Enhancing Social Media Platforms for Educational and Humanitarian Knowledge Sharing: Analytics, Privacy, Discovery, and Delivery Aspects

THÈSE N^{O} 7495 (2017)

PRÉSENTÉE LE 24 FÉVRIER 2017

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
GROUPE SCI STI DG
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Andrii VOZNIUK

acceptée sur proposition du jury:

Prof. J. Huang, président du jury
Dr D. Gillet, directeur de thèse
Dr C. Ullrich, rapporteur
Dr A. Mikroyannidis, rapporteur
Dr I. Vonèche Cardia, rapporteuse

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

Share your knowledge.
It’s a way to achieve immortality.
— apparently misattributed to Dalai Lama,
but still like the idea

To my family…
and all the great people I met along the way

Acknowledgments

Writing the Acknowledgments section is the time when one realizes how many people and factors have contributed to the five years PhD venture. In my case, I was lucky to be able to work in great places that EPFL and Switzerland are, with great colleagues from my dear React group, enjoyed interacting with amazing project partners, and more than that was supported by my beloved family and friends.

Foremost, I would like to take this opportunity to thank the thesis jury president Prof. Jeffrey Huang and the thesis jury members Dr. Isabelle Vonèche Cardia, Dr. Carsten Ullrich, and Dr. Alexander Mikroyannidis for taking their time to engage with my research.

I was really lucky to do my PhD studies in the React group of EPFL, where I enjoyed collaborating and interacting with amazing people. The atmosphere in the lab encouraged open collaboration and free exchange of ideas, and I benefited from that and learned a lot. I would like to thank my supervisor Denis Gillet. Denis, thank you for always encouraging me to try something new and being open to new ideas and risky experimentation and, what is the most important, being tolerant to failures as a part of the learning experience. Thank you for your continuous support in my research, valuable encouraging feedback and motivating discussions. Research is only one part of the things I aimed to learn from you, but there is much more to it! Thank you for being my mentor!

A special thank goes to the fellow PhD students from the group: Evgeny Bogdanov, Na Li, Laleh Makarem, Wissam Halimi, and Ezequiel Debada. I always enjoyed talking to you, discussing ideas both research and life-related, and doing sports together. Your empathy and ability to encourage when it was necessary helped me to keep going. I would like to thank Evgeny in particular for being my friend and at the same time a technology mentor, thank you for spending your time getting me up and running on the projects at the beginning of the PhD. It is always motivating to talk to you, share ideas and receive constructive feedback. Moreover, your passion for sports and things you do is a never-ending source of motivation and inspiration.

I was honored to conduct research, write papers with, and be advised by postdoctoral researchers in the group: Sten Govaerts, Adrian Holzer, and Maria Jesus Rodriguez-Triana. Thank you for showing by your example how the research should be done and for your super helpful suggestions regarding my papers and research lines in general. This communication pushed me to learn and improve and helped to become a better researcher.

I also learned a lot from the software engineering team of our group consisting of Alex Wild and Andre Nogueira. Guys, thank you for helping to build things faster and better!

During my studies, I benefited a lot from working with numerous talented students that did projects within our group. These collaborations pushed me to learn new things and challenged me to understand and help to find solutions to issues as they emerged. Aubry Cholleton, Nikita Grishin, David Sandoz, and Moe Haydar thanks, it was a great pleasure working with you and thank you for committing and pushing from your side!

One of the reasons why things run smoothly during my PhD, is because all of the administrative work was handled by Dominique Espic and Anne Remillet. Thank you for taking care of that and being helpful whenever I needed an administrative advice. Special thank goes to Christophe Salzmann for always being responsive and helping to resolve challenging technical issues encountered in the course of the PhD.

I would like to express my gratitude to numerous project partners I have collaborated with during my PhD. Thank you Jorge Mazuze for facilitating the connection with the MSF users and being open to trying new things. Your insightful feedback from the users (particularly from the field) and the suggested improvements were super useful. They always motivated to keep on improving our tools to better satisfy the needs of end-users. Numerous Go-Lab partners were as well indispensable in getting user feedback and brainstorming ideas to be researched. Thank you all!

From my family side, I feel very happy to be supported by my dear wife, Lina. Thank you for going together through this Swiss PhD adventure, always being on my side and inspiring me to aim higher and develop as a better human being. I am as well extraordinary grateful to my parents: Lesya and Slavik. Mama and tato, most of what I have achieved is because of the values you have conveyed to me via education and hard work together and just by showing it all by your example. Thank you! Sergii, my dear brother and a close friend, you are my companion in many ventures, and you are the one that I can always rely on. My greetings go to my extended Kyiv family. Kostya, Olya, Lyuda and Yarema thank you for your support while studying in Kyiv. I am super happy to meet you every time I am in Kyiv. I was as well encouraged and supported by part of my family from the USA. Volodya, Olya, Eugene, Yura and Anya thank you for being just great!

Communication with my Ukrainian friends always brought my mood up, and their visits made my life brighter. It was always valuable to discuss different perspectives on common topics. Oleg Gumennyi, Katja Voznytska, Oleg Zaezdny, Denis Oborskiy just thank you! I am as well grateful to my Swiss friends with whom we shared great work and great play together. Sergii Buth, Denise Marchi, Hamza Harkous, Walaa Wehbe thank you.

From my studies before joining EPFL, I am particularly grateful to Dr. Yuri Maistrenko, who introduced me to scientific research and showed how international science looks like. My sincere greetings go to Dr. Pavlo Maslyanko who was an indispensable mentor during my time at the Kyiv Polytechnic and always encouraged to improve and seek more.

My research was primarily funded by the European Union in the context of the Go-Lab project

(Grant Agreement no. 317601) under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (FP7). Also, part of the research was supported by the Innogrant and via the collaboration with Médecins Sans Frontières. I express my sincere gratitude to the funding organizations that made this research possible.

Even though I really tried hard to give thanks to everyone involved, it is still possible that I have missed someone. So finally, thank you all the great people that I met along the way!

Lausanne, 10 November 2016 Andrii Vozniuk

Abstract

Social media (SM) platforms have demonstrated their ability to facilitate knowledge sharing on the global scale. They are increasingly often employed in educational and humanitarian domains where, despite their general benefits, they expose challenges peculiar to these domains. Specifically, the research context of this thesis is directed by my participation in the Go-Lab European project and my collaboration with Médecins Sans Frontières (MSF) where SM platforms were used extensively.

In this thesis, we address four challenges regarding *analytics*, *privacy*, *discovery* and *delivery*, aiming to answer corresponding four research questions. How to provide user-oriented analytics in knowledge sharing systems to support awareness and reflection? What privacy management interfaces and mechanisms are suitable for knowledge analytics and learning analytics? How to enable discovery of knowledge relevant to user interests? How to facilitate knowledge delivery into settings where Internet connectivity is limited or absent? Henceforward, we provide an overview of our results.

