IEEE Multimedia Magazine, vol. 25, no. 3, pp. 8-21, July-Sept. 2018. © 2018 IEEE. DOI:10.1109/MMUL.2018.2873495.
Author's accepted version. The final publication is available at https://ieeexplore.ieee.org/document/8494800/

# Open Innovation in the Big Data Era with the MOVING Platform

## An Integrated Working and Training Approach for Data-savvy Information Professionals

Iacopo Vagliano[(a)], Franziska Günther[(b)], Matthias Heinz[(b)], Aitor Apaolaza[(c)], Irina Bienia[(d)], Gert Breitfuss[(e)], Till Blume[(a)], Chrysa Collyda[(f)], Angela Fessl[(e)], Sebastian Gottfried[(b)], Peter Hasitschka[(e)], Jasmin Kellermann[(b)], Thomas Koehler[(b)], Annalouise Maas[(d)], Vasileios Mezaris[(f)], Ahmed Saleh[(a)], Andrzej Skulimowski[(g)], Stefan Thalmann[(h)], Markel Vigo[(i)], Alfred Wertner[(e)], Michael Wiese[(d)], Ansgar Scherp[(a)]

(a) ZBW – Leibniz Information Centre for Economics, Germany

(b) Technische Universität Dresden, Germany

(c) University of Manchester, United Kingdom

(d) Ernst & Young, Germany

(e) Know Center, Austria

(f) Centre for Research and Technology Hellas, Greece

(g) AGH University of Science and Technology; Progress & Business Foundation, Poland

(h) Graz University of Technology, Austria

(i) University of Manchester, United Kingdom

**Abstract:** In the Big Data era, people can access vast amounts of information, but often lack the time, strategies and tools to efficiently extract the necessary knowledge from it. Research and innovation staff needs to effectively obtain an overview of publications, patents, funding opportunities, etc., to derive an innovation strategy. The MOVING platform enables its users to improve their information literacy by training how to exploit data mining methods in their daily research tasks. Through a novel integrated working and training environment, the platform supports the education of data-savvy information professionals and enables them to deal with the challenges of Big Data and open innovation.

Open innovation is a distributed innovation process based on knowledge flows across organizational boundaries[1] which involves various actors, from researchers, to entrepreneurs, to users, to governments, and civil society. Many Open Innovation Systems (OIS) already exist, e.g., Innocentive (Innocentive.com) and Hypios (hypios-ci.com). They mainly support collaborative idea generation and problem solving. However, the generation of ideas is not the biggest challenge of open innovation. Research and innovation staff in academia and industry needs to effectively obtain an overview of publications, patents, products, funding opportunities, etc., to derive appropriate innovation strategies. For instance, researchers and students need to find, understand, and build on top of a large and steadily increasing number of previous publications and other online educational resources (video lectures, tutorials, etc.). Similarly, financial auditors need to monitor a constantly evolving set of regulations pertinent to their daily work. In the Big Data era, such information is usually available and freely accessible in digital resources (text and media). However, students and professionals typically lack the time, strategies and tools to efficiently extract useful knowledge from all these resources.

The MOVING project (moving-project.eu) is developing a platform to enable people from all societal sectors (companies, universities, public administration) to improve their information literacy by training how to exploit data and text mining methods in their daily research tasks. Thus, the MOVING platform's users can more efficiency identify and process relevant information by knowing how to deal with data and text mining methods, and then use this information to contribute to open innovation, as any innovation is based on previous knowledge.

The MOVING platform is a combination of data and tools. In terms of data, it integrates various kinds of educational resources: unstructured data in the form of documents, structured data in the form of metadata, as well as video material and social media data; some of these resources are automatically selected and collected from the Web and social networks. In terms of tools, the platform supports actions such as cross-media search on these resources, and exploits video processing techniques that enable automatic concept annotation and search on the videos. Through an integrated training and working environment, the MOVING platform provides two main contributions:

1. The working environment provides tools for the analysis of large amounts of structured, semi-structured, and unstructured data, notably text and other media. Two aspects of Big Data are addressed, volume and variety.
2. The training environment supplies a training program to use these tools and boost open innovation processes.

Finally, the combination of the working and training environment with a community of practice (currently being formed) will allow users to share ideas and challenges, as well as communicate their experiences and learn from them.

## MOVING BEYOND THE STATE OF THE ART IN OPEN INNOVATION SYSTEMS AND TECHNOLOGY ENHANCED LEARNING

MOVING is an interdisciplinary project bringing together unique expertise from computer science and media didactics. We conducted an extensive literature research and identified different fields of research related to OIS, based on existing classifications like Hrastinski et al.[2] These include OIS, Expert Search Systems (ESS), Recommender Systems (RS), Adaptive Hypermedia Systems (AHS), Decision Support Systems (DSS) and Technology-Enhanced Learning (TEL). We briefly discuss each of these related fields of research and show how they relate to MOVING.

OIS are concerned with the facilitation of open innovation processes and the transfer of knowledge from the crowd into organisations.[1, 2] Typically, an organisation describes a problem to be solved and provides a tool that allows individuals to submit proposed solutions. Hrastinski et al. identified typical OIS features: idea submission (users submit an idea, often within predefined categories), problem submission (organisations submit a problem, users suggest solutions), proposal evaluation (users assess the quality of proposed solutions), expert directory (describing and locating experts), and marketplace (connecting innovators with innovation seekers).[2] In contrast, MOVING addresses the question of how managers, researchers and employees can be trained to initiate, maintain and support open innovation. Furthermore, the existing OIS mainly support collaborative idea generation. However, the time, strategies and tools to efficiently extract the necessary knowledge from existing, background information is usually missing. MOVING addresses this challenge by providing tools for analysing large amounts of text and media (working environment – Sections *Data Acquisition* to *Data Visualization*) and training programs for these tools (training environment, notably the Adaptive Training Support – Section *Adaptive Training Support*).

