search, the Semantic Search analyses the results and orders them based on their relevancy. Thus, users are emancipated from performing the time-consuming work of browsing through the retrieved results in order to find the content they are looking for.

[3] http://onki.fi/
[4] http://dbpedia.org/
[5] http://www.opencalais.com/

The Semantic Search service is implemented as a Java web application composed of three main components (see Figure 6):

- RESTful Web Service: based on Apache CXF framework, it represents the semantic search service front-end. It receives the search queries from the HARMONISE platform and returns the list of search results provided by the underlying components;
- Semantic Search Service Core (SSS Core): component based on Java/Maven project, customized to manage all the core processes (data indexing, content search, content retrieving, results formatting);
- Semantic Search Engine: component based on Apache Solr [25] enterprise search platform, in charge of the indexing and the search processes. When a new content is uploaded to the HARMONISE platform it reads from the Virtuoso database the data produced by the semantic content enrichment service in order to create the index to query on. When a user submits a query the semantic search engine queries the index in order to find the documents that best match the user request parameters.

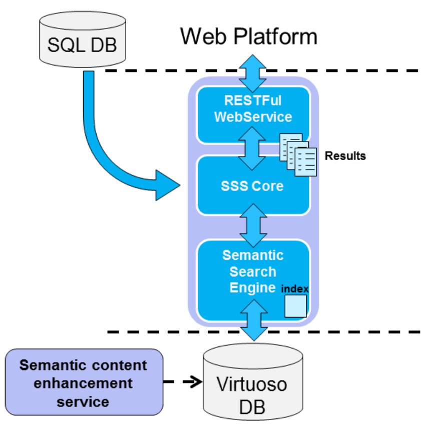

Figure 6. Logical architecture of the semantic search service

The Semantic Search service relies on the Solr search engine [26] in order to search across large amount of content metadata and pull back the most relevant results in the fastest way. The Solr component is a web application developed in Java and provided by the Solr open source enterprise search platform from the Apache Lucene project [25]. Solr is a document storage and retrieval engine and every piece of data submitted to it for processing is a document composed by one or more fields. Internally Solr uses Lucene's inverted index to implement its fast searching capabilities. Unlike a traditional database representation where multiple documents would contain a document ID mapped to some content fields containing all of the words in that document, an inverted index inverts this model and maps each word to all of the documents in which it appears. Solr stores information in its inverted index and queries that index to find matching documents.

According to the data structure of the contents uploaded to the Virtuoso database by the Metadata Enrichment service, the document fields shown in Table I have been defined for the construction of the Solr index.

TABLE I. SOLR INDEX FIELDS

Field	Description
Id	Content identifier on Virtuoso DB
Upload date	Date when the content has been uploaded
Topics	List of topics from the Thematic Framework
Resilience Tasks	List of Resilience Cycle tasks
Permissions	List of user groups allowed to view the document
Title	Title of the content
Description	Textual description of the content
Keywords	List of keywords (inserted by the uploader)
Tags	List of tags added by the metadata enhancement service.

The search results provided by the Semantic Search service are ranked according to the relevancy scores that measure the similarity between the user query and all of the documents in the index. The results with highest relevancy scores appear first in the search results list.

The scoring model is composed by the following scoring factors:

- Term Frequency: is a measure of how often a particular term appears in a matching document. Given a search query, the greater the term frequency value, the higher the document score.
- Inverse Document Frequency: is a measure of how "rare" a search term is. The rarer a term is across all documents in the index, the higher its contribution to the score.
- Coordination Factor: It is the frequency of the occurrence of query terms that match a document; the greater the occurrence, the higher is the score.
- Field length: the shorter the matching field, the greater the document score. This factor penalizes documents with longer field values.
- Boosting: is the mechanism that allows to assign different weights to those fields that are considered more (or less) important than others.