Analytics. To enable awareness and reflection for an SM platform users, we propose the embedded contextual analytics model where the analytics is embedded into the interaction context and presents information relevant to that particular context. Also, we propose two general architectures materializing this model respectfully based on real-time analytical applications and a scalable analytic back-end. Using these architectures, we provided analytics services to the SM platform users. We conducted an evaluation with the users demonstrating that embedded contextual analytics was useful to support their awareness and reflection.

Privacy. To address the privacy concerns associated with the recording, storage, and analysis of user interaction traces, we propose a novel agent-based privacy management model. Our proposal uses a metaphor of physical presence of a tracking agent in an interaction context making the platform user aware of the tracking and allows to manage the tracking policy in a way similar to the physical world. We have implemented the proposed privacy interface in an SM platform and obtained positive evaluation results with the users.

Discovery. Due to a large number of content items stored in SM platforms, it can be challenging for the users to find relevant knowledge. Addressing this challenge, we propose an interactive recommender system based on user interests enabling discovery of relevant content and people. We have implemented the proposed recommender in an SM platform and conducted two evaluations with platform users. The evaluations demonstrated the ability of the approach to identify relevant user interests and to recommend relevant content.

Delivery. At the moment of writing in 2016, near half of the world's population still does not have reliable Internet access. Often, the places where humanitarian action is needed have limited Internet connection. We propose a novel knowledge delivery model that relies on a peer-to-peer middleware and uses low-cost computers for local knowledge replication. We have developed a system implementing the model and evaluated it during eight deployments in MSF missions. The evaluation demonstrated its knowledge delivery abilities and its usefulness for the field staff.

Keywords: social media, knowledge sharing, knowledge analytics, learning analytics, activity tracking, awareness and reflection, privacy interface, interests mining, knowledge discovery, recommender system, content analytics, information retrieval, knowledge delivery, peer-to-peer, humanitarian knowledge management.

Résumé

Les plates-formes sociales ont démontré leur capacité à faciliter le partage de connaissances à une échelle globale. Elles sont de plus en plus souvent employées dans le domaine éducatif et humanitaire où, malgré leurs avantages généraux, elles présentent des défis particuliers. Cette thèse s'inscrit dans le contexte de recherche associé au projet Européen Go-Lab et à une collaboration avec Médecins sans Frontières (MSF) qui exploitent intensivement les plates-formes sociales.

Quatre défis sont traités dans cette thèse liés à l'analyse, la confidentialité, la découverte et la diffusion de connaissances, afin de répondre à quatre questions de recherche associées. Comment fournir des analyses orientées utilisateurs dans les systèmes de partage de connaissances pour aider à la sensibilisation et la réflexion? Quelles interfaces et mécanismes de gestion de la confidentialité sont adaptés à l'analyse de connaissance et d'apprentissage? Comment permettre la découverte de connaissances alignées avec les intérêts des utilisateurs? Comment faciliter la diffusion de connaissances dans des environnements où la connexion à Internet est limitée ou absente? Ci-dessous, nous présentons un aperçu de nos résultats.

Analyse. Pour faciliter la sensibilisation et la réflexion des utilisateurs dans les plates-formes sociales, nous proposons un modèle d'intégration d'analyses au contexte d'interaction et de visualisation d'informations pertinentes associées. Nous proposons également deux architectures de matérialisation du modèle proposé basées respectivement sur des applications d'analyse en temps réel et des services évolutifs d'arrière plan. En exploitant ces architectures, nous fournissons des analyses aux utilisateurs des plates-formes sociales. Une évaluation réalisée avec des utilisateurs a montré que des analyses contextuelles intégrées aident à la sensibilisation et la réflexion.

Confidentialité. Pour faire face aux préoccupations associées à la capture, au stockage et à l'analyse des traces d'interaction, nous proposons un nouveau modèle de gestion de la confidentialité basée sur un agent. Notre contribution repose sur la métaphore de la présence d'un agent d'observation et permet de gérer la politique de capture comme dans le monde réel. Nous avons mis en oeuvre le modèle proposé dans une plate-forme sociale et obtenu des résultats d'évaluation positifs de la part des utilisateurs.

Découverte. En raison du grand nombre d'éléments stockés dans les plates-formes sociales, il peut être laborieux pour les utilisateurs de trouver du contenu pertinent. Pour relever ce challenge, nous proposons un système de recommandation interactif basé sur les intérêts des utilisateurs qui permet la découverte de contenu et d'utilisateurs. Nous avons mis en oeuvre le système de recommandation proposé dans une plate-forme sociale et avons réalisé deux évaluations avec des utilisateurs. Les évaluations ont montré la capacité de l'approche proposée d'identifier les intérêts des utilisateurs et de recommander du contenu pertinent.

Diffusion. Au moment de la rédaction, quasi la moitié de la population mondiale n'a pas accès à une connexion Internet fiable. Généralement, les endroits où l'action humanitaire est requise bénéficient également d'une connextivité réduite. Nous proposons un nouveau modèle de diffusion du savoir qui repose sur un intergiciel de communication entre pairs et exploite des ordinateurs à bas prix pour la réplication locale de connaissances. Nous avons réalisé un prototype basé sur le modèle proposé et l'avons évalué dans huit missions de MSF. L'évaluation confirme les capacités de diffusion de connaissances et son utilité pour le personnel de terrain.

Mots-clés: médias sociaux, partage de connaissance, analyses de connaissance, analyses d'apprentissage, sensibilisation et réflexion, interface de confidentialité, exploration des intérêts, découverte de connaissances, systèmes de recommandation, analyse de contenus, exploration de textes, suivi d'activités, récupération d'informations, diffusion de connaissances, partage entre pairs, gestion de connaissances humanitaires.

Contents

List of Figures

List of Tables

Glossary

- **CDN** - Content Delivery Network
- **CD** - Compact Disk
- **CERN** - Conseil Européen pour la Recherche Nucléaire - The European Organization for Nuclear Research
- **CSCL** - Computer Supported Collaborative Learning
- **CSCW** - Computer Supported Collaborative Work
- **DNS** - Domain Name System.
- **DSL** - Digital Subscriber Line
- **DVD** - Digital Versatile Disc
- **EDM** - Educational Data Mining
- **ELK** - A stack of technologies consisting of Elasticsearch, Logstash, Kibana
- **HCI** - Human Computer Interaction
- **HDD** - Hard Disk Drive
- **ILS** - Inquiry Learning Space
- **ISP** - Internet Service Provider
- **IT** - Information Technology
- **KMS** - Knowledge Management System
- **KM** - Knowledge Management
- **KS** - Knowledge Sharing
- **LA** - Learning Analytics
- **MOOC** - Massive Open Online Course
- **MSF** - Médecins Sans Frontières