Related to the expert directory of OIS, ESS identify people with relevant expertise on a topic of interest. Balog et al. reviewed this area.[3] Typically, ESS create profiles of candidate experts by associating a set of documents to them, to represent their expertise. In the context of open innovation, ESS are used by companies to search in the profiles and invite experts to submit solutions. The companies can then select the best contribution and acquire the rights to use it. Thus, ESS can boost a company's problem-solving activity.[2] However, finding experts in a given topic is often not sufficient since more diverse solutions in terms of domains of knowledge and perspectives on the problem are needed. Innovation often comes from experts in topics not directly related to the problem who can transfer the knowledge from one domain to anoth-

er. In MOVING, advanced search and visualisation functions such as network graphs (Sections *Data Indexing and Search* and *Data Visualization*) enable users to find key literature and experts on a topic and potentially related topics.

RS are other related tools which suggest interesting items, e.g., movies, news, scientific papers. Typically, RS are classified into content-based, collaborative-filtering, knowledge-based, or hybrid.[4] Content-based RS make suggestions that take into account the items a user liked in the past. Collaborative-filtering RS generate recommendations to a user based on the items that similar users liked. Knowledge-based RS infer similarities between user requirements and item features described in a knowledge base. Hybrid RS combine one or more of these techniques. With the evolution of the Web toward a global data space known as the Linked Open Data cloud (linkeddata.org/), Linked-Data-based RS have emerged. They suggest items by exploiting knowledge on the LOD cloud.[5] In MOVING, recommendations go beyond suggesting items to the user: the Adaptive Training Support (ATS), described in Section *Adaptive Training Support*, recommends platform features based on the users' behaviour.

In line with providing recommendations and personalized information, AHS aims to automatically adapt the organisation, presentation and interaction of personalized hypermedia content to its users.[6] To this end, AHS observe the users' interactions with the system and react to it. They maintain three interconnected models: diagnosis, educational and expert. The diagnosis model comprises assumptions and information about the level of knowledge of the user in a specific domain. The educational model provides a didactic concept of how to convey and present the content to users. The expert model contains relevant domain-specific knowledge. In MOVING, the ATS is based on the concepts underlying to AHS. It gives feedback regarding the user's context and activities on the MOVING platform (stored in the corresponding user profile). Based on this, it provides reflective questions to users, increasing their awareness of how they use the platform. Furthermore, it recommends new features to improve their search behaviour and train them to use the platform more effectively.

DSS are another field of research related to MOVING, specifically to the ATS. Their goal is to provide decisional advice to enable faster, better, and easier decision-making.[7] Central to open innovation systems and open educational systems, and thus to MOVING, are the two dimensions of *invocation* and *timing*. Invocation refers to how guidance is invoked,[7] i.e., whether users are automatically notified by the system based on predefined events, receive feedback only when users actively request it, or based on some context. In MOVING, the ATS considers all three previously mentioned forms of guidance. We analyse the users' behaviour on the MOVING platform to automatically provide reflective questions, which are informed by the users' behaviour. For example, if one often uses a specific feature then the ATS assumes that one likes such feature and asks why, in contrast, when a feature is not used, its use is suggested. Finally, users can also actively request guidance. T*iming* refers to when guidance is invoked.[7] Guidance can be triggered during the actual user activity, before a user actually conducts an activity, and after a user performed an activity. MOVING focuses on triggering training support during and after user activities.

Finally, TEL[8] is highly relevant to MOVING since using technological tools to support learning and knowledge acquisition is the central feature of our training environment. From TEL, MOVING borrows computer-supported reflective learning, i.e., the mechanism to learn from experience.[9] Reflective learning happens both directly within a work process (*reflection-in-action*) and more systematically outside operative work processes (*reflection-before-action*, *reflection-on-action*).[10] In the social context of an organisation, reflective learning must be understood not only as a cognitive process of the individual worker (individual reflective learning) but also as a social process (collaborative learning). Regarding computer-support for work-related reflective learning, activity logging supports reflection by providing accurate data. A transfer of these results to work settings is often not easy to implement for multiple reasons. First, it is often not obvious what data that constitutes relevant aspects of work can be automatically captured. Second, these data needs to be closely related to relevant entities in the work domain (e.g., customers or artefacts). Finally, even the best-educated users have difficulties in gaining actionable knowledge out of those data. Our key insight from the previous work is that reflection guidance needs to be designed into computer-mediated reflection tools, and embedding reflective learning into business processes is crucial.

Table 1 summarizes the discussion of the research fields presented above and their comparison with the MOVING platform, which introduces some new features. We identify the most relevant features in the related works previously mentioned[2-7, 9, 10] and classify them into the three key areas of MOVING: working environment, training environment, and community of practice. OIS cover the community of practice, while they lack all the features of the training environment. Only the OIS feature of expert directory is addressed regarding the working environment. This is also the only feature supported by ESS, although they may also profile their users to personalize the search. RS are limited to content recommendation and profiling. AHS focus on training, but do not support the community of practice and typically do not have all the working environment's features. DSS mainly recommend content and features. They also profile users and may provide visualizations. TEL focuses on the training environment, but does not address the working environment. To the best of our knowledge, while the research fields discussed focus on a specific area or a set of few features, MOVING supports all of them. However, the community of practice in MOVING is currently being formed and only partially available. The platform's key feature is integrating the working environment with the training: this allows users to improve their information literacy by training how to exploit data mining methods in their daily research tasks.