E. Content Recommendation

Similar to the Semantic Search, the Content Recommendation Service (CRS) is based on semantic modelling of content resources. The aim of the content recommendation service is to improve user experience in terms of the search functionality and the filtering of relevant information through the utilization of collaborative filtering. As the volume of content continues to increase, the development of recommendation systems (RS) have become essential to handle large volumes of data. They are widely used across diverse domains to predict, filter and extract content for users [27][28]. Examples of commercial applications of RS include Amazon, Twitter, Facebook and Ebay. A popular type of RS is collaborative filtering. This type of RS analyses information on users' preferences and predicts content to present to users based on their similarity to other users of the system [29]. Due to the information stored for users in their user profile, collaborative filtering was a natural selection to predict content to users based on similarities in their profiles. The developed CRS utilizes user profiles which are created and maintained by the HARMONISE platform

An overview of the CRS algorithm is provided in Figure 7. Figure 7 illustrates how user preference and user profile similarity are fused together along with a weighted sum to provide a ranked list of recommendation tailored to the user.

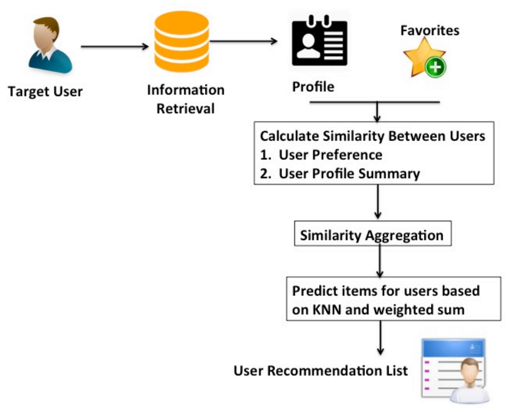

Figure 7. Overview of the CRS Algorithm

Figure 7 details the various stages of the recommendation process which includes the extraction of user details from the HARMONISE system through to the list of recommendations presented to the user at the end. A detailed overview of the CRS algorithm can be found in [30]. The user profiles contain information about user's preferences and favorite content. It also includes the content item IDs that have been already recommended for that particular user. This information is then utilized when content recommendations are created for different users. The CRS is triggered by the HARMONISE platform through the 'get recommendation' method provided by the Semantic Layer REST API. The ID of the user is transmitted as a method parameter. Once the recommendation service receives the ID, it retrieves the user profile of the user from the database and analyses the information it contains. Various pieces of information are stored in the profiles such as topics of interest, group membership, languages, lines of investigation. Furthermore, content that the user has marked as favorite is stored as user preferences. The CRS then combines all the information about users' preferences and profile information to find a target user neighborhood, and recommend new urban resilience tools/resources that might be of potential interest to him/her. To do this, firstly the ordered weighted average and uniform aggregation operators are applied to fuse user information and obtain global degrees of similarity between them using the formula below:

$sim_p^{i,j} = OWA_w(sim_I^{i,j}, sim_G^{i,j}, sim_L^{i,j})$ profile between users i and j , and sim_I, sim_G, sim_L are the profiles line of investigation, interests and groups respectively which are aggregated using ordered weighted sum.

The similarity between the preferences of users u_i and u_j as $sim_f^{i,j} \in [0,1]$ is also calculated using the Jaccard index $J_F^{i,j} = J(F_i, F_j)$ among the sets $F_i, F_j \subset I$ where I are the items marked favorite by u_i and u_j .This is schematically presented in Figure 8 from [30] below.

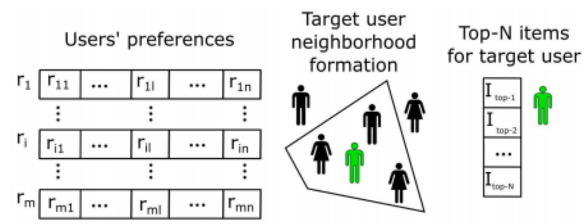

Figure 8. Schematic overview of User Preferences approach from [30]

When the profile similarity and preference similarity between users have been calculated, these are then fused together resulting in a global degree of similarity between the target user u_i and the rest of users in the system. We apply a uniform aggregation function to obtain the global similarity $sim^{i,j} \in [0,1]$ between u_i and u_j.