- **NGO** - Non-Governmental Organization
- **NRT** - Near real-time
- **OLE** - Online Learning Environment
- **P2P** - Peer-to-peer
- **PLE** - Personal Learning Environment
- **SBC** - Single Board Computer
- **SRL** - Self-Regulated Learning
- **STEM** - academic disciplines of Science, Technology, Engineering, and Mathematics
- **SUS** - System Usability Score. A survey often used to evaluate the usability of a system or a component
- **SM** - Social media
- **Space** - a container for user activity, may enclose content, applications and users
- **UI** - User Interface
- **USB** - Univeral Serial Bus
- **UX** - User Experience
- **cMOOC** - connectivist MOOC

1 Introduction

Knowledge plays a significant role in value creation in the post-industrial economy. Knowledge is acquired and enriched in life-long learning, which can take place in various forms and locations, in an educational setting or at a workplace, online or offline, often through communication and collaboration with peers. The development of information technology moves us closer to the ultimate goal of universal knowledge sharing anytime anywhere. One example of such technological advances often employed to share knowledge on a global scale is social media platforms. Social media platforms are increasingly used to support knowledge sharing, both when learning in educational contexts and when collaborating with colleagues within organizations. Yet, when deploying social media for collaborative learning and knowledge sharing in such authentic settings a set of issues emerge that need to be addressed to enable adoption.

In this thesis, we take a close look at individual parts of the knowledge sharing process, identify issues encountered in practice when using social media to support knowledge sharing and state six core requirements to be met by a social media platform to become suitable for knowledge sharing in the educational and humanitarian domains. Building on these requirements, this thesis contributes new approaches to address the identified challenges and presents their validation in a knowledge sharing platform called Graasp. In addition, we discuss deployment and exploitation of the proposed solutions in authentic settings. This thesis contributes to addressing the challenges of (1) providing knowledge analytics for awareness and reflection, (2) allowing contextual privacy management with an agent-inspired interface, (3) enabling discovery of knowledge corresponding to user interests and, (4) facilitating knowledge delivery into environments with limited Internet connectivity.

1.1 Graasp

In this section, we describe Graasp, a social media platform that serves us as a test bench when prototyping and validating the contributions discussed in this thesis.

Graasp[1,2] (also known as Graspeo[3]) is a social media platform designed to support collaboration, learning and knowledge sharing [9, 10, 111]. Graasp supports people in their personal and collaborative activities. Graasp facilitates many types of activities, but most often it is used as a personal learning environment. Users can organize their activities into public or private contextual spaces, where they can share relevant resources and tools with other users. Unlike mainstream social media platforms, Graasp focuses on activity-based relations (e.g. people studying English), instead of individual social networks (e.g. people who I know). Due to this emphasis, Graasp is particularly suitable for decentralized structures such as schools, universities [37] and humanitarian relief organizations [111].

1.1.1 Space as interaction context

Graasp is built around the concept of a **space**, which defines the **context of knowledge sharing**, or more generally speaking the **collaborative interaction context**. In practice, spaces are used for blended learning sessions, project workgroups, topic-based knowledge sharing groups, communities of practice, organizing teamwork, courses, MOOCs, etc. The user interface of Graasp is presented in Figure 1.1. The "Green Energy" space contains "Solar Energy" and "Wind farms" subspaces as well as a "BBC Documentary" video, a link to a Wikipedia page, a "Team Discussion" discussion thread, a pdf report and two images.

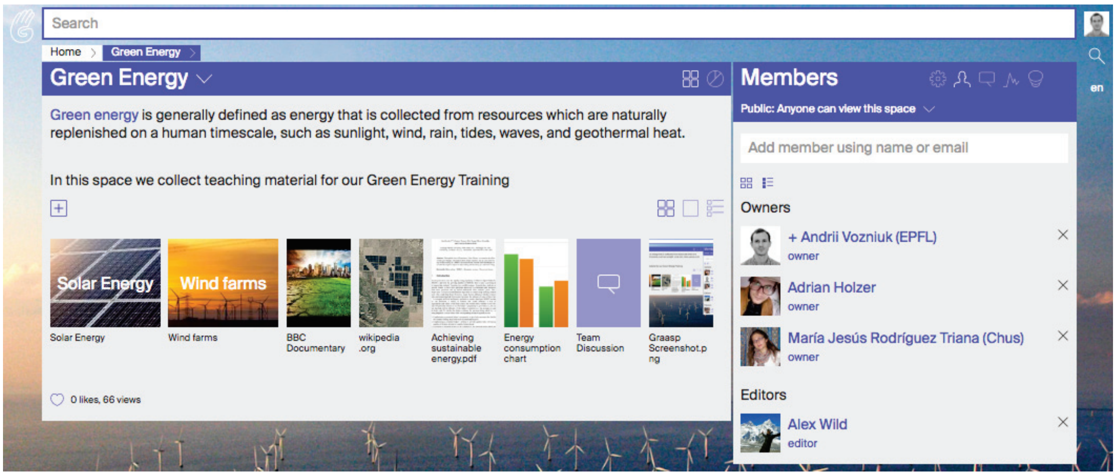


Figure 1.1 – The "Green Energy" space in Graasp.

Below, we formally define the core concepts in Graasp.

Spaces. Let s denote a space. The hierarchical structure of Graasp consists of a set of spaces $\{s_1, ..., s_q\}$ organized in trees, where every space can have a set of sub-spaces. We say $s' < s$ if s' is part of the sub tree where s is the root. We use the

[1] Graasp.eu https://graasp.eu (last accessed 10 November 2016)
[2] Graasp.net https://graasp.net (last accessed 10 November 2016)
[3] Graspeo.org https://graspeo.org (last accessed 10 November 2016)

family tree metaphor and say that s' is a descendant space of s and that s is an ancestor space of s'. If there is no space s'' such that $s' < s'' < s$ then s' is called the child of s and s is the parent of s'. There is a special system-level root space, which is an ancestor of all spaces.

Items. Let i denote an item. Items in Graasp can be of five main types: (1) spaces, (2) content (coming from the local disk or the cloud), (3) applications (e.g. OpenSocial apps), (4) discussion threads and (5) alias items providing dynamic links to other items. A space s in Graasp acts as a container for *items* $I(s) = \{i_1, ..., i_n\}$ located in this space and each item in Graasp has a space, where it is located.

Members. Each space s has a set of members $M(s) = \{m_1, ..., m_k\}$. Each member m of a space has a defined permission in this space. The member permissions, denoted by p, are from the set of permissions {`owner`, `editor`, `viewer`}, where owners have more rights than editors, which have more rights than viewers, i.e. `owner` > `editor` > `viewer`).

Activities. When space *members* interact with space *items*, they generate a set of *activities* $A(s)=\{a_1, ..., a_p\}$ in the space describing these interactions. The activities can be captured and stored in Graasp for further usage.

Settings. The space has as well a list of *settings* $S(s) = \{s_1, ..., s_l\}$ including the space type, the space description, the background image and the space thumbnail icon. Each space can be shared with others, either by making it *public* and accessible to anyone on the Web, or by keeping the space *private* accessible to only the space members.