Table 1. Comparison of the different fields of research with MOVING. Legend: not supported: -, partially supported: (X), fully supported: X.

| **Field of research** / **Features** | OIS | ESS | RS | AHS | DSS | TEL | **MOVING** |
|---|---|---|---|---|---|---|---|
| **Working environment** | | | | | | | |
| Semantic search | - | - | - | X | - | - | X |
| Faceted search | - | - | - | - | - | - | X |
| Multimodal search | - | - | - | - | - | - | X |
| Expert directory[2, 3] | X | X | - | - | - | - | X |
| Advanced visualisation[6] | - | - | - | X | (X) | - | X |
| Evolution-aware search | - | - | - | - | - | - | X |
| Content recommendation[4] | - | - | X | X | X | - | X |
| **Training environment** | | | | | | | |
| Guidance with tutorials[7] | - | - | - | (X) | (X) | - | X |
| Videos[6] | - | - | - | (X) | (X) | - | X |
| Profiling[4, 6] | - | (X) | X | X | X | X | X |
| Reflective learning[9, 10] | - | - | - | X | - | X | X |
| **Community of practice** | | | | | | | |
| Idea submission[2] | X | - | - | - | - | - | (X) |
| Problem submission[2] | X | - | - | - | - | - | (X) |
| Proposal evaluation[2] | X | - | - | - | - | - | (X) |
| Marketplace[2] | X | - | - | - | - | - | (X) |

# PLATFORM OVERVIEW

The MOVING platform exhibits two main novelties. First, it integrates heterogeneous components and technologies for data analysis, visualization, search, etc., in a combination that cannot be found in other platforms, as shown in Table 1. These components are described in Sections *Data Acquisition* to *Data Visualization*. Second, it combines working and training in a single environment by merging analytical tools and visualisation techniques with a qualification and training concept. This is implemented by the ATS (Section *Adaptive Training Support*).

Figure 1 illustrates the MOVING platform's architecture. Data (HTML pages, tweets, scientific papers, videos, and more) are acquired from the intended sources and processed to be used by the search engine. Data processing includes features such as text and video analysis and author disambiguation. Social community functionalities include core features for the community of practice, e.g., blog, wiki, and the possibility to contact other users. On top of these modules the visualization technologies display the search results in different ways. User logging tracks the users' behaviour on the platform by capturing UI events from data visualization, while the ATS analyses the collected user behaviour data through the analysis framework (WevQuery) to select appropriate training material depending on the use patterns.

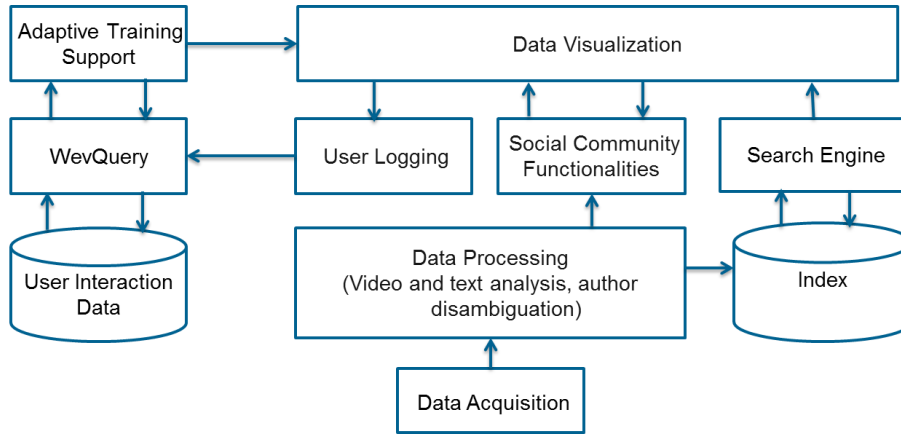



Figure 1: Architecture of the MOVING Platform

## Data Acquisition

The MOVING platform processes huge amounts of text coming from different sources. These datasets contain different document types, e.g., bibliographic data from scientific publications, crawled web pages, and video metadata. To integrate data from these sources, we use different crawlers.

To crawl social media, we adapted the Social Stream Manager (github.com/MKLab-ITI/mklab-stream-manager) to the MOVING architecture. This crawler monitors several social streams (Twitter, Google+ and Youtube) to collect incoming content relevant to a keyword, a social media user or a location, using the corresponding API of each service. It stores the items (tweets, posts, etc.), extracts webpage and multimedia links, and indexes them. To retrieve data from the Web, the platform needs to perform topic-based search and also crawl websites. We exploit web search APIs to search for particular topics in the Web, including the Google custom search, Faroo and Bing Search. Crawling specific websites requires a crawler limited to a specified web domain. A crawler based on Scrapy (scrapy.org) provides this functionality. The first prototype of the MOVING platform contains 19,638 crawled web pages.

Moreover, we harvested bibliographic metadata from the Linked Open Data cloud. We relied on an adaptive index model to cope with heterogeneous knowledge representations in different data sources. The initial version of MOVING contains 181,235 metadata records from a snapshot of the Linked Open Data cloud previously crawled (km.aifb.kit.edu/projects/btc-2014).