The actual recommendation generation process is carried out by comparing the user profile data with the semantic content metadata descriptions. Similarly as in the search algorithm described in the previous section, the content items whose metadata is associated with e.g., terms, topics or research areas as contained by the user profile are included to the initial recommendation results. Of course, the content items that have already been recommended for the user are excluded from the results list. Subsequently, the recommendation results are analyzed using the ranking model introduced by the Semantic Search. Using K-nearest neighbors, the content items that gets the highest score is returned to the platform as the most highly recommended content item. This involves the generation of a recommendation list $R_i = \{C_i, \dots, C_n\}$ consisting of content C of size $h \in \mathbb{N}, h \ll n$ is presented to users on the HARMONISE interface as illustrated in Figure 9. This list contains items $C_k = i_l \in I$ with the highest values for $p(u_i, i_l)$. This results in the user receiving a list of content ordered by rating value.

IV. USE CASE EXAMPLE

The functionality of the SLH is demonstrated with a Use Case example in which a user uploads a document into the HARMONISE content management platform and tries to retrieve it with the search functionality. Additionally, the recommendation service is verified by creating a user profile that is interested in topics relevant to the uploaded content. The content item used in the Use Case example is an electronic manual that presents tools to help assess the performance of buildings and infrastructure against terrorist threats and to rank recommended protective measures. This kind of guidance document is a typical representative of a content item managed by the HARMONISE platform.

Once the user has provided necessary input in the upload form the content description is transmitted to the Metadata Enrichment component that processes the collected data and forms an RDF metadata description of the content. It was noted that the returned semantic content metadata contained five keywords that are enriched with 81 broader or narrower and 26 similar terms. Moreover, the content is annotated with several categories defined by the Thematic Framework.

Once the enriched metadata is stored to the Semantic Metadata Repository, and indexed by the Semantic Search service, it can be tried to be retrieved with the search functionality. The content retrieval is tested with the 'Resilience Search Wizard' feature provided by the SLH. The wizard allows to define keywords and to select those categories from the Thematic Framework that are considered as relevant to the uploaded content. The utilized search parameters are shown in the search wizard screenshot illustrated in Figure 9.

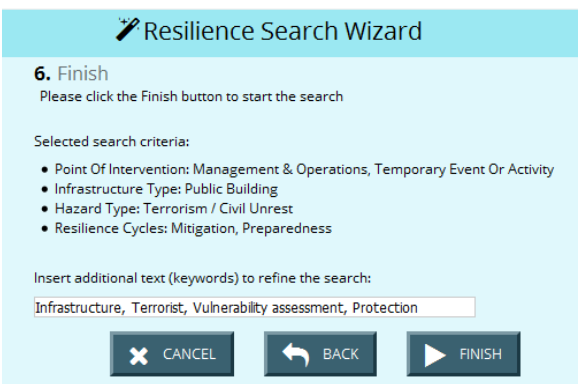

Figure 9. Search parameter definition

As earlier explained, the search functionality is able to sort the results based on their relevancy. Figure 10 represents the most highly ranked search results returned by the search service. As can be seen, the applied ranking algorithm identified the uploaded electronic manual document as the second relevant search result for the given search query. In total, the search functionality found 24 results with the defined search parameters.

In the final phase of the use case example, the Content Recommendation service is tested by creating a user profile and obtaining personalized recommendations. The user profile was created with 6 topics of interests of interest from a total of 13 topics namely: Point of Intervention, Management and Operation, Infrastructure type, Commercial Center, Hazard Type and Man Made Hazard. The user then marked 10 items of favorite content from a total of 156 items in the database.

Figure 10. The ranking of search results

These included content such as "Tools of Regional Governance" and "Flood management in Linares Town". For the first step in the recommendation algorithm, Jaccard index is utilized to compute the degree of similarity between the favourite content and profile information of the user entered and all the users of the HARMONISE system. In the second step, a KNN algorithm is applied to identify the 5 most similar neighbors. Based on neighbor users, we compute for each item not marked as a favorite by the user, a predicted rating. This is used to construct an ordered recommendation list to the target user, which in this case study was a list of 5 recommendations including documents based on "Key issues of Urban Resilience", "Building urban resilience Details" and "Resilience: how to build resilience in your people and your organization".