Presentation. The space has a *presentation*, that is the way it is displayed to the user in Graasp, i.e. its graphical user interface as shown in Figure 1.1. Some parts of the presentation are configurable, including the space background and thumbnail, the way items are displayed (as a list or as a grid) and are defined in the space settings.

1.1.2 Applications of Graasp

The core concepts of Graasp were developed in the framework of three consecutive European initiatives in Technology Enhanced Learning (TEL) aiming at defining online support for communities of practice (CoPs) [92], personal learning [38, 72, 39] and inquiry learning at school [27]. During these projects, the need to provide a versatile and agile social media platform strengthening digital education has emerged. Recent developments were as well motivated by knowledge sharing requirements in large-scale humanitarian relief organizations [111].

The current version of Graasp was designed and developed from the ground up based on requirements of knowledge sharing in inquiry-based learning in the framework of the Go-lab project[4] and to meet requirements of knowledge sharing in humanitarian organizations [111] in the context of our collaboration with Doctors Without Borders[5]. The current version has added such features as advanced search, discussions, fine-granular activity tracking, interactive contextual analytics, innovative privacy management, recommendations, and offline content access. Some of these new features were built as an outcome of the investigations carried out in this thesis. The features of the current version and its core ideas are summarized in [41].

In the Go-Lab project teachers use the Graasp platform to support blended inquiry learning activities. In inquiry-based learning, the learner is expected to go through a set of phases, e.g., Orientation, Conceptualization, Investigation, Conclusion, Discussion, sometimes sequentially, but also moving back and forth between phases [26]. Graasp provides a special type of space, called Inquiry Learning Space (ILS) with the aforementioned inquiry learning phases as sub-spaces [93]. An ILS has a teacher view and a student view also called “a standalone view”. The former is used by the teachers to assemble an ILS, and the latter is accessed by students through a secret URL that the teacher shares with them. Figure 1.2 shows an example of the teacher view.

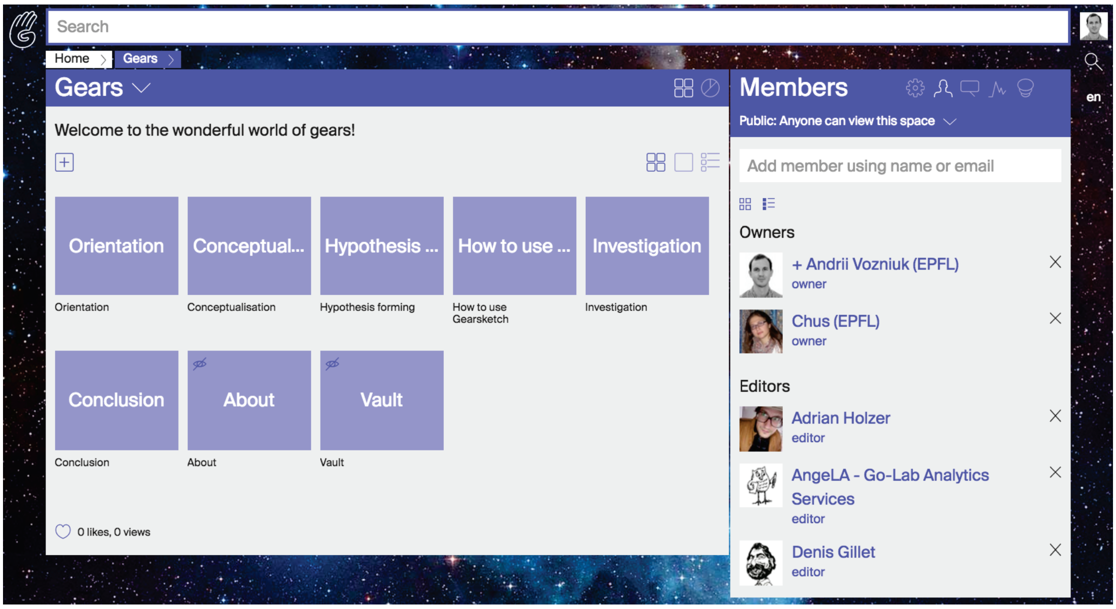


Figure 1.2 – The teacher view of an inquiry learning space in Graasp.

When used by MSF, the headquarters (HQ) staff members typically upload essential documents into the appropriate space, organize them and the field staff members download the content when needed. As an example, Figure 1.3 shows the MALARIA space containing content relevant to the malaria thematic.

[4]The Go-Lab project http://www.golab-project.eu (last accessed 10 November 2016)
[5]MSF http://www.msf.org (last accessed 10 November 2016)

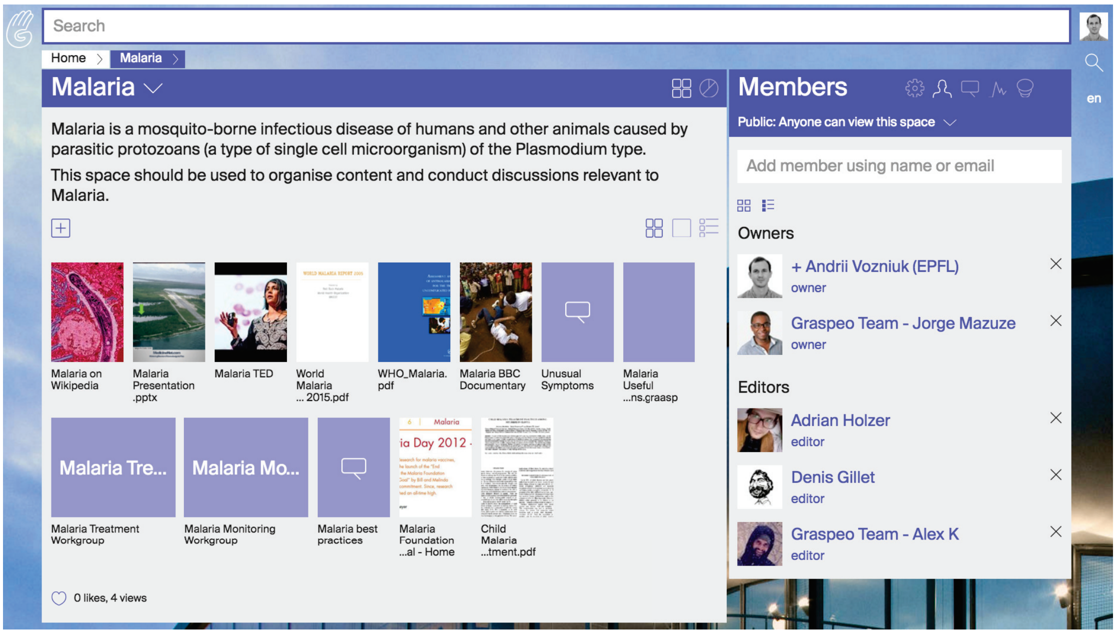


Figure 1.3 – The "Malaria" space in Graasp employed to organize collaboration and share knowledge related to the malaria topic.