Additionally, the first prototype of the MOVING platform contains the following datasets:

- **Videolectures.net**. It consists of 19,756 metadata records of educational video lectures with transcripts.
- **ZBW Economics.** It consists of 4.6 million metadata records and 413,097 full-text scientific publications in Economics.
- **GESIS.** It includes 2.8 million metadata records and 5,400 open access documents in social sciences.
- **Laws and regulations**. It collects 1,693 metadata records, each containing information about a law and its changes over time. It mostly covers sea-related regulations in Germany and the European Union.

We are continuously integrating further data to extend our databases.

## Video Processing

A key feature of the data processing module is video analysis. Two general classes of video that can contribute to the educational functionalities of the MOVING platform are considered: lecture videos and non-lecture videos.

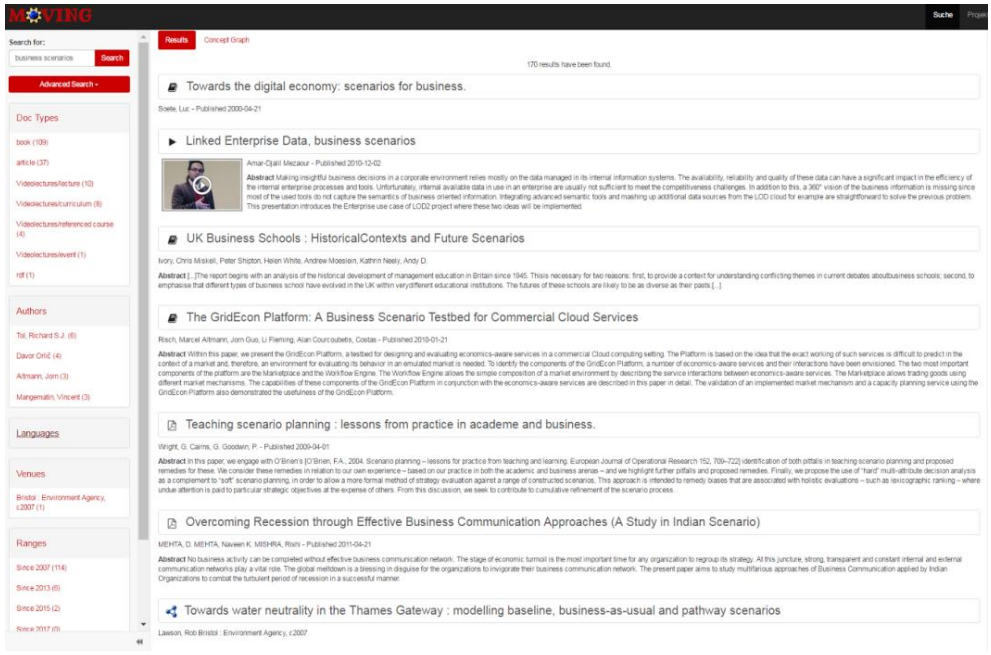
For lecture videos, audio is most important: just by listening, one can understand the topic of the lecture and infer a wealth of metadata for indexing. Thus, in the MOVING platform we ingest the audio transcript of each lecture video (from an off-the-shelve speech recognizer that exhibits a moderate word error rate of 23.4% on our data), and automatically translate it into a set of high-level concepts (from a rich, pre-specified concept pool). To this end, a Transcript Language Model (TLM) and Concept Language Models (CLM) are built, similarly to Tzelepis et al.[11] A TLM is a set of N keywords extracted from the transcript of a video: the transcript is transformed in a Bag-of-Words (BoW) representation and the N most frequent keywords are selected. Similarly to TLM, a CLM is a set of M words or phrases most relevant to a specific concept (a separate CLM is built for each member of our pre-specified concept pool). CLMs are built using Wikipedia: a Wikipedia query is automatically issued for a concept, and the top retrieved articles are transformed in a BoW representation, from which the top-M words are kept. After building the TLMs and CLMs, we calculate a single value per concept-transcript pair, denoting the semantic relation between the two. Specifically, for a CLM-TLM pair we initially form a $N \times M$ distance matrix $W=[w_{i,j}]$. Each element of this matrix captures the semantic relatedness between pairs of words appearing in the TLM and CLM, calculated using the Explicit Semantic Analysis measure.[12] $W$ is then transformed into a scalar value using the Hausdorff distance, defined as $D_H=\text{median}_i(\max_{(I;\ i=\text{const})}(w_{i,j}))$. For a given TLM, a single score is calculated per concept repeating the above process for the corresponding CLM, and the k concepts with the highest scores represents the video transcript.

For non-lecture videos, the visual modality is the most important: to convey its message, a documentary or a piece of user-generated video most often shows something, rather than orally describing it. Thus, instead of attempting transcript analysis of audio possibly captured in uncontrolled environments (e.g. outdoors), we analyse the visual modality. The video is decomposed into elementary temporal segments (shots) with the method of Apostolidis et al.[13] Then, each shot is annotated with high-level visual concepts coming from the same pre-specified concept pool used for describing the lecture videos. This pool comprises the 346 concepts defined in the TRECVID SIN task (as in Markatopoulou et al.[14]), but is easily extendible to additional concepts for which training data are available (e.g., ImageNet). We use state-of-the-art deep-learning techniques such as Deep Convolutional Neural Network (DCNN) architectures. We adapted GoogLeNet to our selected concept pool by fine-tuning strategies that involve not only the traditional replacement of the classification layer of the DCNN, but also the insertion of one or two additional extension layers.[15] The output of each fine-tuned network is one score in [0,1] for each of the 346 concepts. Fusing in terms of arithmetic mean the output of the different networks for a given concept results in a single score for each concept; and the k concepts with the highest scores represent the video in the concept space.

