MOVING

Deliverable 1.2: Initial implementation of user studies

Sabine Barthold, Irina Bienia, Till Blume, Sebastian Gottfried, Franziska Günther, Angela Fessl, Annalouise Maas, Ahmed Saleh, Ilija Šimić, Iacopo Vagliano, Michael Wiese

22/12/2017

Work Package 1: Requirements and use case development

TraininG towards a society of data-saVvy inforMation prOfessionals to enable open leadership INnovation

Horizon 2020 - INSO-4-2015

Research and Innovation Programme

Grant Agreement Number 693092

Dissemination level	PU
Contractual date of delivery	31/12/2017
Actual date of delivery	22/12/2017
Deliverable number	D1.2
Deliverable name	Initial implementation of user studies
File	MOVING_D1.2_v1.0.pdf
Nature	Report
Status & version	*Final v1.0*
Number of pages	37
WP contributing to the deliverable	WP1
Task responsible	EY
Other contributors	TUD, ZBW, KC, GESIS
Author(s)	Irina Bienia, Annalouise Maas, Michael Wiese *EY* Sabine Barthold, Sebastian Gottfried, Franziska Günther *TUD* Till Blume, Ahmed Saleh, Iacopo Vagliano *ZBW* Angela Fessl, Ilija Šimić *KC*
Quality Assessors	Markel Vigo, Aitor Apaolaza *UMAN*
EC Project Officer	Hinano SPREAFICO
Keywords	User studies, initial implementation, requirements

Executive summary

This deliverable includes three main sections that describe the implementation and priority status of the platform requirements and specifications (section 2), a typical navigation of the platform by target users of both use cases (section 3) and a look at the future user studies with tasks the use case partners consider applicable (section 4). In section 2, the platform requirements and specifications' status and outlook are described for the current version of the platform. Since the platform will continue to improve technically and content-wise, these are subject to change. As described in the DoA, deliverable D1.4 will provide an update on this. Section 3 expands on section 2 with examples of how the user should navigate through the platform in order to maximise the use of it. Section 4 includes possible tasks for future user studies regarding both use cases. Deliverable D1.3 will extensively describe the methodology and results of the evaluation of the user studies. Moreover, deliverable D1.3 will include an updated list of requirements and specifications.

Table of contents

List of Figures

List of Tables

Abbreviations

Table 1: Abbreviations

Abbreviation	Explanation
API	Application programming interface
ATS	Adaptive Training Support
DFN-AAI	Authentication and Authorisation Infrastructure in the German Research Network (Deutsche Froschungsnetz - Authentifizierung und Autorisierung Infrastruktur)
FDC	Focused web-Domain Crawler
ISA	International Standards on Auditing
LDAP	Lightweight Directory Access Protocol
MOOC	Massive Open Online Course
PEST analysis	Political, economic, social and technical analysis
SEC	Search-Engine based web Crawler
SSM	Social Stream Manager
SWOT analysis	Strengths, weaknesses, opportunities and threats analysis
VIA	Video Analysis

1. Introduction

1.1. History of the document

Table 2: History of the document

Date	Version
18.10.2017	v0.1: initial table of contents (ToC) draft
03.11.2017	v0.2 the revised ToC draft and an early description of content send to Quality Review
05.12.2017	v0.3: review-draft ready for QA
15.12.2017	v0.4: review-revised draft
21.12.2017	v0.5: partners' feedback on the draft
22.12.2017	V1.0: final version of the deliverable

1.2. Purpose of the document

The purpose of the present deliverable is to describe the status of the platform requirements' implementation, which was the result of the requirements analysis performed for deliverable D1.1. Moreover, the priority of the requirements will be discussed in regard to the two use cases and, where applicable, an outlook on future implementation improvements will be provided. Finally, a close look is taken at the typical use of the platform by target users and possible tasks to be utilised during future user studies, whose results will be the main focus of deliverable D1.3.

1.3. Structure of the document

This document is structured in 4 sections including an initial platform design evaluation plan, specifically addressing the implementation status of the requirements, as well as an outlook towards improving the platform to support those requirements. Section 3 describes the navigation of the platform by the targeted users according to the user scenarios also addressed in deliverable D1.1 for the requirements' analysis. Furthermore, in section 4, examples of tasks for future user studies are described.

1.4. Use cases, user studies, and the MOVING platform and systems

In order to set the scene for the present document, we feel it is beneficial to start by clarifying the relation between the MOVING use cases, user studies, and the MOVING platform and systems. As documented in D4.2, the main system-level output of the project is the integrated working and training environment, the MOVING platform. For the application of this generic platform in each of the two specific use cases of the project, certain customisations of it are needed, in accordance with the list of requirements that resulted from the requirements' analysis performed and described in deliverable D1.1. These customisations centre around the different content that the platform needs to be populated with, which is specific to each use case, and gives rise to two customised editions of the platform that we call "systems". For instance, the platform can use different datasets as sources for the search queries. In the system that supports use case 1 (scenario 1), the laws dataset is introduced to allow financial professionals to perform queries related to compliance with laws and regulations. The use of the MOVING platform/systems by target users of the two use cases is described in more detail in section 3 of the present document. However, to avoid any misunderstandings, we would like to clarify upfront that, at the software implementation level, we are not developing two separate systems but rather a single, widely-applicable, integrated software platform: the MOVING platform.

2. Initial review of implementation

In this section of the deliverable, we first present the working process of the requirements implementation (subsection 2.1), continuing with a description of the requirements' prioritisation (subsection 2.2). Furthermore, subsection 2.3 describes the work allocation and other connecting details, while in subsection 2.4 the status of the implementation and an outlook of further steps on the implementation are provided.

2.1. Working process of requirement implementation

The process of implementing the user requirements consists of the following iterative steps: prioritising the requirements, discussing feasibility and allocating the work. The first step is done by the use case partners, whereas the 2nd and the 3rd step is done in close cooperation with the developers and the entire consortium. The process is repeated until a conclusive list is created, which is accepted and implemented by both the use case partners and the developers. In order to instrument this process in a collaborative manner, we have decided to use the project management software Trello[1]. The management of this process in Trello enables the consortium to have an open and transparent cooperation not only regarding the implementation of requirements but also the feasibility and responsibilities. In Figure 1, a screenshot of the different boards in Trello is shown. Each board represents a functionality and is organised as a Kanban board, which means there are different areas for each step of the implementation, leading to the subsections: TODO, WORK IN PROGRESS, DONE, CANCELLED and BACKLOG, located on the board (Figure 3).

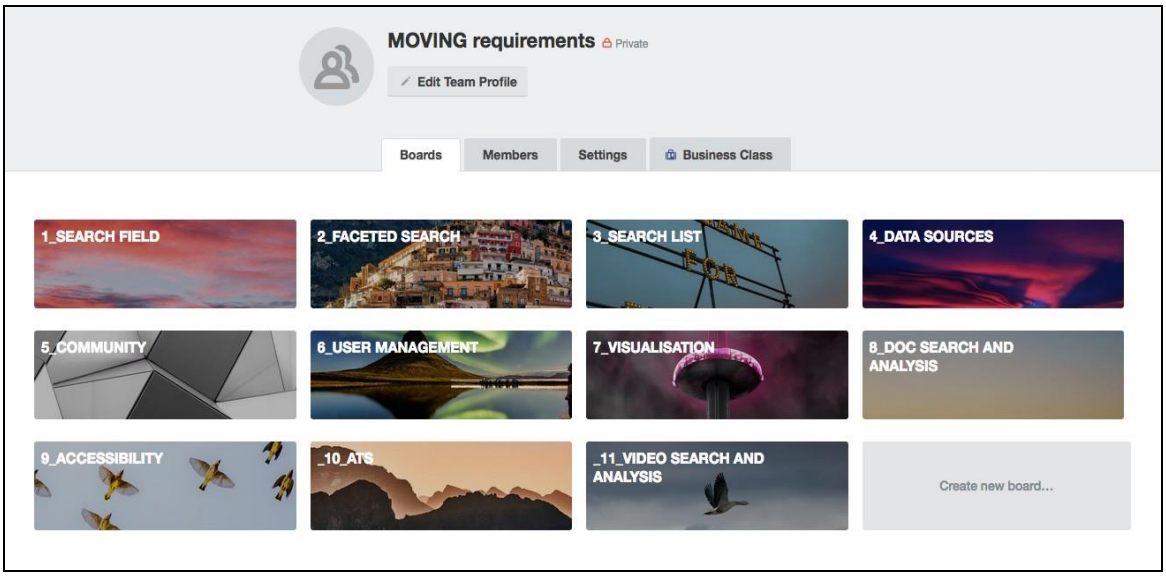

Figure 1: Trello boards for requirements management

[1] https://trello.com/ last accessed at 14.12.2017.

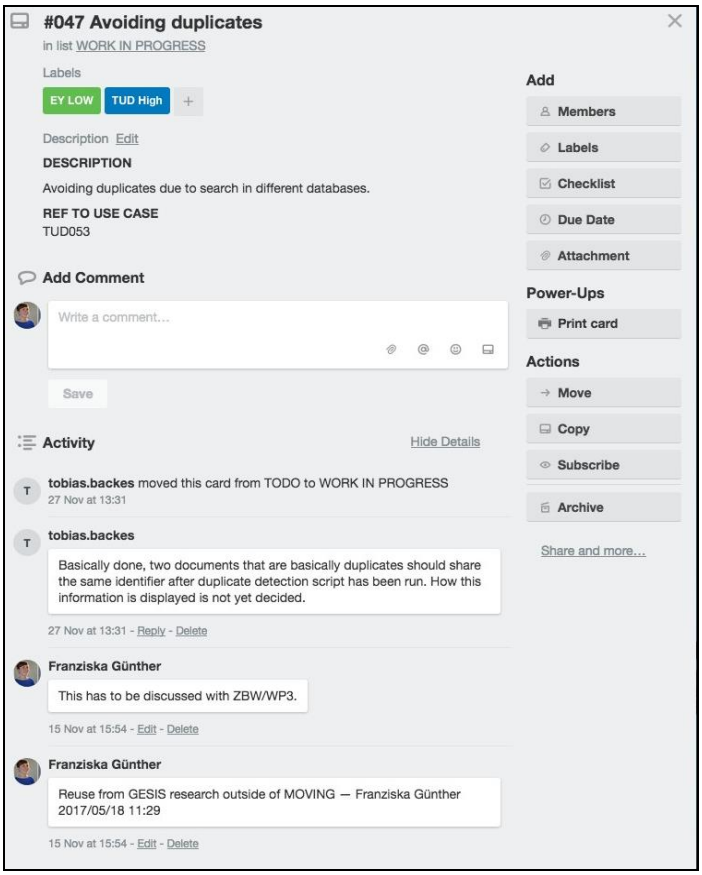

Figure 2: Trello card view of requirement #047

These subsections are also called lists. The cards within the lists can be dragged and dropped into other lists. On each card (Figure 2), a member has the possibility to comment on the request, but also to attach a file or a link.