1.1.3 Architecture of Graasp

From the technical point of view, Graasp is a MEAN web application [6] using JavaScript end-to-end [81]. As presented in Figure 1.4, it is composed of the front-end part, running in a browser and the back-end part installed on a server. The front-end is developed with the AngularJS[7] framework that allows minimizing the data sent to the clients since only templates are exchanged instead of complete HTML pages. This saves bandwidth and decreases latency leading to a more responsive web page in places with low bandwidth. On the back-end, a Node.js[8] server is handling the business logic and the data is persisted in a MongoDB[9] database.

1.2 Knowledge Sharing in Educational and Humanitarian Domains

In this thesis, we refer to knowledge sharing as an activity through which knowledge, i.e. information, skills, or expertise, are exchanged among people, friends, families, communities or organizations [3]. When investigating approaches supporting knowledge sharing we focus on two domains: educational and humanitarian. Both of these domains are knowledge-intensive, and the efficiency of the activities conducted there strongly depends on how well the knowledge sharing is performed [68, 79]. As such, these domains are of particular interest

[6] The MEAN stack http://mean.io/ (last accessed 10 November 2016)
[7] AngularJS https://angularjs.org/ (last accessed 10 November 2016)
[8] Node.js http://nodejs.org/ (last accessed 10 November 2016)
[9] MongoDB http://www.mongodb.org/ (last accessed 10 November 2016)

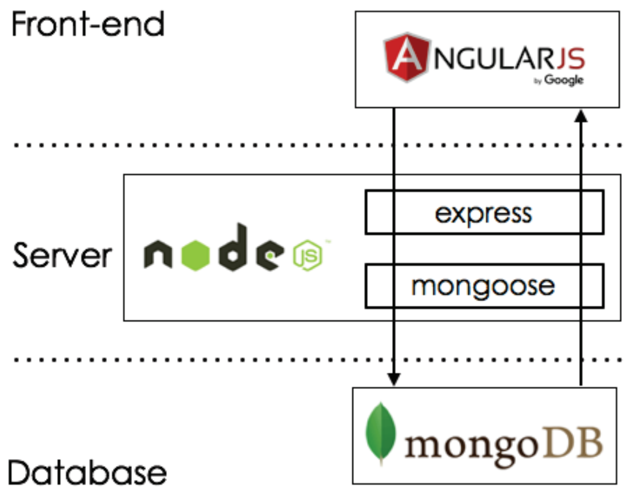


Figure 1.4 – A schematic representation of the 3-tiers of Graasp architecture.

when analyzing knowledge sharing needs, proposing solutions and verifying their validity.

Identifying and understanding typical user scenarios is usually the first step when designing a suitable information technology [5]. Below, we discuss in details scenarios determined during participatory design sessions carried out with the core stakeholders involved in the knowledge sharing process. For the educational domain, we worked with teachers from the Go-Lab project when we studied how to strengthen educational knowledge sharing with SM platforms. In the case of the humanitarian knowledge sharing, we communicated with knowledge managers and field staff of MSF when investigated the application and extension of SM platforms to support humanitarian knowledge sharing. We cover the participatory design methodology in Section 1.3. Below, to define the context of this thesis, we describe each of the settings and typical knowledge sharing scenarios we have identified.

1.2.1 Educational Knowledge Sharing: the Go-Lab Project

In its essence, education is about the creation, transformation, and transmission of knowledge [68]. The knowledge sharing between the teachers and the students, between the teachers and between the peer students are the fundamental parts of the educational process. This sharing is particularly relevant for self-regulated learning or when learning in networked structures like in connectivist MOOCs (cMOOCs), where often students learn from the knowledge shared by peers, compared to traditional classrooms where most of the knowledge comes from the teachers. Improving the technology support of knowledge sharing in such settings is considered to be a promising way to improve teaching and learning leading to better education.

In this thesis, we consider the case of the Go-Lab European project [25], where the author has participated, as an example of large-scale knowledge sharing initiative in education. Below we provide more details about the project itself.

Go-Lab Project.

Go-Lab is a European project from the EU's Seventh ICT Framework Programme (FP7) and is motivated by the need to sustain the growth of the knowledge economy that requires more and better-educated people in science, technology, engineering and mathematics (STEM). The project aims to encourage and orient students from an early age to study STEM subjects and select these fields in their future educational and career path [42]. The project applies inquiry learning using online labs and allows the students to engage in scientific topics and get acquainted with scientific inquiry methodologies through the use of remote laboratories, virtual experiments, and datasets. Blended inquiry learning relying on online labs to enhance classroom activities is considered as a promising approach to increase the skills and the interest of students in STEM [27].

Inquiry Learning Space (ILS).

Inquiry learning typically leads students through various phases (demonstrated in Figure 1.5), e.g. Orientation, Conceptualization, Investigation, Conclusion and Discussion, where students formulate hypotheses, evaluate them through experiments and then reflect on them, possibly repeating the cycle. This type of learning shows benefits over usual lectures or demonstrations [82]. In Go-Lab, the inquiry learning is structured with the help of inquiry learning spaces (ILSs), presented in Figure 1.5. Such ILSs are designed to support the study of STEM topics and can be considered as an educational resources [93]. The ILS itself enforces inquiry learning phases to promote the study of science by employing a scientific methodology.

ILSs can be created and enriched by using the ILS teacher view in Graasp, which allows to populate spaces with relevant content (e.g., PDFs, instructions & videos), applications (e.g., concepts maps, chat & drawing tools) and labs (e.g., chemistry simulations and remote hardware such as a telescope). Figure 1.2 presents the teacher view of the Gears inquiry learning space. When an ILS is presented to students, it takes the content and apps from the space and arranges them using tabs in the ILS student view, making the content easy to navigate (shown in Figure 1.5). The upper bar contains the title of the ILS and the student's nickname. The tabs of the ILS represent the inquiry phases and, in the case shown in Figure 1.6, the Conceptualisation phase explains how to use the Gearsketch gears simulating application. To access an ILS, the student just opens in the browser the ILS URL provided by the teacher. The browser-based access allows the student to interact with the ILS online from any device without the need to install any custom software or plug-ins.