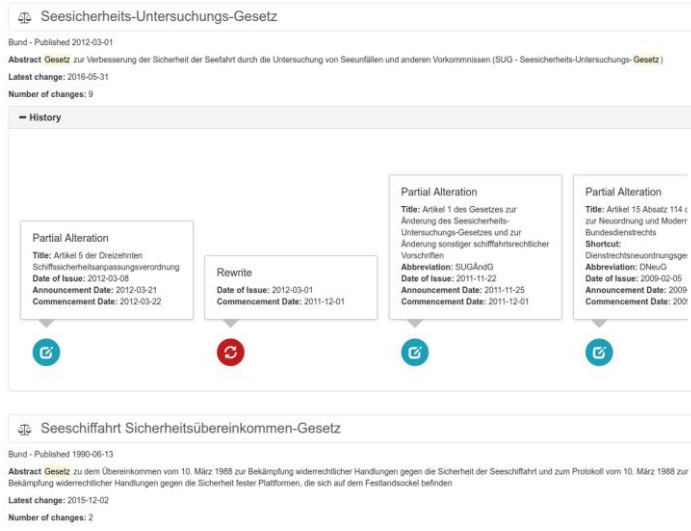
Finally, we match the generated concept-based representations of lecture and non-lecture videos using semantic word embeddings to support not only the concept-based retrieval of lecture and non-lecture videos, but also the detection of associations between such videos. This means finding which non-lecture videos are most closely related to a given lecture video. This is realized in a direct analogy to how Markatopoulou et al.[14] use semantic word embeddings to match the concept-based representations of textual queries and videos for performing video retrieval.

## Data Indexing and Search

To efficiently handle the various data acquired, the MOVING search engine provides scalable, real-time, multimodal and faceted search and handles multiple document types. The facets enable users to filter the search results based on different criteria (document type, author, date, venue, etc.). Figure 2a shows the search results page and the faceted search widgets. We developed a novel ranking method, HCF-IDF, which ranks the search results based on their relevance to the user query relying only on titles.[16] This is an important feature as often only document's titles are available. HCF-IDF can obtain results comparable to state-of-the-art techniques based on full text. Another interesting feature is viewing the history of related documents, e.g., laws and regulations of a specific topic (Figure 2b). This feature can help users like auditors to track the evolution of these documents over time and refer to a specific version.




(a)

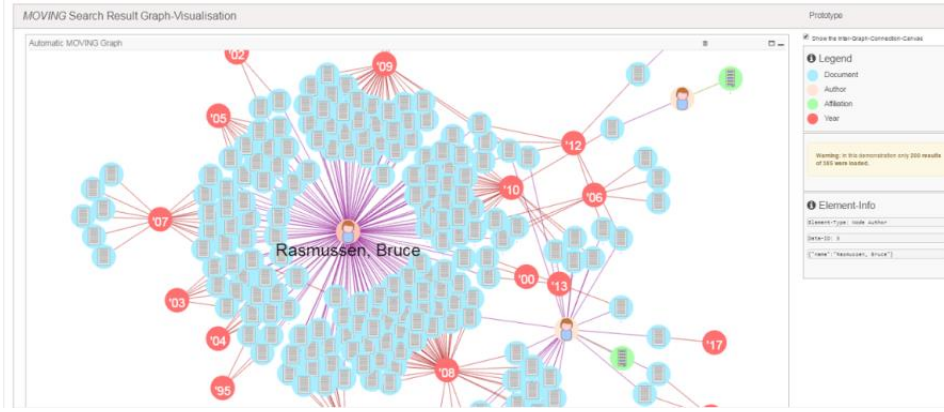



(b)

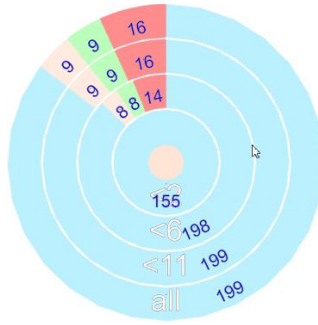
Figure 2: The search page of the MOVING platform (a) and a law history example (b)

## Data Visualization

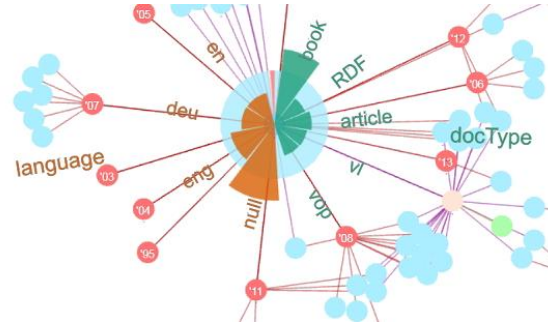
Browsing through a long list of documents and reading parts of their content to locate the needed information can be an exhausting task. Data visualization assists users to more easily find valuable information in search results. Graphs are used to represent different entities and their relations. Entities are visualized as nodes, which are connected by links. Each graph is represented by a specific visual layout, which specifies the positions of the nodes (e.g., through force-directed placement algorithms[17]) and the geometry of the links (e.g., edge bundling methods[18]). Different types of entities and relations can be visualized through different visual variables. In our case, entities such as documents, authors, and locations are represented through nodes of different colours containing a particular icon, while relations vary in colours and thickness.