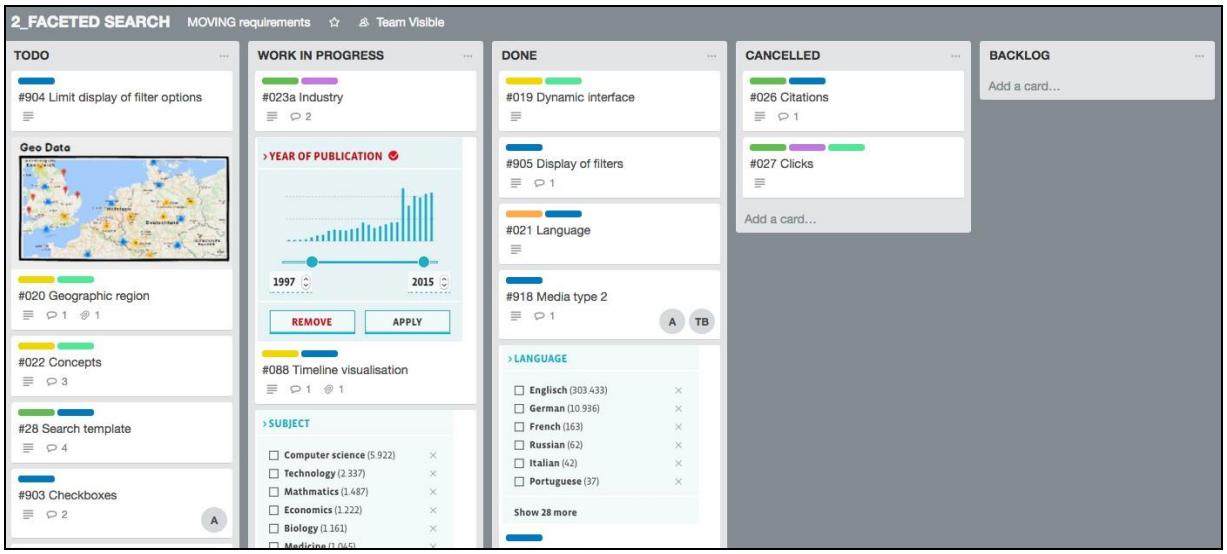

Figure 3: Trello board "Faceted Search"

2.2. Requirement prioritisation

The requirements analysis performed in Task 1.1 and reported in deliverable D1.1 resulted in an extensive list of user requirements, which are subdivided into different functionalities (e.g., ATS and visualisation). As a next step, their implementation needed a prioritisation into three classes (i.e., low, medium, high) in order to ease the process of implementation for the technical partners. Being the result of two requirements analyses, for the two different use cases, a number of requirements have been assigned different priority classes by the two use case partners, EY and TUD. This is in concordance with the scenarios and the user studies performed in Task 1.1. As an example, the priorities of the visualisation requirements, in connection to the two use cases, range from "low" to "high" for use case 1 (Research on business information by public administrators) and hold a constant "high" across the requirements for use case 2 (Managing and mining research information).

Specifically, most of the visualisation requirements have a high importance for use case 1 and the user tasks that are needed to perform as a financial professional while following the ISAs on an audit engagement. Nevertheless, specific requirements are just as important for target users of use case 2, hence their priority was set to "high" by both use case partners. The top priority for further implementation within the visualisation development will be given to a tag cloud functionality to display and interact with most relevant keywords extracted from the search result contents. These shall primarily be used to filter the results set, however, we also plan to use them for the re-ranking of results. Finally, we will be adding visualisations supporting filtering along temporal, geographical and statistical dimensions.

2.3. Wording, feasibility and allocation of work

When building requirements with appropriate characteristics (e.g. unambiguous, complete, correct, mandatory, readable), the development partners have to be involved "to ensure that the requirements are feasible and verifiable" (Firesmith, 2007, p. 18). Discussing the requirements has led us to decide to change, adapt and reformulate several requirements. The wording, too general in some respects, i.e., a more detailed/technical explanation is needed, was closely reviewed and revised. For example, the description of requirement no. 004 "Combination of words" was extended from "it should be possible not only to search for one word but for combination of words" to "Compound words (e.g. computer science, data mining) should be recognised as single not separate search term, for example, by using a thesaurus or Boolean operators". In addition, requirements that would lead to functions with a high degree of complexity and high resource consumption were detected. Here we need to limit the function to the most important ones, whose implementation is also feasible within the project period. This includes, for example, the requirement no. 131 "Reference management system on the platform" with the description "Organising and comparing the search results in a reference management on the platform". The requirement would be a new comprehensive functionality. Hence, this could not be implemented within the project period, while the focus is, currently, being placed on other functionalities. Potential users of the MOVING platform were approached and it was found that a plug-in for a reference management outside the platform

would be sufficient. Nevertheless, the idea of integrating a literature management system into the MOVING platform is not completely cancelled but postponed.

2.4. Status of implementation and outlook

This section summarises the status of implementation of requirements and functionalities of the platform. Moreover, it provides an outlook for further implementation improvements and activities, where applicable. An extensive description of the platform prototype is provided in deliverable D4.2.

Responsiveness

The technical aspects of the MOVING platform's responsive design were discussed in detail in deliverable D4.2 and will be summarized here in short. As originally described in deliverable D4.1, the Bootstrap framework[2] is used to implement the responsive design making. This allows the platform to be easily and efficiently accessible via multiple types of devices, ranging from desktop web clients to smartphones and tablets. The mock-ups of the MOVING platform, developed with the Balsamiq Software[3] and described in deliverable D1.1, were used as a starting point for the implementation of the current version of the responsive design. Figure 4 shows the search environment of the platform designed in Balsamiq. In accordance with these mock-ups, we organise the different functionalities in the same way in all the views of the application. Particularly, the faceted search is placed on the left side of the MOVING platform (Figure 4), the visualisations in the middle and the ATS on the right site. Also, each responsive design view hosts on the upper part the main navigation buttons (“search”, “projects”, “learning” and “community”), and the button “My account”. In response to initial usability tests conducted at TUD, some parts of the mock-ups were not implemented due to overlapping corresponding functionalities. Nevertheless, most mock-ups were implemented exactly as documented in deliverable D1.1. More details regarding the implementation and the differences between the early version of the mock-ups and the implemented responsive design are elaborated in detail in deliverable D4.2 (subsection 3.3). All views adapt to different screen sizes, where the layout is automatically changed according to the size and capabilities of the device. For example, on a PC screen, the users see the content in a three-column view, as depicted in Figure 9, on a mobile phone, the content is presented in a single-column view, and on a tablet the same content is displayed with the menus on the top of the screen, as shown in **Figure 5** (left) and (right).

[2] http://getbootstrap.com, last accessed at 22.09.2017.

[3] https://balsamiq.com, last accessed at 22.09.2017.