In addition to being a means to structure learning materials and an interface to online laboratories, ILSs provide tools such as scaffolds to improve the learning process. The scaffolds are tools that aim to help students to stay in the zone of proximal development, providing guidance and help when needed. The scaffolds can employ learning analytics to analyze activities of students and teachers recorded during their interaction with the ILS. The recorded

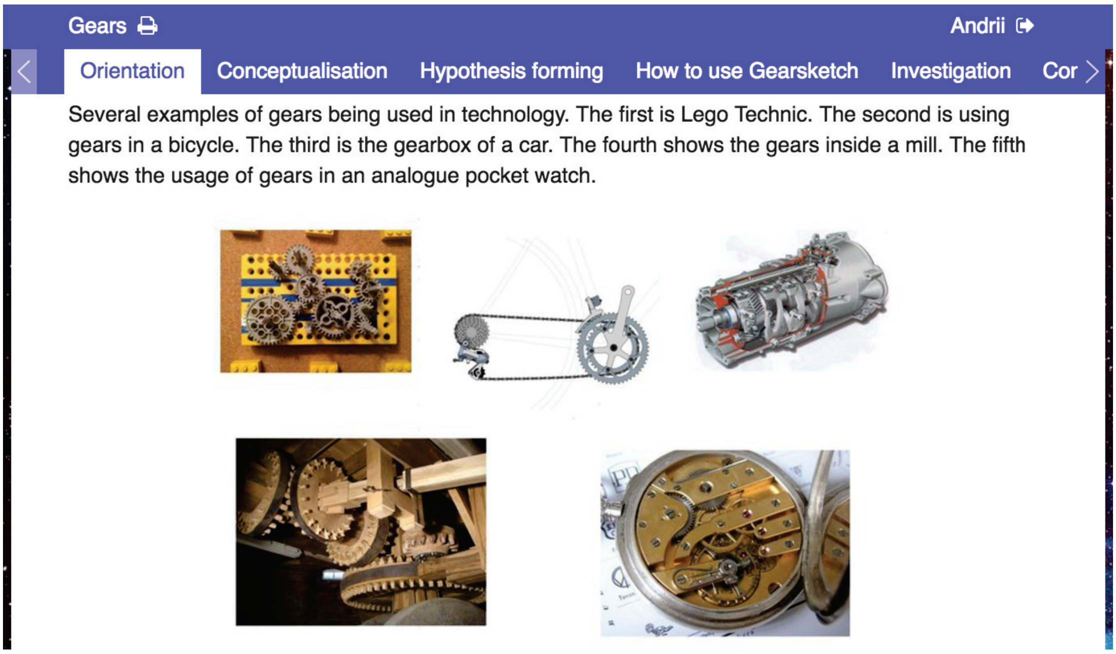


Figure 1.5 – The student view of the "Gears" inquiry learning space. The inquiry learning phases are displayed as tabs on the top bar.

traces can be used for building learning analytics dashboards [110] that can help a teacher to gain a better understanding of the learning happening in an online environment by presenting the relevant information at a glance.

Golabz. After the teacher has constructed an ILS in Graasp, she can use the Golabz repository (shown in Figure 1.7) to share the ILS with other teachers so they can reuse or repurpose the ILS if necessary. The Figure 1.8 depicts an ILS available in Golabz that can be copied and used by teachers. In addition to ILSs, Golabz stores other types of educational resources as well as apps and labs immediately available to the teacher when constructing her ILS. On Golabz, teachers can discover, use and enhance online labs appropriate for their courses. Through the Golabz repository, Go-Lab aims to establish a federation of online labs where lab owners can promote their labs, and teachers can find labs to support their activities and share their resources with others [42].

With Inquiry Learning Spaces, Graasp and Golabz, Go-Lab provides a common, ubiquitously accessible ILS platform, so schools do not need to spend resources on installing and administering the software. To better understand the needs of the teachers and how they can be supported in the knowledge sharing, below we consider a scenario typical to Go-Lab.

Blended Learning Scenario. Primarily, Go-Lab is targeting two primary types of users: *teachers* and *students*. From the Go-Lab perspective, a teacher is a user who teaches with an ILS, and a student is a user who carries out inquiry learning activities in an ILS. In Go-Lab, knowledge sharing can happen among the teachers at the ILS constructing stage or from the teacher to

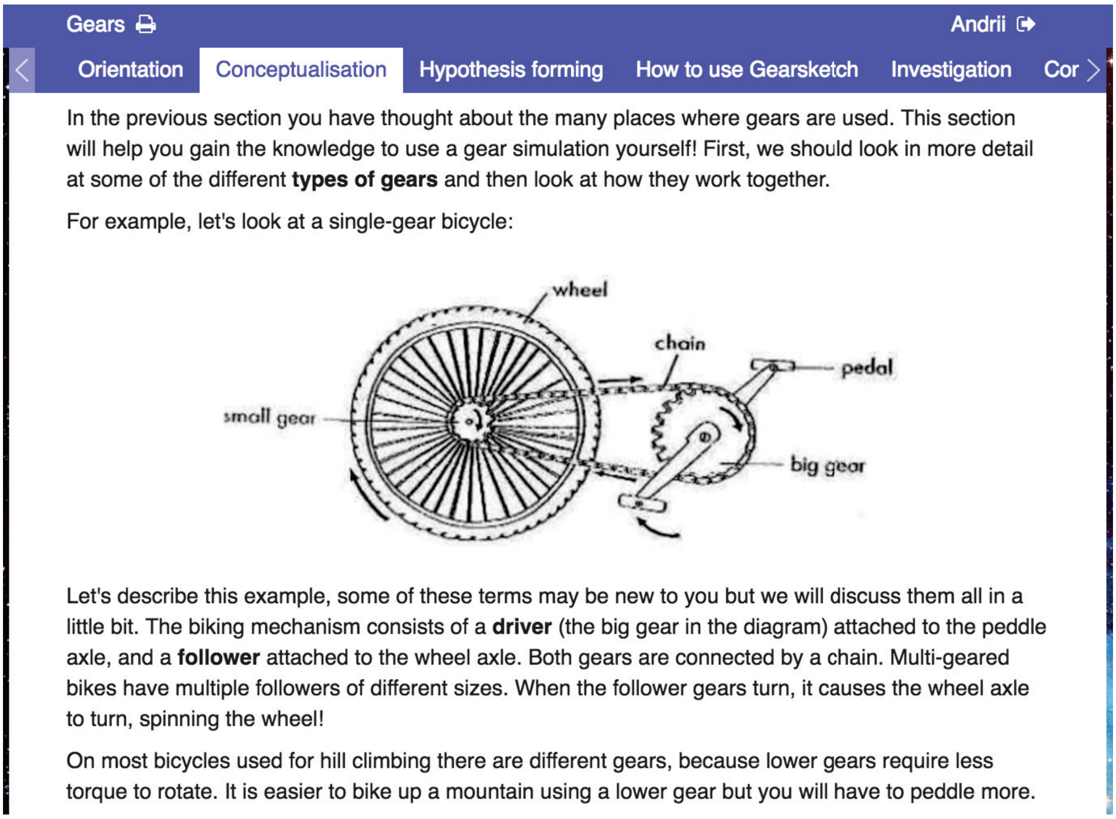


Figure 1.6 – The student view of the Conceptualisation phase of the “Gears” ILS.

the students during a blended inquiry learning session. The following user story provides additional context.

Imagine John, a school physics teacher. John searches on the Golabz repository (shown in Figure 1.7) for engaging inquiry learning spaces for his physics course to explain how gears work. While browsing through the “Big Ideas” in the science section, he finds an appropriate topic about energy, where he selects the “Gears” inquiry learning space (presented in Figure 1.5) explaining the gears concept and containing interactive application “Gearsketch” allowing to draw and experiment with gears. John copies the ILS into his social media platform Graasp to try it out.