V. A REVIEW OF EQUIVALENT TOOLS AND APPROACHES

As earlier discussed, at present there exist no similar content management tools that would address the special requirements set by the domain of urban resilience. However, over the past few years approaches that provide equivalent functionalities as the HARMONISE platform and the SLH have been delivered by research community. Although these tools are designed for different application areas, they have many similar end-user requirements and technological characteristics. In this chapter a selection of these tools are being reviewed and analyzed.

To start with, [31] introduces a platform for curation technologies that is intended to enable human experts to get a grasp and understand the contents of large and heterogeneous document collections in an efficient way so that they can curate, process and further analyze the collection according to their sector-specific needs. Furthermore, the platform aims at automating such tasks as looking for information related to and relevant for the domain, learning the key concepts, selecting the most relevant parts and preparing the information to be used. As in the HARMONISE project, the platform is extended with a semantic web-layer that provides linguistic analysis and discourse information on top of digital content.

The target audience of the platform for curation technologies is knowledge workers who conduct research in specific domains with the goal of, for example, preparing museum exhibitions or writing news articles. Currently, the focus in the platform is on written documents but in future the aim is to improve the platform by improving its abilities to convert non-textual data into text. Similarly as in HARMONISE platform, the semantic layer of the platform for curation technologies facilitates semantic annotation, enables connecting interlinked representation to external information sources, implements a semantic triple store and provides search functionalities.

As discussed above, the platform for curation technologies provides similar functionalities compared to the HARMONISE platform and the SLH. Moreover, the platform uses many of the same technologies. However, the platform for curation technologies is not as well optimized to address the needs of a special domain. For example, it does not provide an application area specific taxonomy that is often needed for realizing effective domain-specific content annotation and categorizing functionalities. Moreover, the HARMONISE platform and the semantic layer provide more comprehensive support for the management of heterogeneous content. However, the abilities of the platform for curation technologies to support natural language and multilingual text processing are more advanced compared to the HARMONISE platform and the SLH.

A second approach providing similar content management functionalities as the HARMONISE platform and the SLH is the Ondigita platform [32] that is developed for the management and delivery of digital documents to students enrolled in bachelor's courses within the field of engineering. The platform implements a cloud-based repository to allow educational organizations to create a digital collection of their educational materials and enable students to store and access these resources in their computers, tablets, or mobile phones. Moreover, the Ondigita platform supports the managing of heterogeneous learning material including books, audio and video, for example, and enables students to manually annotate content elements by highlighting important text passages and adding textual notes. The resulting annotations can be shared with others students.

The main components of the Ondigita platform include a course materials repository, an application server and a web application to access the content repository. The platform also offers adapters that enable integrating the infrastructure with such external file hosting services as Dropbox or Google Docs. Additionally, a mobile application available for Android OS is provided. When comparing to the HARMONISE platform and the SLH it can be concluded that the search services provided by the Ondigita platform are more constricted. Moreover, the Ondigita platform does utilize semantic technologies or include any domain specific taxonomies or ontologies. Also, the manual annotation of content items can be considered as relatively time-consuming and labour-intensive. However, the Ondigita platform offers better support for interoperability with widely used file hosting services and allows utilizing its services through a mobile application. Both of the aforementioned features are currently missing from the HARMONISE platform.

The final approach to be reviewed here is Sentic Album [33]. Sentic Album is a content-, concept-, and context-based online personal photo management system that exploits both data and metadata of online personal pictures to annotate, organize, and retrieve them. Sentic Album utilizes a multi-tier architecture that exploits semantic web techniques to process image data and metadata at content, concept, and context level, in order to grasp the salient features of online personal photos, and hence find intelligent ways of annotating, organizing, and retrieving them.