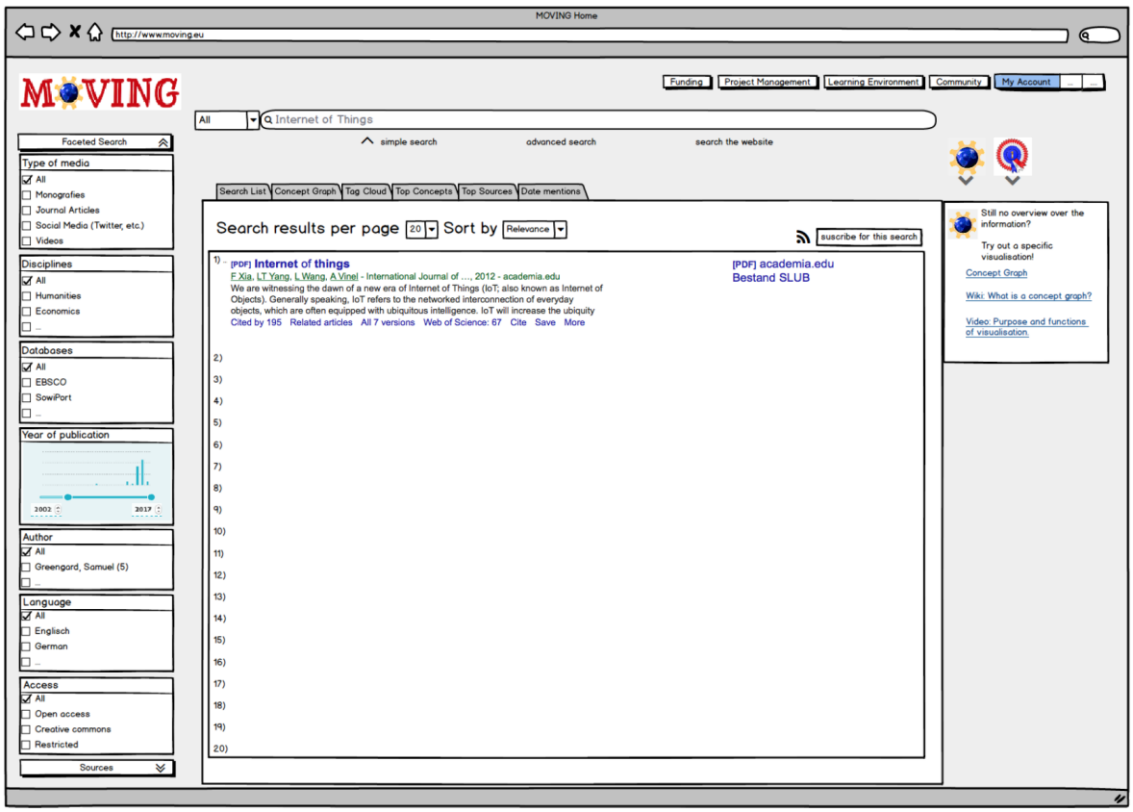

Figure 4: Mock-up of the search environment

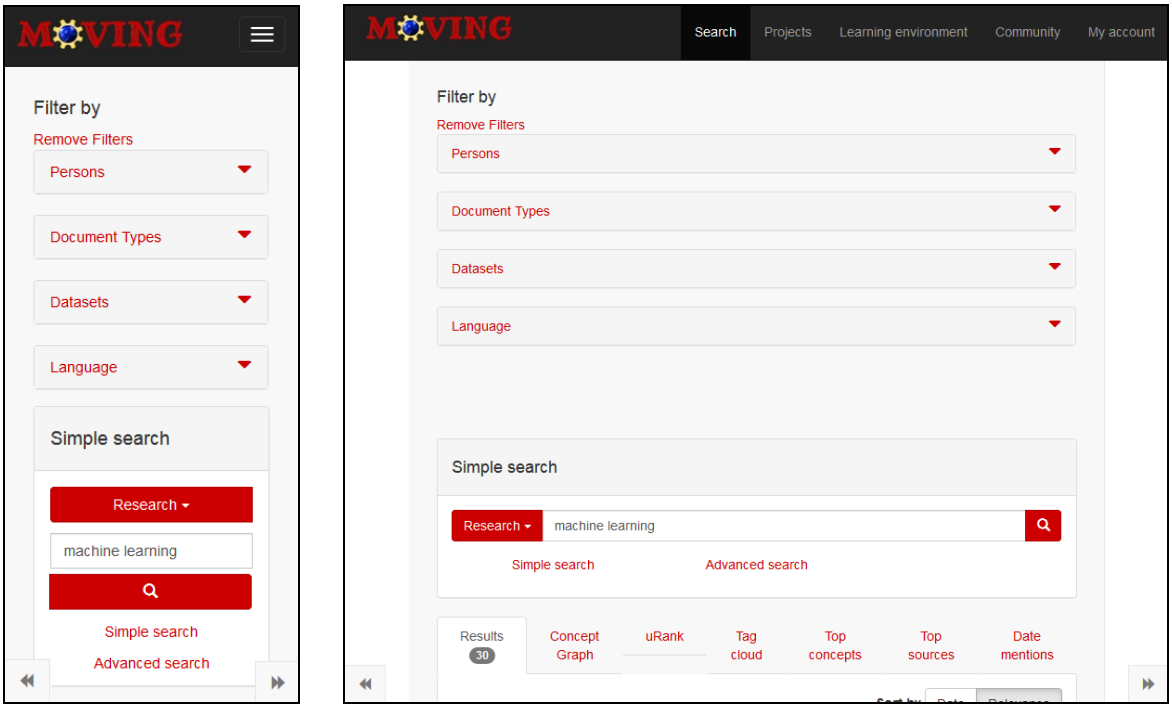

Figure 5: Example of the MOVING responsive design on a smartphone (left) and tablet (right) screen

Adaptive Training Support

The ATS pursues the overarching learning goal "acquire digital information literacy skills and competencies". This goal can be divided into the following two sub-goals: support for learning on how to use the MOVING platform and support for learning content. To be able to support both goals, the ATS consists of two different widgets integrated into the MOVING platform's user interface:

- "Learning-how-to-search" widget
- Curriculum Reflection widget

A first version of the "Learning-how-to-search" widget is already implemented in the current version of the MOVING platform. This widget presents reflection interventions that visualise search behaviour and aim to challenge users to experiment with different search functionalities and to reflect on search efficiency (feature-based adaptation). Detailed information about the widget, its goals and its implementation can be found in deliverables D2.1 and D4.2.

The development of the "Learning-how-to-search" widget is almost finished. Only minor improvements and bug fixes are planned in the 2nd half of the project. With regard to the use cases, this widget is kept in a very general manner so that there will be no individual implementation for the different use cases as the features provided by the MOVING platform will be relatively the same.

The development of the "Curriculum Reflection" widget has started immediately after the first release of the MOVING platform. It will present interventions that visualise on the one hand the learner's progress with respect to the digital information literacy curriculum and aim to challenge users to reflect on own proficiency in terms of the curriculum (content-based and feature-based adaptation). On the other hand, it will present recommendations referring to the curriculum in order to offer the learning content tailored to the learner's learning progress. In the first version of this widget, we will focus on the TUD use case. The focus will be put on supporting the following three major skills and the corresponding subskills: search, communication and collaboration, and content creation. It is planned that the "Curriculum Reflection" widget is implemented in such a way that the general functionality can be easily transferred to the curriculum of EY.

The previous knowledge self-assessment presented in Figure 6 shows the user interface used for the opt-in to the user interaction tracking. If the user decides to use the ATS, and here especially the "Curriculum Reflection" widget, s/he is asked to answer a short questionnaire, which captures the prior knowledge that the user has with regard to the three digital competence areas (Figure 7) – "Digital information and data literacy", "Digital communication and collaboration" and "Digital content creation" – that will be part of the MOVING curriculum. This curriculum will be specified further in deliverable D2.2 after month 24. The score will be stored in the user profile and will be used to present the individual's learning progress with respect to the curriculum and the skills already gained. As the widget cannot directly track the learning process of a user, it will collect on the one hand, which suggested resources the user has clicked on (knowing that we cannot extract learning progress from this but just get some ideas the user is interested in), and on the other hand, the widget will ask from time to time if the user has the impression that s/he has gained a subskill.

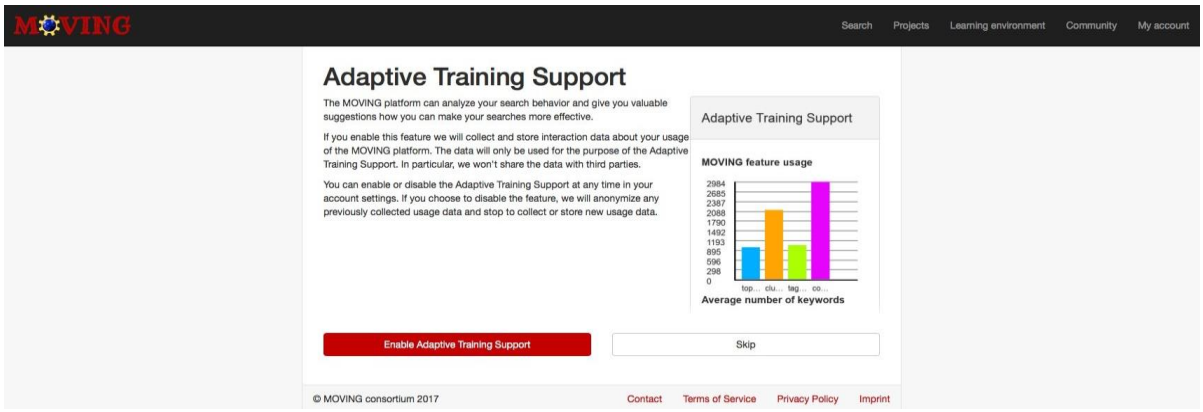

Figure 6: Opt-in for the ATS as part of the MOVING platform on-boarding process

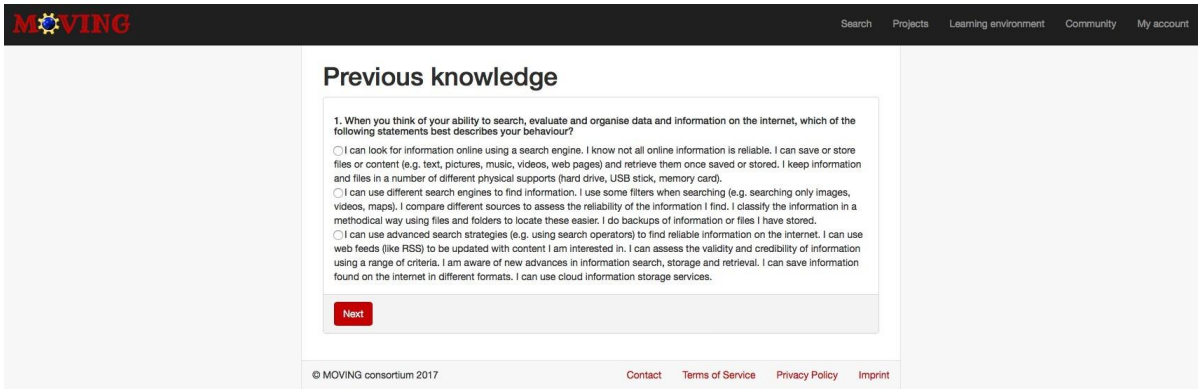

Figure 7: Previous knowledge assessment for ATS as part of on-boarding

Community

The MOVING web application is based on the existing eScience[4] platform. It is a working environment for the advanced digital initiation, organisation, implementation and documentation of projects in the academic field. Using eScience as a basis of the MOVING platform allowed us to (1) reuse and extend its features to meet the requirements derived from the use cases defined in deliverable D1.1, and (2) create an integrated working and training environment. Users of the platform can create projects on the MOVING platform, in which they can collaborate with other users. For each project, the user can choose from a set of modules to enable specific functionalities. Currently, there are modules providing a project wiki, task management, time tracking, document management, blogs, file storage, polls and message boards. Additionally, users can maintain a profile page with their research fields and competencies that are accessible to other users via search. Finally, there is a messaging system, which allows users to connect with other researchers registered on the MOVING platform.

[4] https://www.escience-sachsen.de/, last accessed at 12.01.2017.

Each user is assigned an individual profile page where s/he can specify a set of individual skills and indicate research areas in which he or she is active. This information can be utilised to search and connect with other researchers on the platform. Users also have the option to upload a profile picture and individualise their dashboards with projects and activities (Figure 8). The community function of the MOVING platform will be updated according to the findings of the user studies after month 24.

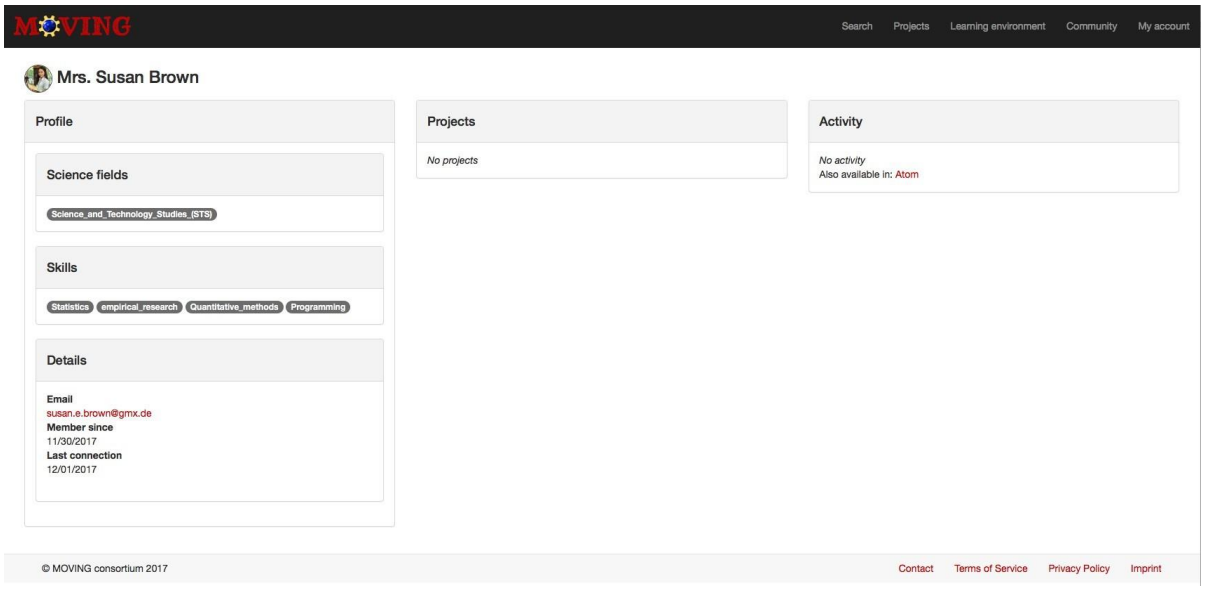

Figure 8: User Profile view

Crawling and video processing

Three web crawlers are implemented for the MOVING platform to provide data from three web sources: user-specified websites, social media and Google search. The crawlers run as background services continuously collecting and indexing the crawled data. The **Focused web-Domain Crawler (FDC)** collects HTML pages from user-specified websites, such as companies' websites and online courses. The crawler supports duplicate detection and allowing/blocking URL patterns for each website.