(a) Graph visualization




(b) Summary of a node context




(c) Aggregated graph region

Figure 3: Rich visualizations provided by the MOVING platform

The Graph Visualization Framework (GVF, github.com/PeterHasitschka/gvf_core) is the primary module responsible for data visualization in MOVING. It supports interactive analysis of large, complex networks consisting of various entities and relationships which arise from co-occurrences, hierarchies, reading orders, etc. GVF focuses on visual representations of metadata and novel graph aggregation metaphors conveying relevant properties of nodes and relations in sub-graphs. We introduce powerful interaction models for explorative navigation, filtering and visual querying of graph data. The graphs to visualize contain multiple types of nodes connected by different types of links. Additionally, this graph may grow large and complex when many nodes are shown, leading to an information overload. GVF enables users to focus on the desired information by summarizing the rest of the graph in a way that allows them to identify and explore other potentially relevant graph areas. Figure 3a shows a graph visualization of the search results.

Users can initiate the exploration of the graph beginning from a selected node. Clicking on it, its context, the network surrounding it, is summarized (Figure 3b). This allows users to identify related nodes depending on their properties (as their type or other metadata) and their distance from the original node. Users can explore the rest of the graph by clicking on a sector, which triggers the expansion of the visible portion of the graph by showing nodes and relations which surround the current node. Additionally, users can aggregate graph regions which are out of their focus to view a less complex summary (Figure 3c), which provides information on what users can expect to find in that region, e.g., documents' kind and language.

## Adaptive Training Support

The ATS implements reflective learning technologies from the TEL domain to the domain of learning how to search. It supports users to learn how to search and complete the selected curriculum through the *learning-how-to-search* widget, reflects a user's search behaviour regarding the functionalities used on the platform, provides questions to reflect on the progress, and recommends documents or activities to do next.

Through the learning-how-to-search widget, users can become experts in searching with the MOVING platform. The widget provides a bar-chart representing the feature use (e.g., how often the graph visualisation was exploited) and reflective questions related to its use, or recommend a new feature. It also tracks users' experience in using the platform´s features, e.g., stores their search queries. Depending on this experience, prompts are selected and presented to users. Additionally, the user´s context is taken into account, which may be related to the user's search topics or curriculum.

The ATS consists of input, analysis, and output steps. The input step consists of context and user data from our recommender system and semantic profiling modules. The context information is the user's workplace activities, notably the MOVING platform's use during work. These activities are automatically tracked by the activity logging and analysis framework (WevQuery) whereby activity data is stored in the corresponding user profile. The analysis step exploits the data provided by the input step and prepares them for the visualization of the reflection guidance. The analysis step takes into account the context and user data in order to derive information with regard to the user´s behaviour and search patterns, including the functionalities of the MOVING platform used. It computes data for being presented as visual performance indicators and guidance in the form of questions for reflective learning. The output step provides support for learning how to use the platform and opportunities on how to improve the search behaviour by suggesting not used or unknown features. Similar approaches have been successfully deployed elsewhere[19].

# EVALUATION

The MOVING platform is developed following a user-centered design paradigm, which focuses on the users' needs conducting evaluations at all stages of the development process. We employ a mixed-methods approach to acquire an understanding of the user's needs and feed it back into the platform.

At the requirements gathering stage, we interviewed 26 auditors about their day-to-day challenges with unstructured data. The interviews highlighted the need of tools to support the analysis of unstructured data, which typically includes contracts, laws, emails, and search engine results. The strategies employed to effectively handling this data included the extensive use of the within-document search functionality and critical screening. To address researchers, nine MSc and PhD students in the humanities and social sciences were interviewed about their strategies for dealing with unstructured information. The interviews' analysis revealed twelve themes that can be grouped into three main activities, namely strategies for search and information retrieval, knowledge generation and management, and collaboration and cooperation (moving-project.eu/wp-content/uploads/2017/04/moving_d1.1_v1.0.pdf). The strategies previously mentioned informed requirements gathering and evaluations of the platform in the following development iterations. For instance, based on the interviews, low-fidelity prototypes were created and iteratively refined by the stakeholders to map the requirements into functionalities and drive the development of the platform.

In addition to the formative evaluations described above, summative evaluations of the MOVING platform provide deeper insights into the effectiveness of our approach. The usability of the platform was evaluated in a laboratory setting where users performed a set of typical tasks that were derived from the formative evaluations. The 27 participants (20 PhD students and 7 auditors) to the evaluations carried out three tasks that were representative of their use case: Get and overview of a new research topic; Narrow down the exploration using advanced search functionalities and Use visualisations to find related topics. We took a mixed-methods approach to evaluate the usability. We used standard questionnaires, such as System Usability Scale (SUS), objective measures, such as completion times, and our own observations. Both user groups reported that the platform's usability is acceptable (overall SUS score of 71.39) with averages between 65.36 and 73.50 for auditors and PhD students respectively – see the distribution of scores in Figure 4. Nevertheless, we identified workarounds that were indicators of suboptimal functionalities. For instance, we observed that some participants used the browser's search functionality when a large number of results was yielded, which suggests that participants found more efficient ways to find what they were looking for. Future iterations will be devoted to generating fewer and more meaningful results.

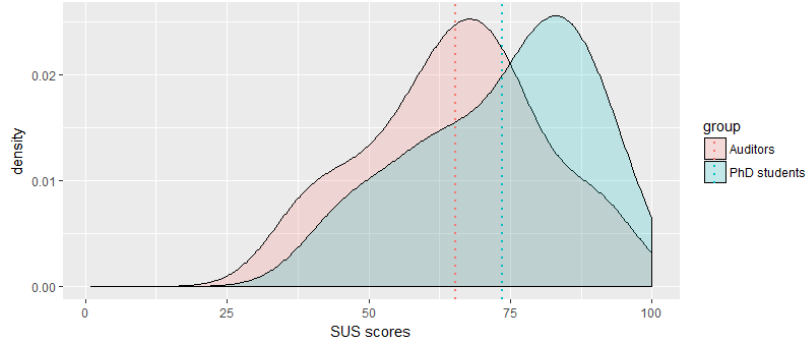



Figure 4: SUS scores per user group. Dotted lines indicate the group's average. Scores between 50–70 indicate usability is *OK*, above 70 is *good*.