Similar to the HARMONISE platform and the SLH, Sentic Album includes semantic databases, automatic annotation features and a search and retrieval module. The search functionality provides users an UI that allows them to manage, search and retrieve their personal pictures online. Moreover, users are able to assign multiple categories to an image object, enabling classifications to be ordered in multiple ways. This makes it possible to perform searches combining a textual approach with a navigational one. The combination of key-word based search and content classification based search is similar to the search feature provided by the SLH. However, in HARMONISE platform the classifications are based on pre-defined domain knowledge which enables addressing more task specific needs and requirements. In general, the main difference between Sentic Album and the HARMONISE platform is their targeted group of end-users. The HARMONISE platform aims at serving a specific group of end-users including planners, urban designers and building security personnel, whereas Sentic Album is targeted to more heterogeneous group of end-users.

VI. System evaluation

In order to test and demonstrate the viability and effectiveness of the HARMONISE Platform and the SLH, the developed system was applied in five different case study contexts under the project activities. The selected case study cities are Dublin, Ireland; London, United Kingdom; Genoa, Italy; Bilbao, Spain and Vantaa, Finland. Each of these case studies incorporates a large scale urban built infrastructure project, at different scales and contexts. Moreover, the case studies incorporate a combination of urban built infrastructure systems at various stages from completed, operational projects, to as yet unrealized proposals at design and planning stage.

The actual evaluation process was started by testing the platform and the SLH with a range of built environment professionals (including architects, urban designers and town planners) from the HARMONISE project consortium organisations. Subsequently, the developed system was demonstrated and assessed in case study specific workshops where the platform and the SLH were presented to the key stakeholders including various urban resilience professionals and policy makers. Moreover, the workshop participants were invited to experiment and criticize the system.

The evaluation process also included creating a set of standardised questionnaires. The questionnaires were to be used when interacting with the end-users in the case studies. The purpose of the questionnaires was to elicit feedback from users and associated stakeholders as to the performance (actual or intended) of the HARMONISE platform and the SLH being tested. Overall, this task aimed to discover '*what* *works'* on the ground in practice and learn about the user focussed process of implementation (what users want).

The survey was designed to focus on both the technical aspects of the platform as well as on the usability of the system. Importantly, the adopted approach incorporated a balance between quantitative and qualitative feedback. The results from this survey were to be used as feedback to designers so that improvements can be made to the platform.

The actual evaluation was divided into two phases: In phase 1 (between March and April 2015) a workshop was held in each case study area, led by the Case Study Leads (CSL's). The purpose of these workshops was to present the HARMONISE concept and the platform to key stakeholders and to gather feedback on the work-in-progress version of the platform. Furthermore, the phase 2 (between mid-October and early December 2015) a second workshop was held in each case study area. During these second workshops, an improved version of the platform was presented to stakeholders.

For the first testing phase, the created questionnaires were circulated to each CSL's in advance of the case study workshops as an online survey link. The survey was designed to be self-completed by the end user so each CSL then requested that all case study workshop participants complete this online questionnaire in their own time following the event. The survey was opened for responses for a period of one month following the workshop held in the case study location.

However, in some cases more emphasis was placed on gathering feedback within the context of the stakeholder workshops (rather than using the online questionnaire after the workshop). As such, feedback was gathered in two ways:

- During the stakeholder workshops only – In this case, some questions from the online questionnaire were used to stimulate discussion among the stakeholders. The discussion was then recorded by the HARMONISE facilitators in a manner which closely matched the format of the questionnaire (to ensure that feedback could be analyzed in a relatively consistent manner).
- During the stakeholder workshops and using the online questionnaire – In this case, stakeholders reported their 'first thoughts' during the workshops, with some also choosing to report more detailed feedback after the event through the use of the online questionnaire

Following the completion of phase 1 of the testing process, many CSL's reported that they had received the richest feedback during discussions as part of the workshops, with less feedback provided through the online survey. Indeed, many CSL's reported that some stakeholders felt that the online survey was too lengthy, a factor which discouraged them from inputting feedback in a detailed manner. As a result, the testing approach for the second phase of testing was adjusted in two minor ways – 1. The questionnaire survey was further edited (with stakeholders encouraged to focus on providing qualitative comments) 2. A soft copy version of the survey was circulated to each workshop participant during the various workshops – with a request for them to complete the survey during the workshop.