The **Social Stream Manager (SSM)** collects URLs from social media posts related to user-specified topics and indexes those HTML pages. The SSM also supports the indexing of social media videos (title, metadata, etc., as well as additional video annotations generated by invoking the MOVING platform's Video Analysis service (VIA), discussed below). The social media platforms currently supported are Twitter, Google+ and YouTube. Future improvements will include the indexing of social media posts and support for location information extraction, crawling of whole YouTube playlists, monitoring specific social media accounts, and adding Facebook to the list of supported social media.

The **Search-Engine based web Crawler (SEC)** exploits the Google API to search for user-specified topics on the web and index the retrieved HTML pages. The VIA is a web service hosted in CERTH that

performs video analyses (temporal segmentation and visual concept annotation). At this point, the SSM utilises the VIA by issuing requests to analyse social media videos and receive the visual annotations from the service, which are then added as metadata to the indexed video document. Future plans include the extension of the video annotation feature for videos embedded in web pages crawled by the other crawlers.

Data sources

The MOVING platform provides access to a large variety of documents coming from different data sources. Currently, we distinguish between the document type (book, article, video, etc.) and the document source. For example, crawled websites or professionally created metadata from the ZBW economics corpus. All types of documents are stored in the search engine based on a common data model and are subject to quality tests before being indexed. The MOVING platform currently provides access to economics literature (ZBWEconomicsDataset), social sciences literature (OAFulltexts, PublicationMetaData), laws and regulations from Wolters Kluwer, cross-domain literature crawled from the Web of Data (BTC2014), various crawled websites specified by the use case partners containing learning material, funding opportunities, information about organisations, transcripts of the videolectures.net videos, as well as information about projects (ProjectMetaData). About 99% of the documents contained in the above-mentioned datasets passed our quality tests and can be accessed via the MOVING platform. Additional required data sources, e.g., containing information about research datasets, library catalogues, or potential project and business partners, will be integrated until the end of the project depending on possible technical and legal limitations.

Search functionalities and document analysis

The MOVING platform needs to process huge amounts of text data coming from different data sources that contain different document types (see the previous paragraph). To handle these data efficiently, the MOVING search engine was implemented on top of Elasticsearch[5], which provides a scalable real-time search and supports multiple document types, different file formats and different programming languages. We developed a novel ranking model, called HCF-IDF, as a plugin in Elasticsearch for the ranking of search results based on their relevance to the user query (Nishioka & Scherp, 2016).

[5] https://www.elastic.co/products/elasticsearch

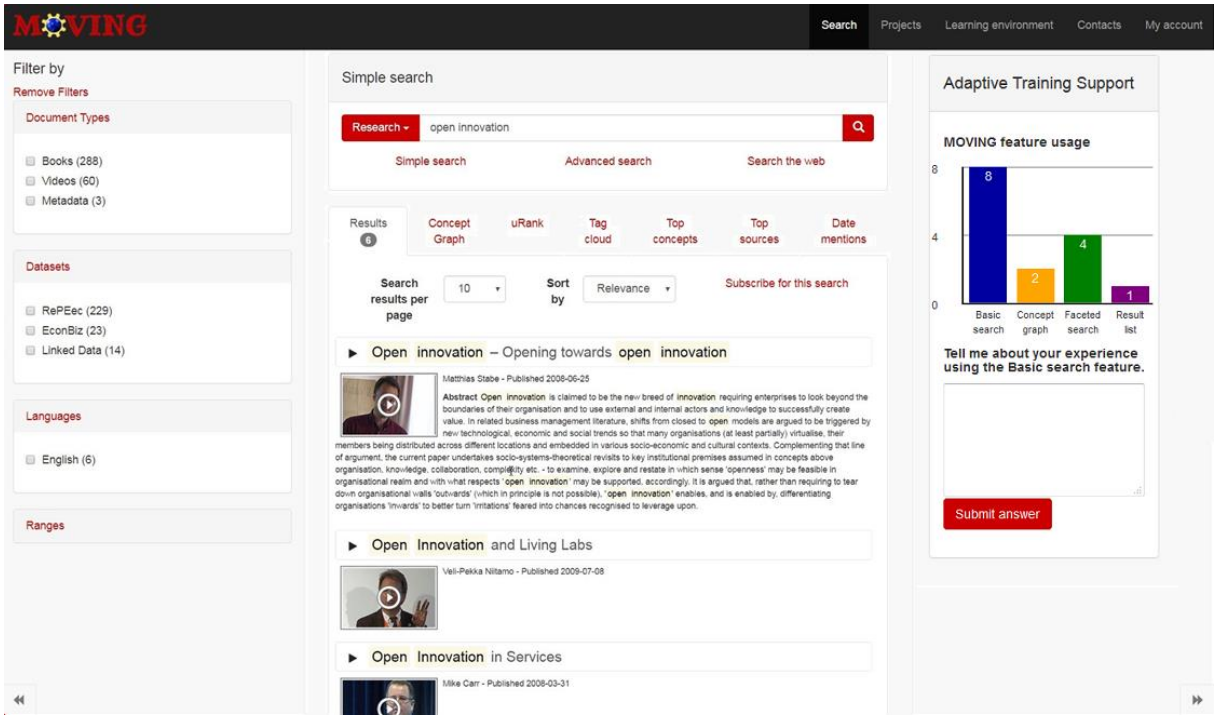

Figure 9: The MOVING search page

HCF-IDF extracts a set of concepts from the text. These concepts are defined in controlled vocabulary thesauri and extracted from the documents by text matching. Similarly, we plan to extract entities (e.g., persons, locations and organisations) from the documents. For this purpose, we prepared a new version of the common data model of the MOVING platform which allows storing different types of entities. More details about the ranking models used are available in deliverable D3.1, while D2.1 contains additional information on HCF-IDF and concept extraction. The upcoming deliverable D3.2 will more extensively explain the updates on the common data model and the progress on concept and entity extraction. The search page (Figure 9) provides a basic search through a classical search field. Advanced search features include the possibility to search for specific fields as title, author, and abstract. The search results can be refined by means of the facets. The current version of the search page is based on the MOVING platform's requirements and mock-ups, which are described in detail in deliverable D1.1.

The search field was implemented as specified in the related requirements (see deliverable D1.1, section 6.1), but some advanced features need to be improved. For instance, support for Boolean operators has still to be implemented (e.g., searching for "machine learning" AND "data science"), and when searching for terms entered in multiple fields, only results which satisfy both searches are returned (it is possible to search for *title* = "Genetic Implications of the structure of Deoxyribonucleic Acid" AND *author* = "Watson", but not *title*=" Genetic Implications of the structure of Deoxyribonucleic Acid" OR *author*="Watson"). These features will be included in the next version of the platform. The faceted search component (deliverable D1.1, section 7.1) has been implemented through drop-down menus by using multiple checkboxes for every single facet. At this stage of the

projects, the facets are not yet complete. An updated version will be implemented after month 24 when the updated version of the mock-ups will be available (e.g., filtering by discipline is desired but the corresponding facet is currently not implemented, as this information is unavailable in the data integrated so far). Another new interesting feature is the possibility to view the history of laws and regulations, as displayed in Figure 10. For example, this may help audit professionals (one of the user groups addressed in use case 1, see section 4.1) to track the evolution of these documents over time and refer to a specific version if needed. The laws and regulations dataset was provided by Wolters Kluwer within the H2020 ALIGNED project[6]. To sum up, most of the requirements for the search have been implemented as specified. Additional information on the status of the implementation is available in deliverable D4.2. Some requirements, specified in deliverable D1.1, assumed the availability of the full text in a specific format, e.g., PDF. Since most of the documents are not openly accessible, currently the platform does not host them but points to their source, where they can be possibly accessed. We plan to implement a document upload function to enable users to add their own documents to the platform and analyse them. This scenario has to be further defined and the requirements adjusted consistently (e.g., make the uploaded document private only or available to other users) in order to subsequently implement them.

The author name disambiguation is implemented and integrated with the other MOVING functionalities. Initially, all author mentions with the same name are considered to be one author. The disambiguation script can be applied regularly in order to check if some names need to be split into different authors. The parameters of the methods still need to be fine-tuned. Unique identifiers provided alongside the authors in the metadata are available for a subset of documents in the MOVING platform and can be for this task. Currently, the disambiguation information is used by retrieving all documents authored by the disambiguated person if the user clicks on the author's name in the document's metadata.