When preparing the materials for his new class using the selected ILS, to adapt the ILS to his needs and knowledge of his students, John rewrites some of the explanations and tips. John also would like to reuse relevant materials uploaded by other physics teachers, but he is not able to find them. John thinks that it would be nice if the system automatically suggested content relevant to his interests or the ILS content, so he can save time and build more rich ILSs. After tweaking each inquiry learning phase, the ILS is ready to be used by John in class.

While conducting a blended learning session with his ILS, John wants to see live how his students are performing, on which part of the ILS they are currently working and how the

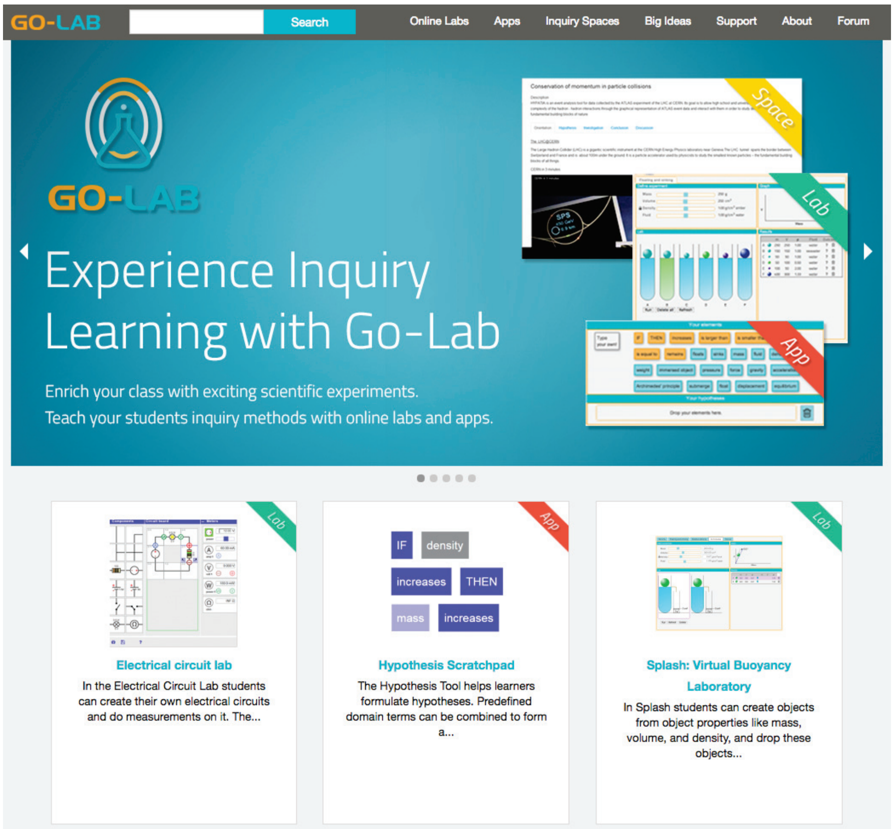


Figure 1.7 – The Golabz repository contains apps, labs, and inquiry learning spaces.

progress of each of the student compares to the overall class progress and to the time allocation that John had in mind. For this purpose, he would like to have analytics in ILS that shows the performance of the students. Moreover, when the session ends, John would like to analyze how students interacted with the learning resources and explore the interactions in order to identify points that may require better explanation so John can adapt the ILS accordingly. When John feels that his ILS is polished, he decides to publish it from Graasp back on the Golabz repository to share it with his colleagues.

Since John teaches physics not only in his school but also volunteers to visit developing countries, he often finds himself in a situation where there is only a limited Internet connectivity in the school. Hence he carries the materials on a USB stick and distributes them manually in the classroom. John needs better technology to support his teaching in such environments where the cloud access is not guaranteed.

Apart from teachers, the primary target users of the Go-Lab project are students from 10 to 18 years old. Due to the age of the students, their data privacy is of particular importance and is subject to special legal regulations. For instance, the European Union provides a data privacy framework, through the EU directives 95/46/EC (Data Protection Directive) and 2006/24/EC (Data Retention Directive). Knowing that, John would like to be in control of the data being

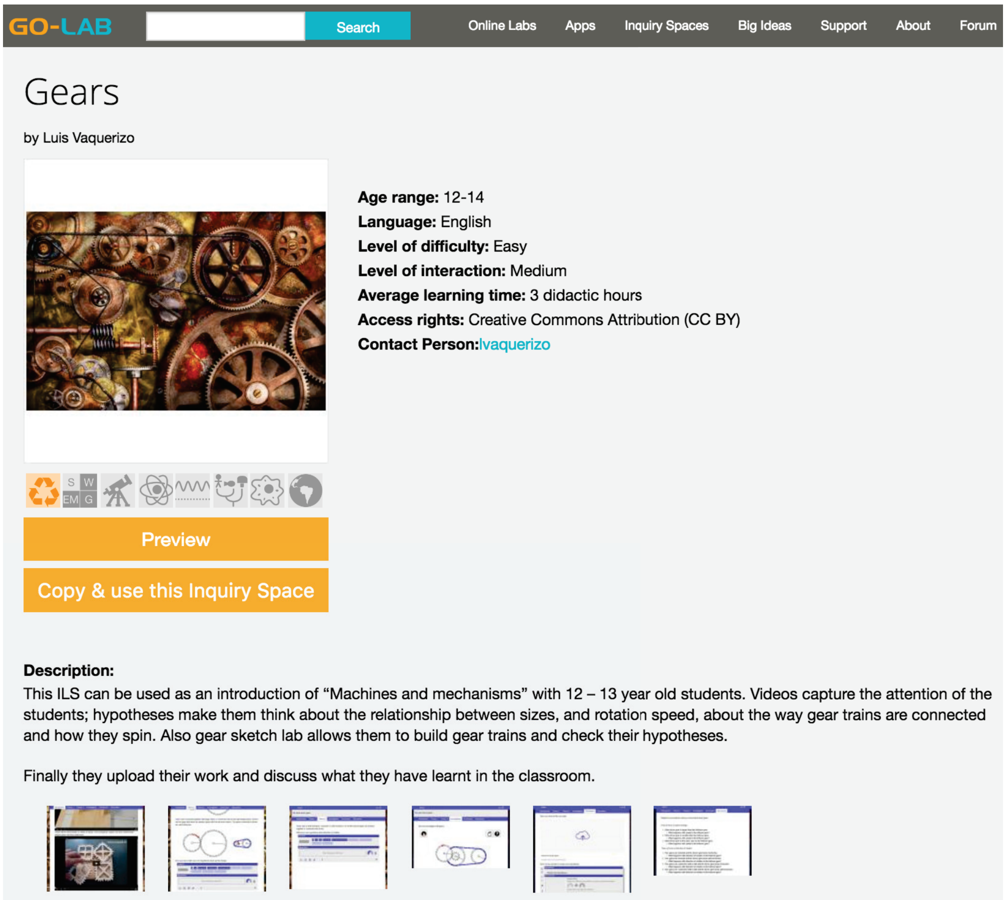


Figure 1.8 – The "Gears" ILS available in the Golabz repository.

collected by the learning tools so he can disable the collection when needed.