As discussed above, the evaluation was divided into two phases. In both parts questionnaires and end-user discussions were used to collect information about the abilities of the system to support urban resilience professionals and policy makers. The questionnaire included questions, for example, about perceived ease-of-use and perceived usefulness of the tool. Perceived usefulness that is defined as "the degree to which a person believes that using a particular system would enhance their job performance" and perceived ease-of-use that is defined as "the degree to which a person believes that using a particular system would be free from effort" were considered as important factors as they can determine whether people will accept or reject an emerging information technology [34].

The questionnaire used a five-level grading system where "Strongly agree" was the best and "Strongly disagree" the worst grade, with "Neither agree nor disagree" being average. Figure 11 summarizes the feedback on the platform, as gathered during evaluation phase 1.

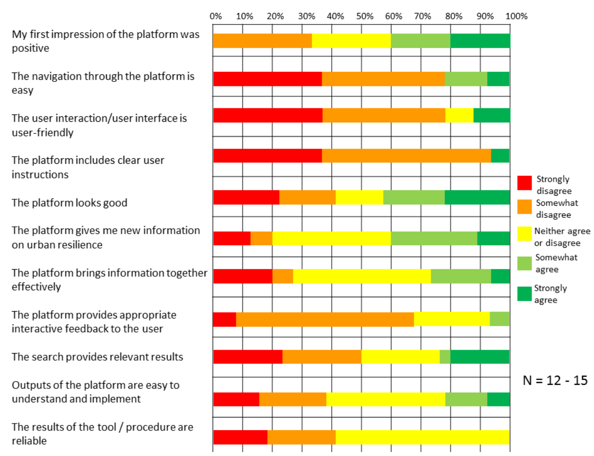

Figure 11. Questionnaire results from the first evaluation stage.

In general, the first evaluation was useful and revealed some extremely interesting points about the constructed system and enhanced the discovery of which features users find useful and easy to use, and which parts of the application still need to be improved. As depicted by the questionnaire results presented above, the usability and graphical user interface of the system was found to require improvements. Moreover, it was perceived as difficult to navigate through the platform. The search feature was also identified as deficient. In general, the first evaluation stage indicated that the platform and the SLH still require further development to reach its full potential.

The knowledge and experience gained from the first evaluation stage was utilized in the subsequent development of the HARMONISE platform and the SLH. The major improvements included, for example, re-designing the visual appearance, user interface and navigation of the platform. Moreover, the metadata management services of the SLH were completely redeveloped and optimized to better support the requirements set by the search feature. Finally, the utilized search algorithm was improved and integrated with the Thematic Framework based content classification mechanism.

The new versions of the HARMONISE platform and the SLH were tested in the second evaluation phase. Again, the actual testing was carried out in workshop sessions were end-users were allowed to test the system and fill in a questionnaire. Additionally, verbal feedback was collected from the workshop participants. Figure 12 summarizes the results from the second evaluation round.

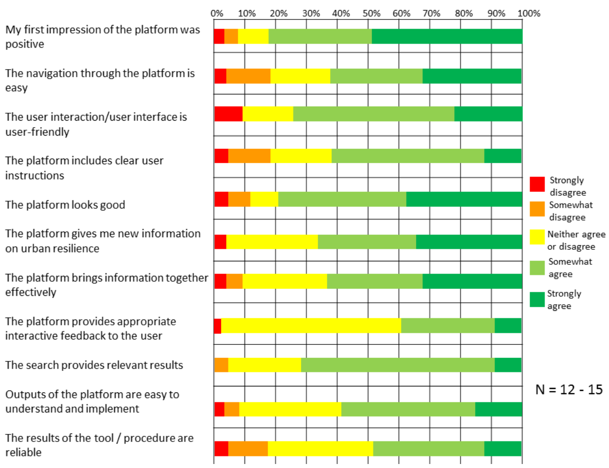

Figure 12. Questionnaire results from the second evaluation stage.