Since the MOVING platform focuses on the training aspect, longitudinal evaluations will assess the learning progress made by the users of the platform. This will be leveraged by WevQuery,[20] the interaction analysis framework of the MOVING platform, which enables MOVING administrators to test complex hypotheses of use. Since WevQuery can handle large amounts of data, these hypotheses can be refined in an iterative fashion. For instance, one can query the number of users that in the last *6 months* entered the term *Deep Learning* in MOVING search engine and then clicked on the *first result*. While these queries can be built by the administrators as needed, there is a core set of queries already exploited by the ATS, which uses the output of WevQuery to profile the users of the platform and provide feedback accordingly. Further extensive user studies are planned, which will take place in the context of building the community of practice.

# CONCLUSIONS

The MOVING platform supports the analysis of large amounts of videos and text, and can train data-savvy information professionals to face the challenges of Big Data and open innovation. In the next months, the platform will be publicly available and tested with hundreds of students and researchers in social sciences from the Technical University of Dresden within respective research training programmes on both postgraduate and doctoral level (www.edu-tech.eu), and with the auditors of Ernst & Young in their real research tasks by supporting the different purposes and needs of these users at the same time. We will improve the platform based on such extensive evaluations. The community of practice is being build and used as pool for drawing participants for the evaluations and to enable sharing ideas and challenges on the platform.

# ACKNOWLEDGMENTS

This work was supported by the EU's Horizon 2020 programme under grant agreement H2020-693092 MOVING. The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG. The laws and regulations dataset used in the MOVING platform was provided by Wolters Kluwer within the H2020 ALIGNED project.

# REFERENCES

1. Chesbrough, H. Open Innovation: A New Paradigm for Understanding Industrial Innovation. In: Open Innovation: Researching a New Paradigm, Oxford University Press, 2006, pp. 1–12.
2. Hrastinski S., Kviselius N. Z., Ozan H., Edenius M. A Review of Technologies for Open Innovation: Characteristics and Future Trends. HICSS 2010: pp. 1-10
3. Balog K., Fang Y., de Rijke M., Serdyukov P. and Si L., "Expertise Retrieval", Foundations and Trends. Information Retrieval: Vol. 6: No. 2–3, 2012, pp 127-256.
4. Ricci F., Rokach L., Shapira B. Recommender Systems: Introduction and Challenges. Recommender Systems Handbook. Springer, Boston, MA, 2015
5. Figueroa, C., Vagliano, I., Rocha, O. R., and Morisio, M. A systematic literature review of Linked Data-based recommender systems. Concurrency Computat.: Pract. Exper., vol. 27, 2015, pp 4659–4684. doi: 10.1002/cpe.3449.
6. Scherp, A. Authoring of Multimedia Content: A Survey of 20 Years of Research, Semantic Multimedia Analysis and Processing, CRC Press, 2014.
7. Morana, S., Schacht, S., Scherp, A., Maedche, A. A review of the nature and effects of guidance design features. Decision Support Systems vol. 97, 2017, pp 31-42.
8. Goodyear, P. & Retalis, S. Technology-enhanced learning. Rotterdam: Sense Publishers, 2010.
9. Boud, D., Keogh, R., & Walker, D. Promoting reflection in learning: A model. Reflection: Turning experience into learning, 1985, pp 18-40.
10. Schön, D. A. The reflective practicioner: how professionals think in action, New York: Basic Books, 1983.
11. Tzelepis, C., Galanopoulos, D., Mezaris, V., Patras, I. Learning to detect video events from zero or very few video examples. Image and Vision Computing Journal vol 53, 2016, pp 35-44, Elsevier.
12. Gabrilovich, E., Markovitch, S. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. IJCAI 2007, pp 1606-1611.
13. Apostolidis, E., Mezaris, V. Fast Shot Segmentation Combining Global and Local Visual Descriptors. ICASSP, 2014. Software available at mklab.iti.gr/project/video-shot-segm.

14. Markatopoulou, F., Galanopoulos, D., Mezaris, V., Patras, I. Query and Keyframe Representations for Ad-hoc Video Search. ACM ICMR 2017, Bucharest, Romania.
15. Pittaras, N., Markatopoulou, F., Mezaris, V., Patras, I. Comparison of Fine-tuning and Extension Strategies for Deep Convolutional Neural Networks. MMM'17, Springer LNCS vol. 10132, 2017, pp. 102-114.
16. Nishioka, C., & Scherp, A. Profiling vs. time vs. content: What does matter for top-k publication recommendation based on Twitter profiles? JCDL, ACM, 2016.
17. Fruchterman, T. M. Graph drawing by force-directed placement. Software: Practice and experience, 1991.
18. Holten, D. a. Force-directed Edge Bundling for Graph Visualization. Berlin, Germany: The Eurographs Association; John Wiley; Sons, Ltd. 2009.
19. Malacria S., Scarr J., Cockburn A., Gutwin C., and Grossman T. Skillometers: reflective widgets that motivate and help users to improve performance. ACM symposium on User interface software and technology. ACM, 2013, pp 321-330.
20. Apaolaza, A. and Vigo, M. WevQuery: Testing Hypotheses about Web Interaction Patterns. Proc. ACM Hum.-Comput. Interact. 1, 1, Article 4 (June 2017), 17 pages. DOI: 10.1145/3095806

# ABOUT THE AUTHORS

**Iacopo Vagliano** is a postdoctoral researcher at ZBW – Leibniz Information Centre for Economics. His research interests include recommender systems, Semantic Web and Linked Data. Vagliano received his Ph.D. in Computer Engineering from Politecnico di Torino. Contact him at i.vagliano@zbw.eu.