[6] http://aligned-project.eu/

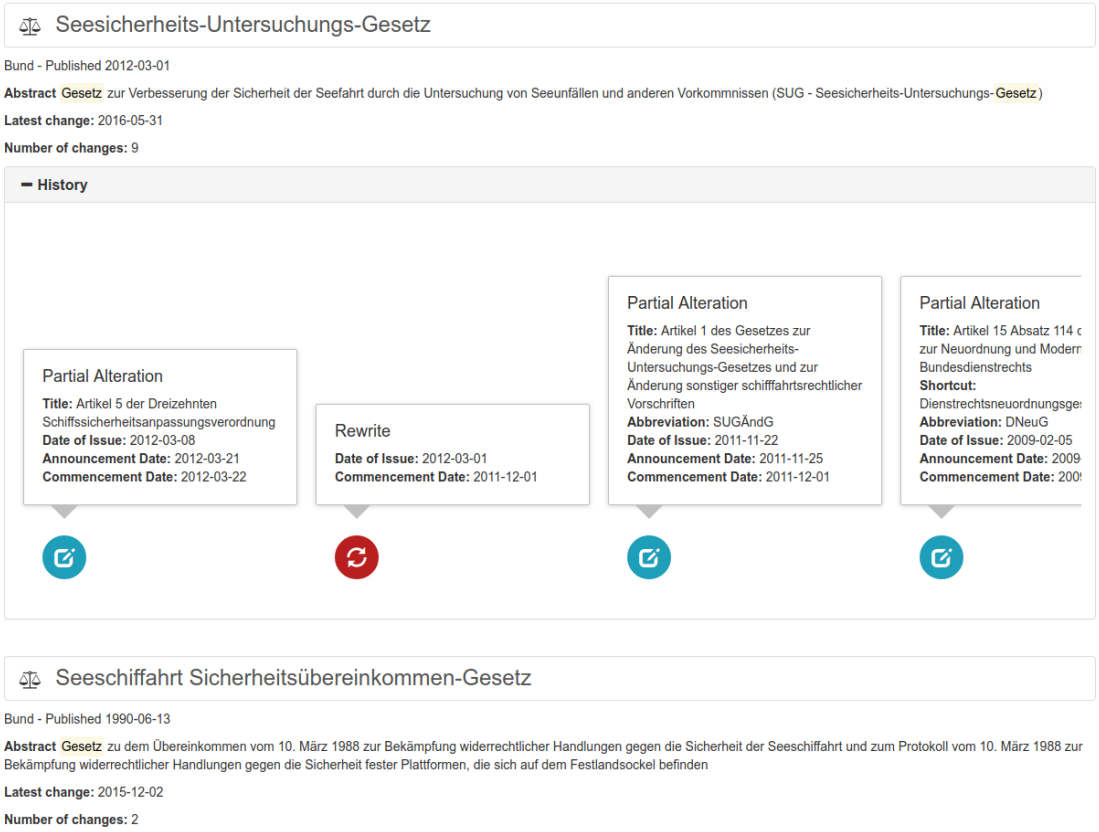

Figure 10: Historical view of laws and regulations – MOVING Platform

Duplicate detection is also implemented and integrated. Initially, all documents have a unique identifier value. The deduplication script can regularly be applied in order to check if some documents need to be assigned the same identifier. It is not yet clear how this information will be displayed to the user in the final version of the platform. With a different set of features, the method can also detect similar documents, which could be displayed to the user, e.g., in a list of most similar documents. However, this would require saving the actual similarity value for each pair of documents, which is an integration step not yet undertaken.

Visualisations

The implementation tasks regarding the concept graph visualisation are nearly finished. The defined interactions on the graph have been added, such as zooming, movement of nodes, additional information on node hovering and graph exploration. Additional features supporting the graph usage have also been completed, which include the node colouring based on their type, the node size based on the number of ongoing and outgoing edges, the graph completion indicator, and the help functionality. The concept graph in its current form is presented in Figure 11. The top results are ordered in a circle based on their relevance for the search term. By clicking on a node, the graph expands to display additional search results. A further interaction was added to the nodes, which allows the user to create a subgraph outgoing from the target node. This can be achieved by using the node ring menu, which gets activated by a right-clicking a node with the mouse cursor. The ring

menu and an example subgraph can be seen in Figure 11. Clicking on the rings will expand the neighbourhood of the node, where the distance of a ring from the node represents the size of the expanded neighbourhood measured in hops (distances between nodes). Furthermore, interaction tracking has been added to the concept graph.

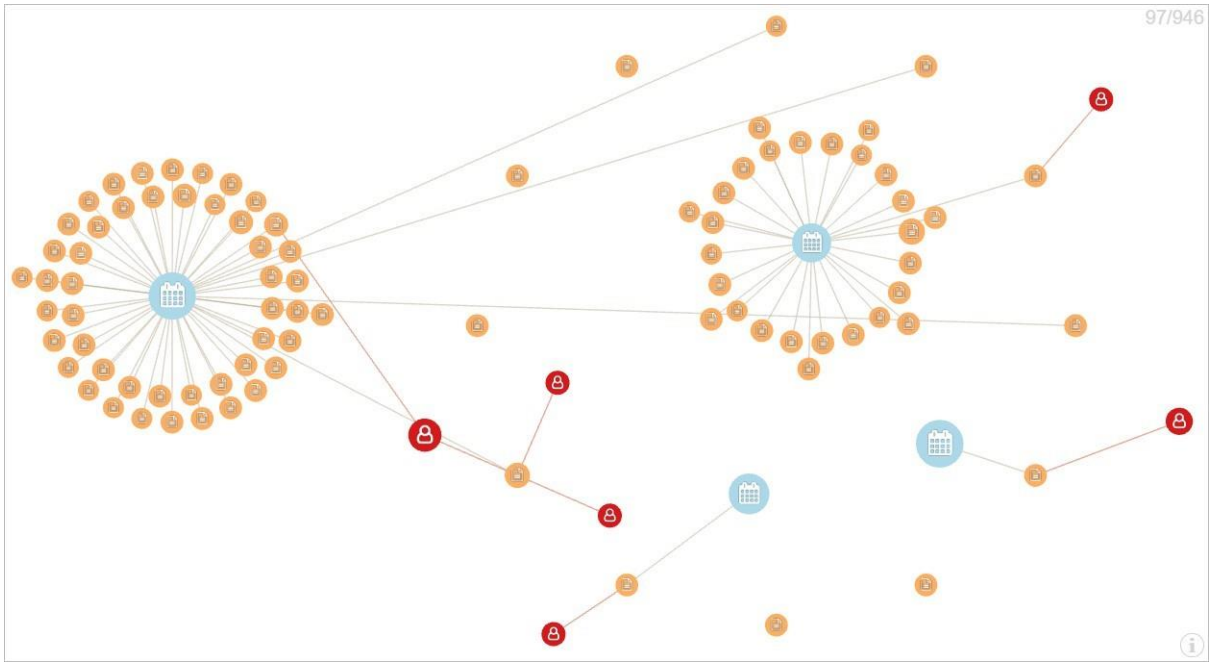

Figure 11: The concept graph with expanded nodes

Because of the limited number of different node types and connections in the platform's prototype, filtering the nodes based on node type is not yet implemented. Otherwise, hiding a node type in the visualisation would cause many nodes to become unreachable, as the intermediate connection would be lost. The co-occurrence of concepts has to be carefully planned since it will add a vast number of additional connections to the graph, which could become more of a hindrance than a support.

The requirements will be continuously worked off, starting with the high priority requirements. In the end, the MOVING platform will be extended with multiple interactive and explorative visualisations. The uRank interface, shown in Figure 13, includes a topical tag cloud and provides means for user interest-driven filtering and the exploration of search results.

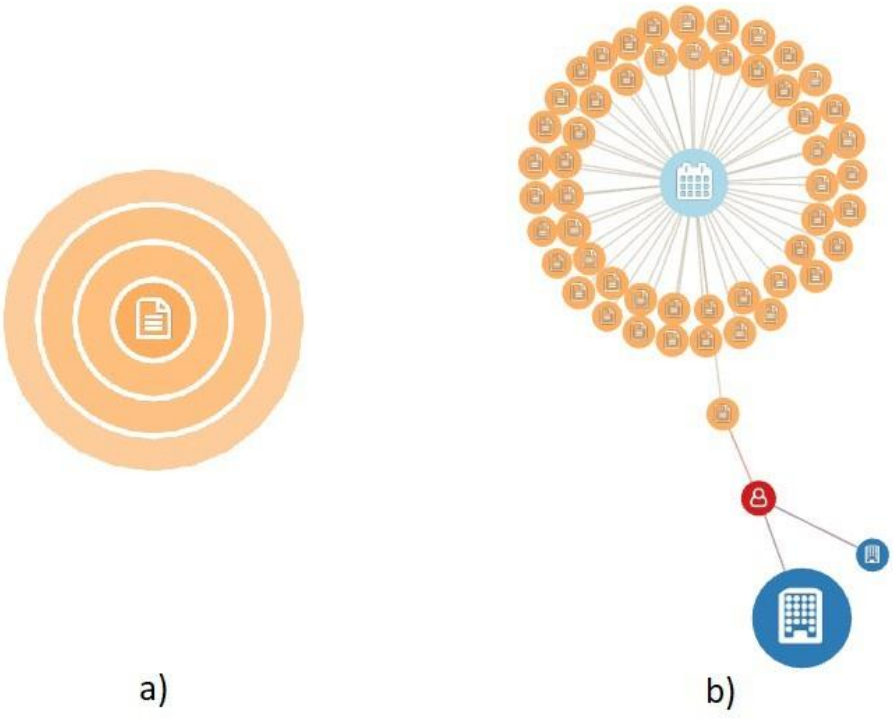

Figure 12: a) the expanded ring-menu of a node; b) the created subgraph from the ring-menu

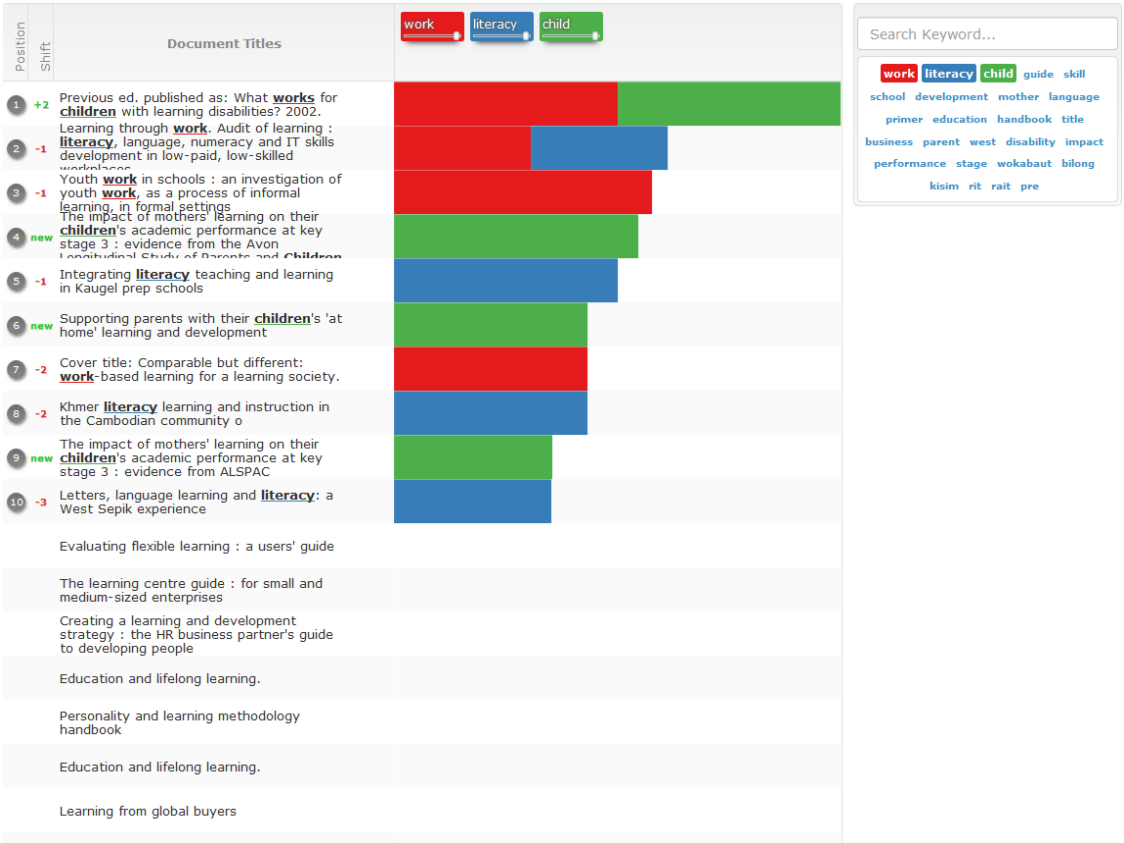

Figure 13: An early stand-alone version of uRank showing search results for "robot" query. On the left side, a tag cloud-like visualisation shows the most relevant keywords extracted from titles and abstracts of the search results. Selecting keywords of interest (e.g., "human", "robot", "interaction", on top right) filters and re-ranks the results accordingly. The contribution of the keywords to each hit is visualised by a stacked bar chart (di Sciascio et. al, 2016).