It should be noted that the chart for testing period 2 is not directly comparable with the results of testing period 1 as most of the test participants were already somewhat familiar with the system in test period 2. Nevertheless, the chart for test period 2 provide a useful indication of progress in meeting stakeholder expectations of the platform, and in addressing some of the stakeholder concerns raised during testing stage 1. In more detail, the stakeholder feedback from testing stage 2 illustrates far greater user satisfaction with the various elements of the platform than was recorded during testing stage 1 – shown by the far higher percentage of stakeholders who 'somewhat agreed' or 'strongly agreed' with some of the key stated aims for the platform functionality.

VII. CONCLUSION

In this work, we introduce a content management platform for the domain of urban resilience. The platform aims at serving as a 'one-stop-shop' for resilience information and guidance offering a holistic view to urban resilience. Furthermore, the platform contains information and tools specifically designed to aid decision makers in enhancing the resilience of large scale urban built infrastructure.

The developed content management platform is extended with an additional information processing layer that utilizes semantic technologies to manage an extensive set of heterogeneous material that comes in different forms including tools, design guidance documentation and specifications. Moreover, the developed semantic layer enables the creation of machine-understandable and machine-process able descriptions of content items. This has resulted in an improved shared understanding of information elements and interoperability.

With the effective utilization of Linked Data based analysis tools and domain specific content annotation mechanisms, the semantic layer offers task specific and personalized content management support for end-users. The enhanced intelligence has provided better understanding of urban resilience content, linkages between unstructured information and tools, and more sophisticated answers to users' various needs. Furthermore, the recommendation service provides the functionality to predict relevant content to the user of the system using our collaborative filtering approach. This approach is able to avail of the rich user data available in terms of profile information and also content preference information resulting in a set of recommendations tailored to individual users.

The developed HARMONISE platform and the SLH have been tested by HARMONISE project partners and other invited domain specialists. Additionally, several case study stakeholders have evaluated the system in terms of usability, perceived usefulness and the relevancy of received search and recommendation results. The performed evaluations have provided valuable information about the deficiencies and strengths of the HARMONISE platform and SLH.

The future work includes further refining the HARMONISE platform and the SLH on the basis of the feedback received from the evaluation process. Additionally, the graphical appearance of the platform's user interface as well as the usability of individual components will be improved.

REFERENCES

[1] I. Niskanen, M Murtonen, F Browne, P. Davis and F. Pantisano, "A Semantic Layer for Urban Resilience Content Management," 8th International Conference on Information, Process, and Knowledge Management (eKNOW 2016), Venice, Italy, April 24-28, 2016.

[2] B. Boiko, Content management bible, John Wiley & Sons, 2005.

[3] S. Das, L. Girard, T. Green, L. Weitzman, A. Lewis-Bowen, and T. Clark, "Building biomedical web communities using a semantically aware content management system," Briefings in bioinformatics, 10(2), pp. 129-138, 2009.

[4] N. W. Y Shao, S. J. H Yang, and A. Sue, "A content management system for adaptive learning environment," Multimedia Software Engineering, Proceedings. Fifth International Symposium on, IEEE, 2003.

[5] N. Mehta, Choosing an Open Source CMS: Beginner's Guide, Packt Publishing Ltd, 2009.

[6] D. Dicheva and D. Christo, "Leveraging Domain Specificity to Improve Findability in OER Repositories," Research and Advanced Technology for Digital Libraries. Springer Berlin Heidelberg, pp. 466-469, 2013.

[7] S. K. Patel, V. R. Rathod, and S. Parikh, "Joomla, Drupal and WordPress-a statistical comparison of open source CMS," Trendz in Information Sciences and Computing (TISC), 3rd International Conference on. IEEE, 2011.

[8] C. Dorai and S. Venkatesh, "Bridging the semantic gap in content management systems," Media Computing. Springer US, pp. 1-9, 2002.

[9] J. L. Navarro-Galindo and J. Samos, "The FLERSA tool: adding semantics to a web content management system," International Journal of Web Information Systems 8.1: pp. 73-126, 2012.