**Franziska Günther** is a research associate at the Technische Universität Dresden, from which she obtained her MSc in communication science. Her research interests include computer-mediated communication and collaboration, and the impact of media coverage on public opinion. Contact her at franziska.guenther1@tu-dresden.de.

**Matthias Heinz** is a research assistant at the Technische Universität Dresden, from which he obtained his MSc in research in further education and organizational development. His research focuses on digital enhanced learning within higher education. Contact him at matthias.heinz@tu-dresden.de.

**Aitor Apaolaza** is a research associate at the University of Manchester, from which he obtained his Ph.D. in Computer Science. His research interests include Web behaviour analysis and systems to ease the access to complex interaction data. Contact him at aitor.apaolaza@manchester.ac.uk.

**Irina Bienia** is a data scientist in the Assurance Research and Development Group of Ernst & Young (EY). Her research interests focus on data analytics. She obtained her Ph.D. in bioinformatics at the University of Duisburg-Essen. Contact her at irina.bienia@de.ey.com.

**Gert Breitfuss** is a senior researcher at Know Center (KC). His research field is open innovation management with special focus on business model innovation. Gert received his MSc in business administration from Karl-Franzens University. Contact him at gbreitfuss@know-center.at.

**Till Blume** is a research assistant at ZBW, and a Ph.D. student in Computer Science Kiel University. His research interests include graph mining and Linked Data. Contact him at t.blume@zbw.eu.

**Chrysa Collyda** is a research associate at the Centre for Research and Technology Hellas (CERTH). Her research interests include image/video processing, human computer interaction, internet and multimedia, bioinformatics and medical informatics. She received her Ph.D. in Medical Informatics from the Aristotle University of Thessaloniki. Contact her at ckol@iti.gr.

**Angela Fessl** received her PhD in informatics from the Graz University of Technology, Austria. She is a postdoctoral researcher at KC and focuses on technology enhanced learning (TEL), specifically reflective learning at the work place. Contact her at afessl@know-center.at.

**Sebastian Gottfried** is a research assistant at the German Aerospace Center in Dresden. He obtained his diploma in information science from Technische Universität Dresden.

**Peter Hasitschka** is a researcher at KC. He is interested in e-learning, social network analysis and visual analytics. He obtained his diploma from the Graz University of Technology.

**Jasmin Kellermann** is a research assistant at the Technische Universität Dresden. There, she is a Master student in European Languages. Contact her at jasmin.kellermann@tu-dresden.de.

**Thomas Köhler** is Professor at the Technische Universität Dresden. He holds a Ph.D. in communications psychology, is Director of the Media Centre, head of the E-Learning Task Force of the Saxonian Rectors conference and President of the International Society of Media in Science. Contact him at thomas.koehler@tu-dresden.de.

**Annalouise Maas** is a team leader in the Assurance Research and Development Group of EY. She manages the teams working on innovation. She obtained her MSc from the University of Duisburg-Essen. Contact her at annalouise.maas@de.ey.com.

**Vasileios Mezaris** is a Senior Researcher at CERTH. His research interests include image/video analysis and retrieval, and machine learning for multimedia and big data analytics. He received a Ph.D. in Electrical and Computer Engineering from the Aristotle University of Thessaloniki. He is a Senior Member of the IEEE. Contact him at bmezaris@iti.gr.

**Ahmed Saleh** is a research assistant at ZBW and a Ph.D. student in Computer Science at Kiel University. His research interests include machine learning, text and data mining. Contact him at a.saleh@zbw.eu.

**Andrzej Skulimowski** obtained his PhD in Automatic Control from the AGH University. He is Professor and Director of the Decision Science Laboratory, AGH University. He is the CEO of the Progress and Business Foundation. His expertise includes multicriteria decision analysis and decision support systems (DSS). He is a member of IEEE. Contact him at ams@agh.edu.pl.

**Stefan Thalmann** is assistant professor at the Graz University of Technology, senior researcher at KC and area manager in Pro2Future. He obtained his Ph.D. in Information Systems from the University of Innsbruck. He works on Knowledge Management, TEL and DSS. Contact him at stefan.thalmann@tugraz.at.

**Markel Vigo** is a Lecturer at the University of Manchester. Vigo is interested in making data-intensive systems easy to use through the analysis of users' behaviours. Vigo received a PhD in Computer Science from the University of the Basque Country. Contact him at markel.vigo@manchester.ac.uk.

**Alfred Wertner** is a senior developer at KC. His main interests include technologies for reflection and learning in working environments. Contact him at awertner@know-center.at.

**Michael Wiese** is Research and Development Leader for the Assurance service line of EY. He is interested in Innovation Management, Fraud Auditing, and Process Mining. He received his Ph.D. from the University of Duisburg-Essen. Contact him at michael.wiese@de.ey.com.

**Ansgar Scherp** is Professor at the Kiel University and ZBW. He received his Ph.D. in Computer Science at the University of Oldenburg, Germany. His research interests include data mining, Semantic Web and Linked Data. Contact him at a.scherp@zbw.eu.