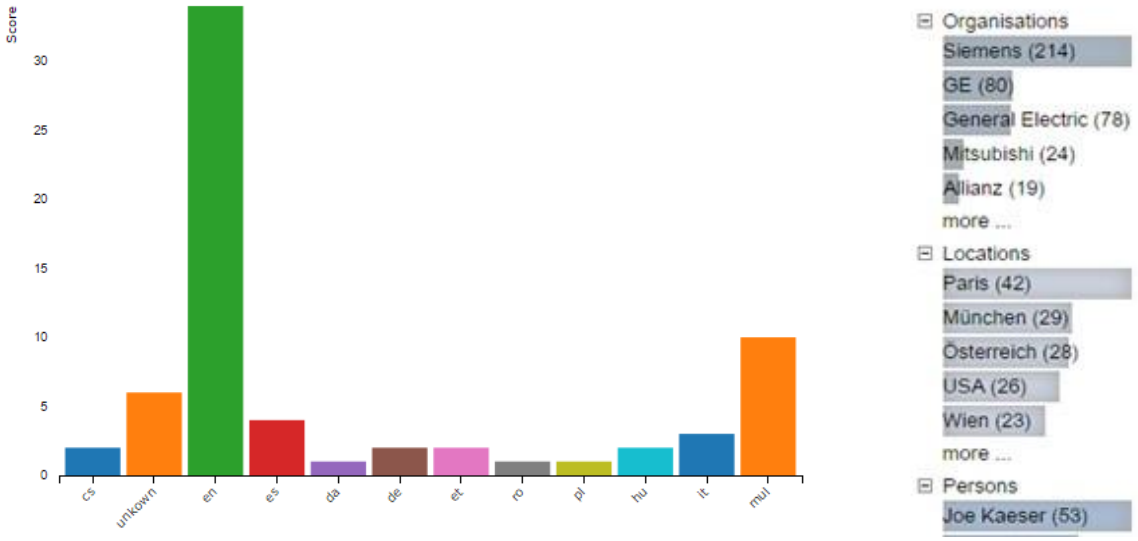

Figure 14: Bar chart on the left side showing the distribution of languages for a given set of search results. Clicking on a bar will filter the search results accordingly for the respective language (Tschinkel et. al, 2016). On the right side, the occurrences of extracted metadata (e.g., organisations, locations, persons, etc.) are shown in a tree view grouping entities along with their types. Selecting multiple entities will filter the search results accordingly (Rauch et. al, 2015).

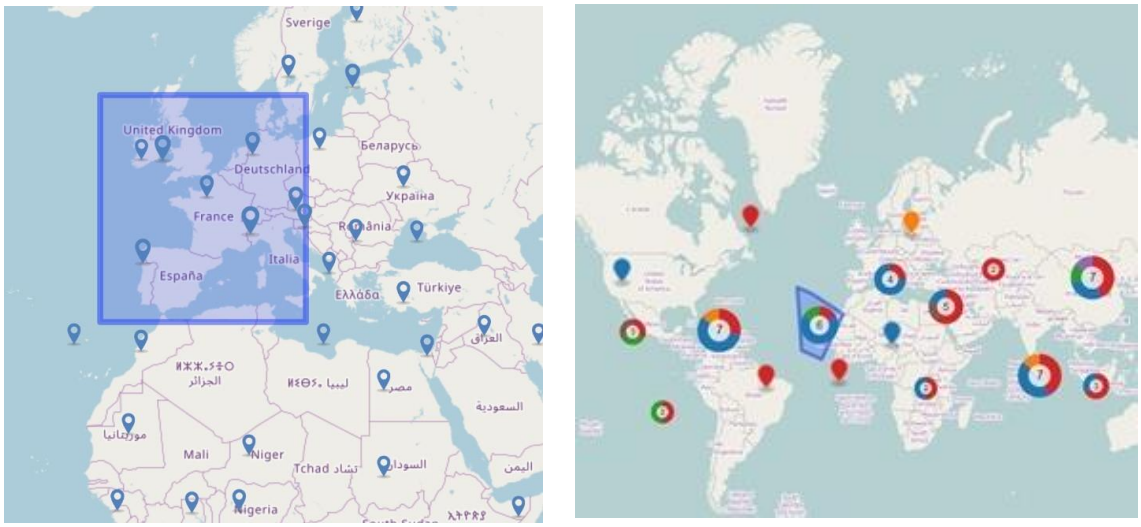

Figure 15: Geo-visualisation showing a geographic distribution of geo-references extracted from the search results (on the left side) (Tschinkel et. al, 2016). The user can filter the search results by selecting an area on the map (blue rectangle) (Rauch et. al, 2015). The distribution of the selected metadata, such as topics encoded by different colours, can be shown by using the embedded doughnut charts (on the right side).

Simple bar chart-based visualisations that are planned to be implemented, as shown in **Figure 14**, show the distribution of metadata and/or extracted entities in the result set. Clicking on the bars will filter the result set. Geo-visualisation, shown in **Figure 15**, can be used for visualizing geographic distributions and filtering of search results along the geospatial dimension. Time based-visualisations, shown in **Figure 16**, display the distribution of time references attached to the search results and support filtering of hits along the temporal dimension.

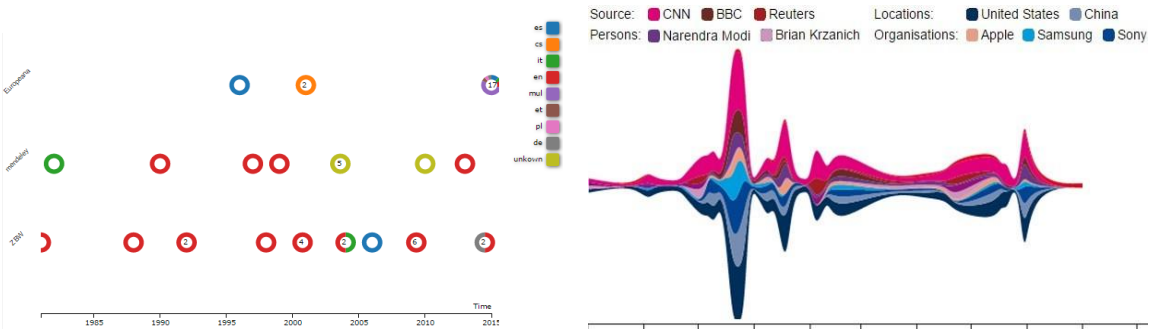

Figure 16: Time-oriented visualisations show the distribution of time references. On the left, creation dates are shown for results from three different data sources (shown on the y-axis). Language is encoded by colour, with doughnut charts aggregating the language distribution for results having equal (or overlapping) time stamps (Tschinkel et. al, 2016). On the right, trends for selected entities (e.g., sources, extracted locations, persons, and organisations) are shown using a stream graph (Rauch et. al, 2015). Entities are colour-coded, with the thickness of each stream corresponding to the number of entity naming (i.e., mentions in the results) at the particular time point.

User management & access

User management systems

It was an initial requirement from the DoA that the platform should have the capability to be integrated into the existing user management facilities of institutional IT environments. The MOVING platform supports multiple forms of authentication. In addition to the in-app user management, it supports LDAP[7] and Shibboleth[8] authentication, both of which enable the platform's integration into existing user management systems. The latter is especially popular among research institutions emerging as the leading single sign-on solution. The DFN-AAI federation, in which German research institutions using Shibboleth organise, currently lists 203 members.[9] LDAP, on the other hand, should be a good fit for the deployment of private institutions as it allows to delegate user management to Active Directory[10] or any other directory services. We assume that this requirement is thereby fulfilled.

User Self-Registration

New users to the MOVING platform are asked to create a user account before they get access to the functionalities of the platform. When a new user is signing into the MOVING platform for the first

[7] https://tools.ietf.org/html/rfc1487, last accessed at 20.01.2017

[8] https://shibboleth.net/, last accessed at 20.01.2017

[9] https://www.aai.dfn.de/verzeichnis/idp-dfn-aai/, last accessed at 13.01.2017

[10] https://technet.microsoft.com/de-de/windowsserver/dd448614.aspx, last accessed at 13.01.2017

time, s/he is asked if he or she wants to allow the data collection process. To comply with the European data protection law, the user interaction tracking is turned off by default. Only when the user agrees with the collection of his or her interaction data, the tracking code is enabled. The user can enable or disable the user interaction tracking at any time via the profile settings. Figure 17 shows the on-boarding procedure for new users. New users have to create a profile and agree to the terms and conditions as well as the privacy agreement. The user also decides whether he or she wants to be found in the community search and if he or she wants the account to be private or visible to other users of the platform (Figure 17).