[10] A. Kohn, F. Bry, and A. Manta, "Semantic search on unstructured data: explicit knowledge through data recycling," Semantic-Enabled Advancements on the Web: Applications Across Industries, 194, 2012.

[11] D. R. Karger and D. Quan, "What would it mean to blog on the semantic web?," The Semantic Web–ISWC 2004. Springer Berlin Heidelberg, pp. 214-228, 2004.

[12] The HARMONISE project (Available online at: http://harmonise.eu/) [accessed: 13.4.2016]

[13] Linked Data (Available online at: http://linkeddata.org/) [accessed: 13.4.2016]

[14] Drupal (Available online at: https://www.drupal.org/) [accessed: 13.4.2016]

[15] WordPress (Available online at: https://wordpress.org/) [accessed: 13.4.2016]

[16] S. Purcell, W. Hynes, J. Coaffee, M. Murtonen, D. Davis, and F. Fiedrich, "The drive for holistic urban resilience," 9th Future Security, Security Research Conference, Berlin Sep. 16- 18, 2014.

[17] A Vailaya, M. A. Figueiredo, A. K. Jain, and H. J. Zhang, "Image classification for content-based indexing," Image Processing, IEEE Transactions on, 10(1), pp. 117-130, 2001.

[18] R. García, J. M. Gimeno, F. Perdrix, R. Gil, and M. Oliva, "The rhizomer semantic content management system," In Emerging Technologies and Information Systems for the Knowledge Society pp. 385-394, Springer Berlin Heidelberg, 2008.

[19] M. Hausenblas, "Exploiting linked data to build web applications," IEEE Internet Computing 4, pp. 68-73, 2009.

[20] ONKI - Finnish Ontology Library Service (Available online at: http://onki.fi/) [accessed: 13.4.2016]

[21] DBpedia (Available online at: http://dbpedia.org/) [accessed: 13.4.2016]

[22] OpenCalais (Available online at: http://www.opencalais.com/) [accessed: 13.4.2016]

[23] Virtuoso Universal Server (Available online at: http://semanticweb.org/wiki/Virtuoso) [accessed: 13.4.2016]

[24] O. Erling and I. Mikhailov. "RDF Support in the Virtuoso DBMS," Networked Knowledge-Networked Media. Springer Berlin Heidelberg, pp. 7-24, 2009.

[25] Solr (Available online at: http://lucene.apache.org/solr/) [accessed: 13.4.2016]

[26] Apache Lucene Core (Available online at: https://lucene.apache.org/core/) [accessed: 13.4.2016]

[27] F. Ricci, L. Rokach, and B. Shapira, Introduction to recommender systems handbook, Springer, 2011.

[28] G. Adomavicius and Y. Kwon, "New recommendation techniques formulticriteria rating systems," IEEE Intelligent Systems, vol. 22, no. 3, pp. 48–55, 2007.

[29] M. Ekstrand, J. Riedl, and J. Konstan, "Collaborative filtering recommender systems," Human-Computer Interaction, vol. 4, no. 2, pp. 81–173, 2010.

[30] I. Palomares, F. Browne, H. Wang, and P. Davis, "A collaborative filtering recommender system model using owa and uninorm aggregation operators," in Intelligent Systems and Knowledge Engineering (ISKE), 2015 10th International Conference on, 2015, pp. 382–388.

[31] F. Sasaki and A. Srivastava, "Towards a Platform for Curation Technologies: Enriching Text Collections with a Semantic-Web Layer," The Semantic Web: ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29–June 2, 2016, Revised Selected Papers, 9989, 65.

[32] R. Mazza, A. Baldassari and R. Guidi, "Ondigita: A Platform for the Management and Delivery of Digital Documents," International Association for Development of the Information Society, 2013.

[33] E. Cambria and A. Hussain, "Sentic album: content-, concept, and context-based online personal photo management system," Cognitive Computation, vol. 4, no. 4, pp. 477-496, 2012.

[34] F.D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS quarterly, 13 (3), pp. 319-340, 1989.