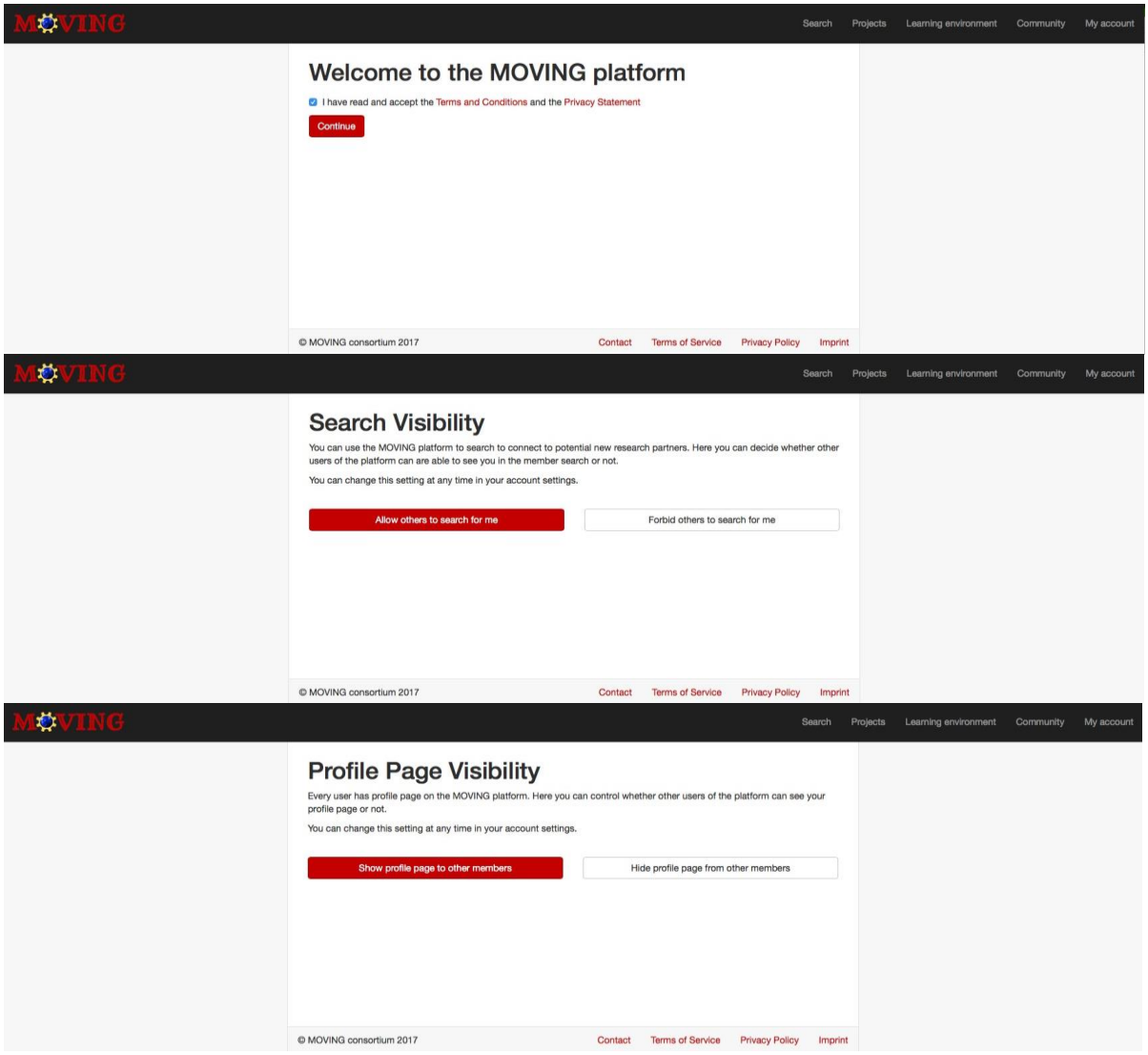

Figure 17: Privacy settings, MOVING platform on-boarding

3. User platform navigation

This section includes a walkthrough of the platform with details on how a typical user should navigate through the platform and use it following typical user scenarios in order to maximise the benefit from the platform's functionalities.

3.1. Use case 1: Research on business information by public administrators

This subsection will provide typical scenarios for financial professionals, target users of use case 1 (EY), and describe how the target users should navigate through the platform in order to maximise the productivity of their assignments. For that purpose, we will follow a number of ISAs and describe which requirements need to be met by the users. The following examples are similar to other ISAs described in D1.1.

ISA 315

This standard requires the professional to identify and assess the risk of material misstatement through understanding the entity and the environment. The MOVING platform supports these risk assessment procedures by providing information about entity's industry, regulatory framework, financial reporting framework, the nature of the entity, such as operations, ownership and governance structures, investments, structure and finances, accounting policies, objectives and strategies, measurement and review of financial performance. Using the search interface to perform a query of the entity is the first step in any analysis. The results of the query should be filtered using the features on the left-hand of the platform. By reducing the number of query outputs, the professional should be able to identify the most relevant results for the task at hand. After filtering, the remaining search results will help with answering the user's questions. The tag cloud and top concepts features (currently not implemented) should be used first, allowing the user to look at the frequency and relevance of words present in the query results. Thus, these features provide an understanding and initial overview of the results. The concept graph feature of the platform should be extensively used to find connections between the entity and other concepts. The user should, therefore, explore the concept graph and zoom into areas of interest to identify those connections. If the query output list is rather complex (a long list), the user should make use of the uRank feature to rearrange the results based on relevance (see subsection 2.2).

ISA 250

This standard requires the professional to responsibly consider laws and regulations and changes therein in the audit of financial statements. As described in scenario 1 of use case 1 from the DoA, the auditor is required to identify changes in laws and regulations in regards to compliance of entities like organisations, companies or research institutes. Some entities operate in highly regulated industries, others are required to comply only with regulations generally related to business operations. The financial professional is required to identify (1) the applicable laws and regulations

impacting the financial statements and (2) any instances of non-compliance. The MOVING platform, with its historical view feature, offers the user the possibility to track the changes of laws and regulations over time and evaluate the entity's compliance. For that purpose, after the search query of specific laws and specific filtering, the user should switch to the historical view feature to identify changes relevant to the financial statement. Next, the user can use the concept graph to search for connections between laws and regulations applicable to the entity, the sector of operations and to identify any relations that indicate instances of noncompliance.

ISA 550

This standard requires the professional to determine related party relationships and transactions in an audit of financial misstatements. It specifically expands on ISA 240, ISA 315 and ISA 330. The nature of related party relationships and transactions may implicate a higher risk of misstatement. Therefore, the complex range of relationships and structures can be visualised and analysed using the concept graph feature of the platform.

3.2. Use case 2: Managing and mining research information

The second use case provided by TUD specifically addresses the information and training needs of young researchers like PhD students and advanced Master students working on their thesis projects. It presents four scenarios highlighting different stages of a research project for which the MOVING platform can be utilised. In the following section, the functionalities of the platform will be illustrated along scenario 1: "state-of-the-art on a research topic" to show how the user requirements that correspond with this use case scenario are already implemented in the current prototype of the platform. This scenario describes how a user ("Ms Brown") interacts with the search feature of the platform to get an overview of a research topic she has little or no previous knowledge about.

Scenario 1: State-of-the-art on a research topic

Ms Brown starts the query with the very broad and unspecific search term "Internet of Things". Because she wants to get a broad overview of the topic she uses the advanced search function to search for "Internet of Things" not only in titles but also within abstracts and full texts. She also wants to see what related concepts often appear together with the search term "Internet of Things". She discovers the adaptive training support (ATS) widget on the right side of the Search view (see section 2.4). It introduces her to the search visualisation features (see section 2.4. Before she enables the search results visualisation she familiarises herself with the feature by watching the linked tutorial video. This video is part of the digital competence and information literacy curriculum of the MOVING platform (see section 2.4). She learns that she can use the uRank visualisation to see the top concepts related to the query. When using uRank she discovers that the top concepts often appearing together with the search term "Internet of Things" are "smart cities", "smart homes", "big data" and "automation". As a Science and Technology Studies researcher she is particularly interested in "smart cities" and "big data", so she uses uRank to re-rank the search results list by defining a new order of concepts according to her interests.

Ms Brown is now interested in recent debates on the topic. To find most recent literature she limits the time range of displayed search results to the last five years. Because researchers often use social media or blogs to disseminate and discuss ongoing research projects long before their findings get published in academic journals she decides to limit the search to social media and websites using the databases filter. She can now browse the results list and check individual entries to get an overview of current research projects and prominent researchers in the field.

She knows, however, that for a scientific literature review she needs to concentrate on published documents, so she sets back the Databases filter and filters the query instead for written documents like books and journal articles and the central authors using the documents filter. She is still dissatisfied with the search results list because there are more results than she can manage. Since Ms Brown has little previous knowledge about the topic, she has difficulties narrowing down the search on their own. The ATS widget helps her to successfully use the topic-based filtering and the faceted search to limit the search to a manageable results list. She is limiting the displayed results to the subject areas “Science & Engineering”, “Computer Science & Technology” and “Architecture & Urban Planning”. As suggested by the recommender system within the ATS, she again uses a visualisation to see the connections between the search results. The graph visualisation helps her discover the structure of her research topic and limit the search to parts she is interested in. This visualisation helps her identify central authors on the topic and browse other publications by the same authors that might also be interesting to her. She can now start working systematically on her literature review by saving and bookmarking individual entries from the results list.

Scenario 2: Finding partners for research projects that are active in a specific research field

Due to the advanced information searching and analysing features of the MOVING search described above Ms Brown has gained an oversight over current debates and latest trends in the research field. Based on the literature review she has generated a number of new ideas and wants to develop them into a research project. She wants to find potential partners that want to work with her on the topic. The MOVING community hosts a wide range of researchers from different disciplines and with a variety of research subjects that might be interested in joining her on an interdisciplinary research project. She searches the community for members that work in the fields of Science and Technology Studies like herself but also for those in “Architecture and Urban Planning, Geography, Social Sciences and Humanities, Science and Engineering and Computer Science and Technology. She wants to find researchers that work on topics related to “Internet of Things”, “Smart Cities” and “Big Data”. The community search provides her with a list of 12 scholars that match her query. She browses the profiles and finds that 5 of them would be suitable partners, so she sends them a contact request which 4 of them accept. After watching an introductory tutorial video about the functions and benefits of the MOVING projects tool Ms Brown creates a group for her venture in the *Projects* environment of the MOVING platform and invites the 4 scholars to this research group. From her previous query on the research topic “Internet of Things”, she identified five central authors who have published extensively on this topic. She finds their contact details through their academic institution and sends them an email with an invitation to join MOVING and the newly created

research group that she started. One of the authors is interested and joins the project. The large collection of MOVING project modules – project wiki, task management, time tracking, document management, blogs, file storage, polls and message boards – makes it easy for the group to develop their initial ideas into a comprehensive research proposal.

Scenario 3: Find suitable funding opportunities for research

Ms Brown and her new colleagues have developed their ideas into a research proposal. Now they want to find suitable sources for financing their project. Ms Brown changes the search mode within the search bar from "research" to "funding" and starts the query with the search term "Internet of Things". The system now displays a list of results with a various call for tenders and call for proposals. Since Ms Brown only searches for funding at a national level she chooses "national" from the drop-down menu "select geographical area" within the faceted search. She narrows the search to "United Kingdom" by clicking on it in the drop-down menu "select nation". The other drop-down menus like "search grant by beneficiary" or "search grant by sectors" are not interesting for her at that moment. Ms Brown does not want to confine herself to one sector, as she is seeking for funds for an interdisciplinary project. She starts the query again. Now she sees only funding opportunities for projects funded by research programs of the United Kingdom for the search term "Internet of Things". Ms Brown can sort the search results in ascending or descending order to see which deadline is the closest. She finds a call, which fits her project idea. She copies and inserts the link to the research group that she has started to discuss with her partners whether they will hand in a proposal.

Scenario 4: Find additional learning materials and training resources

Ms Brown wants to find learning material to prepare for the interdisciplinary research project that she started with a group of scholars on MOVING. She switches the search bar to the Learning domain in the MOVING Search environment. She uses her previous search terms "Internet of Things", "big data" and "smart cities" again to search for learning resources. A list of different learning opportunities appears. She filters the results for her research fields *Science & Engineering, Computer Science & Technology and Social Sciences & Humanities* – and for the type of learning resource *Online Course*. The results contain online courses, MOOCs and Webinars from major universities and research institutions. She browses the list and decides to participate in a MOOC called "The Internet of Things builds smart cities of the future" organised at MIT.

Ms Brown wants to learn more about the advanced features of the MOVING platform and how she can use them in her research routine. Therefore, she goes to the Learning environment of the platform and clicks on "Take a tour through the platform". Here, she can find a variety of small learning units about different functionalities of MOVING. She wants to learn more about the collaboration features that the platform provides because she wants to teach a seminar on the topic "The Internet of Things – how big data is reshaping our cities" for students at her home university and plans on using MOVING as a digital collaboration and communication tool for her seminar. There are a number of short "how-to"-tutorial videos about the different modules of the projects

environments of MOVING and a wiki where all the functions are explained step-by-step. She creates a Project and invites her students to join the platform. The learning units she just found serves as a basic assignment for participants of her seminar so that her students learn how to collaborate in digital environments using the seminar wiki and forum, load documents and files to the repository, assign and schedule tasks to organise their class work.

4. User studies

This section provides a series of possible tasks corresponding to the scenarios from use case 1 and use case 2 that a user of the MOVING platform might perform during future user studies. These studies will evaluate the usability of the current release of the platform and help the further technical improvement of future iterations. Their purpose is to test the platform's functionalities and to show how well the platform prototype fits the user requirements developed in D1.1. Tasks 1.2 and 1.3 of the DoA are ongoing tasks regarding the implementation of requirements, platform testing and development. The participants for these planned user studies will be financial professionals (use case 1) and young researchers (use case 2). The different contents of the platform, such as different data sources and crawling of specific websites, will match the particular needs of both user groups. This will, therefore, create differences in the methodology and tasks of the studies. The studies will be conducted by UMAN and the details of the execution and the results of the user studies will be described in deliverable D1.3 (month 24). The use case partners (EY and TUD) will provide the partners at UMAN with successive task descriptions comparable to those described below.

4.1. Use case 1: Research on business information by public administrators

For a better evaluation of the user interactions during the user studies, the tasks that we propose here aim at the target user group of use case 1, financial professionals and public administrators. The task goals fall into the typical user scenarios of the use case and approach assignments that will enable the user to apply when on the job. The structure of the tasks follows a logical path from search to visualisations in order to provide the proper instruments to enable decision making on a reasonable basis.

User studies Task 1.1

Risk assessment procedures need to be performed by organisations, research institutes, universities and companies likewise to ensure compliance with current laws and regulations. The identification of possible risks can be crucial to the success of the entity or planned projects. An example is a search for administrative procedures required when participating in EU-funded programs. Moreover, extending businesses abroad also requires such risk assessments. A specific illustrative task description could be: S*earch for EU-funded programs; refer to most relevant search results; closely analyse the results for changes relevant to the job task.*

User studies Task 1.2

The use case 1 MOVING platform user needs to constantly update his knowledge and train his risk assessment procedures skills to ensure compliance with laws and regulations. The platform will provide the appropriate tools and will enable the user to efficiently and effectively achieve the goals. In order to test the user's capabilities of using the tools, the following task could be given here: *perform a SWOT analysis, including the necessary PEST analysis.* This task is rather general but it can

be extended and broken down into steps that also include the use of other features and functionalities of the MOVING platform.

User studies Task 1.3

For the analyses referred to in Task 1.2, the platform user should most probably be able to use the filtering features of the platform. The discipline filter feature should allow the user to choose one or multiple options such as political, economic social and technological, which are important for the PEST analysis. The task to be considered here could be formulated as following: *filter your results to perform a PEST analysis; select one or multiple disciplines to filter results; select search results relevant to each step of the PEST analysis.*

User studies Task 1.4

Further filtering features available on the faceted search will help the user to drill down the results and select only those relevant to the job. Language, country, date range, type of publication or industry are filters that can be applied for that purpose. The task that could be given in this case might be formulated as follows: *apply more filters to find more relevant content.*

User studies Task 1.5

In connection with User studies Task 1.1, the current task refers to the visualisations of the search output. Given that the changes in administrative procedures are off key in the risk assessment that the compliance officer needs to conduct, the visualisation over time enables a user-friendly overview and helps keep track of documents of interest. To be able to perform risk-based prognosis, the task to be given here could require the following: *track the evolutions of the documents over time and refer to specific changes of most relevant documents; explore the evolutions over time to determine critical issues to the business.*

User studies Task 1.6

The task formulation "*use the concept graph to further understand the content and relevance of the results to one another*" could be used to ask the user to determine the importance of individual topics and the connection between them. In a different scenario, the user might look specifically for the availability risk of certain materials. Their centrality on the concept graph can indicate a higher risk or not and the user may be able to determine whether sourcing of the materials presents an issue.

4.2. Use case 2: Managing and mining research information

The following examples illustrate tasks corresponding to use case 2 scenario 1 (state-of-the-art on a research topic)[11]: The user is a junior researcher in the field of social sciences, humanities, economics, science and technology studies, computer science, or engineering. She or he is interested in the emerging interdisciplinary research field of the "Internet of Things" and wants to get an overview of existing research and publications on the topic. For this purpose, she or he uses the data and text-mining features of the MOVING platform. Based on the use case scenario, the individual tasks define what the user should do. The following actions function as success indicators and will help the user evaluation of the search and visualisation functionalities of the MOVING platform. The sections below illustrate how possible tasks could be formulated. A task is considered to have been completed successfully if a number of predefined actions have been taken by the users.

User studies Task 2.1

"*Get an overview over the state of the art on the topic "Internet of Things" using the MOVING platform, find related concepts, reduce the results to specific subject areas and use the "concept graph" to visualise relationships between documents, authors and publication years"*.

Successful completion of the task is indicated by the following actions:

- user goes to "search" environment on the platform, type keyword "internet of things" in the basic search bar, clicks search and gets results list; uses the advanced search to expand the query "Internet of Things" to all search fields (title, abstract, full text). Reduces the results to subject areas "Science & Engineering", "Computer Science & Technology", "Architecture & Urban Planning" and "Social Sciences & Humanities" by using the filter "subject area";
- identifies central authors on the topic using the "persons" filter; Uses concept graph visualisation to see which authors have the most publications on the topic "Internet of Things";
- goes to "concept graph" visualisation and identifies bigger nodes, drag-and-drops nodes that appear interesting for the topic aside makes clusters and explores connections, clicks on nodes to see authors and documents connected to them;
- clicks on authors name to open new query to see what else this author has written about, clicks and opens documents that seem interesting to the topic;

User studies Task 2.2

[11] For the not yet implemented functions like the partner search (scenario 2) and funding search (scenario 3) as well as the search for training material (scenario 4) tasks are going to be developed when the functions are ready to use.

"Find new concepts closely related to the topic "Internet of Things" using visualisation features and include related concepts in the initial query"

Successful completion of the task is indicated by the following actions:

- finds related concepts like "smart cities", "smart homes", "big data" or "automation" using either "tag cloud" or "top concepts" to see which keywords are often occurring together;
- uses uRank to see the top concepts of the query and re-rank the search results list according to the new related concepts;
- combines new related keywords ("smart cities", "big data") with initial query "Internet of Things" using basic search: uses Boolean operators to search for "Internet of Things" AND "smart cities" OR "big data" in titles;
- uses the advanced search to search for "Internet of Things" in titles and for "Internet of Things" AND "smart cities" OR "big data" in abstracts and full text;

User studies Task 2.3

"Get an overview of recent debates on the topic "Internet of Things" and identify central researchers on the topic"

Successful completion of the task is indicated by the following actions:

- finds most recent literature by reducing the range to last five years 2013-2018 using "range" filter to retrieve only most recent hits or sorts search results list by date;
- filters the results list for social media and websites to get most recent debates on the topic; "use the filter "data source" to show only websites and social media content or uses the filter "document type" to show only websites and social media entries (Twitter/Google+);
- uses the "persons" filter to see what author has the most results in the query;

5. Conclusion

The goal of this deliverable is to present the initial implementation of requirements and functionalities of the platform. For that reason, the requirements list resulted from deliverable D1.1 was followed and discussed extensively with the partners responsible for implementation. As a management system for this collaborative work we used Trello, to ease the work together and improve transparency and status of the discussions. Furthermore, the requirements were assigned internal prioritisation to structure the work on implementation. The document started with a clarification of the relation between the MOVING use cases, user studies, and the MOVING platform and systems. That is, at the software implementation level, we developed a single, widely-applicable, integrated software platform, the MOVING platform, while at a content level, the customisations of the platform centre around the different content it needed to be populated with, specific to each use case, giving rise to two customised editions of the platform that we call “systems”. In the next sections of the document, along with the currents status of implementation, we included outlooks on possible improvements and changes that may occur in the future of the platform. Further on, we approached the navigation of the platform by the target users of both use cases describing how the users should typically use the platform and its functionalities. Finally, we suggest, for each use case, tasks that may be included in future user studies. This deliverable builds on the results of the deliverable D1.1 and will be update by deliverable D1.4.

References

Firesmith, D. (2007). Common Requirements Problems, Their Negative Consequences, and the Industry Best Practices to Help Solve Them. Journal of Object Technology, 6(1), 17–33.

Nishioka, C., & Scherp, A. (2016). Profiling vs. time vs. content: What does matter for top-k publication recommendation based on Twitter profiles? 2016 IEEE/ACM Joint Conference.

Tschinkel, G., Hafner, R., Hasitschka, P., Sabol, V. (2016) „Using Micro-Visualisations to Support Faceted Filtering of Recommender Results". IV 2016: 318-323

di Sciascio, C., Sabol, V., Veas, E. (2016) "Rank As You Go: User-Driven Exploration of Search Results". IUI 2016: 118-129.

Rauch, M., Klieber, W., Wozelka, R., Singh, S., Sabol, V. (2015) „Knowminer Search - A Multi-visualisation Collaborative Approach to Search Result Analysis". IV 2015: 379-385