The described scenario provides a short overview of the activities a school teachers conducts in Go-Lab. Our earlier works [42, 93] present a complete requirements analysis, the Go-Lab portal and its architecture.

1.2.2 Humanitarian Knowledge Sharing: Médecins Sans Frontières

With international relations becoming more complex, the role of the representatives of civil society is bound to become more important. The increase in the number of non-governmental organizations (NGOs) is witness to this evolution. Currently, the United Nations registered over 22'000 NGOs[10] among which over 3'300 have consultative status with the Economic and Social Council (ECOSOC).

Knowledge is often regarded as one of the most valuable resources available to non-governmental organizations (NGOs) [79]. Hence, non-governmental organizations need to be able to access

[10] UN iCSO System http://esango.un.org (last accessed 10 November 2016)

critical knowledge timely and reliably and build and share knowledge efficiently between different, often geographically dispersed, teams. Such knowledge sharing is often found in NGOs to involve outside stakeholders, for example through blogs and video sharing platforms [65]. Hence, public social media use by NGO employees is growing, while deployment of social media inside of organizations is still in its early stage [119]. Matschke et al. [79] argue that an adequate KM system for NGOs should rely on social media platforms since NGOs and social media have similar characteristics, such as voluntariness, participation, personal relevance and non-formalization.

Médecins Sans Frontières (MSF)

Médecins Sans Frontières (MSF) or Doctors Without Borders was created in 1971 and is one of the leading NGOs in the humanitarian and medical field delivering emergency medical aid quickly, effectively and impartially. The typical activities of MSF include performing surgery, providing health care, carrying out vaccination campaigns, or setting up sanitation systems. In 1999, MSF received the Nobel Peace Prize in recognition of its humanitarian and medical work. It is currently listed amongst the top ten NGOs in general and the top three in the humanitarian category [1].

MSF employs over 33'000 staff worldwide, and its structure can be seen from two points of view: an organizational governance structure and an operationally driven structure. On the organizational level, MSF is organized as an international movement made up of 24 associative organizations, of which 14 are operationally accountable to the five operational centers (France, Switzerland, Belgium, The Netherlands and Spain) overlooking 19 national offices as illustrated in Figure 1.9. On the operational level, MSF has over 70 missions coordinating over 325 projects. The MSF operational culture is rooted in the rapid deployment of emergency medical services on short notice in highly complex contexts. The distributed structure of MSF, the urgency it operates in, and the high turnover of staff working in demanding environments bring to the surface challenges related to large-scale common knowledge sharing infrastructure.

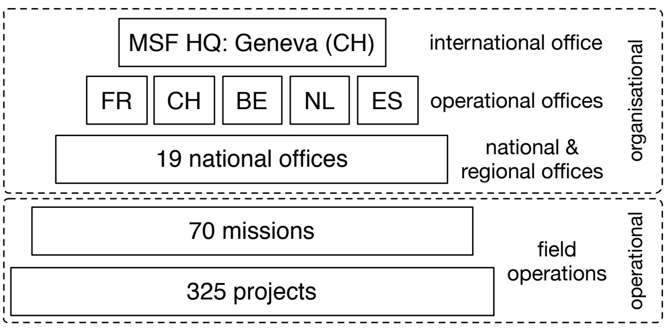


Figure 1.9 – Médecins Sans Frontières organizational chart.

MSF has identified improvement of knowledge sharing as one of the strategic objectives. To achieve this goal, MSF aims to connect experts from the headquarters of the organization and field workers with knowledge relevant to the situations that MSF deals with. To support the knowledge sharing scenarios, at the moment of writing, MSF relies on a total of 61 different web repositories and content management systems making the information scattered and difficult to find. While many systems are available to MSF employees, still most of the communications are done through email, which combined with data duplication leads to a chronic information overload making it difficult to find relevant information rapidly. It was estimated during an internal MSF audit that an equivalent of two hours a day per person is wasted requesting and waiting for information. Furthermore, information from the field is summarized and relayed from missions to the headquarters, which can reduce the information load without cutting out an essential part of the field experience which could serve for subsequent projects as well as reports for donors and auditors. According to an MSF knowledge manager, MSF believes that providing the right knowledge to the right people at the right time can positively affect the organization's operational performance. The same manager highlighted that MSF has tried over 100 different existing platforms to solve this issue, with no success. With this in mind, it seemed crucial to go back to the drawing board and elicit adequate requirements for an effective knowledge sharing system.

Figure 1.10 – An interview with MSF field staff in Niger conducted by an MSF Manager.

In order to better understand the requirements for a knowledge sharing system, our colleagues from MSF conducted in-depth field studies that consisted of a total of 145 hours of interviews including 20 hours in Geneva, 100 hours in Niger and 25 hours in Swaziland between June and August 2013. In the Geneva MSF headquarters, interviews were conducted with the General Director, the head of operations, a head of units, project managers, and international experts. During the field trip to Niger and Swaziland, interviews with heads of mission, medical coordinators, medical, logistics experts, and nurses were done. Figure 1.10 illustrates an interview with field staff in Niger in July 2013. In these interviews, feedback regarding knowledge sharing was collected through discussions and reflection on potential process improvements, followed by the formal conceptualization of work practices through participatory design that led to the development of the first proof of concept that was used to refine requirements further [111]. In Section 1.4 we present the elicited requirements in more details.

Hereafter, we illustrate information flows in MSF through a representative user scenario formulated in participatory design sessions with an MSF Knowledge Manager (see Section 1.3).

HQ-Field Knowledge Sharing

Imagine, Amy, a field doctor, deployed to an emergency project in Zobia, a rural village in Congo, where she will work in a clinic to treat sleeping sickness[11]. Concerning technological infrastructure, field workers can only count on very limited Internet connections, if any. For instance, Figure 1.11 shows a 3G mobile Internet connection employed by the MSF mission in Kampala, Uganda.

Before the mission. The information Amy needs before she travels to Congo includes training material and updated information on the situation on the ground. The headquarters produce the training materials which are maintained by experts from different MSF offices. The information from the field is created by Nermeen, the head of mission based in Zobia. Furthermore, Wikipedia and other online public resources are a rich source of information relevant to the mission context, some of which Amy will want to save to her workspace for quicker access on-site. An example of such workspace aggregating heterogeneous content related to Malaria is presented in Figure 1.3. Regarding the infrastructure, the headquarters is equipped with a broadband Internet link and the Zobia mission has a satellite Internet connection.

During the mission. In the first week of the intervention, some unexpected issues start to appear, and Amy would like to get more information on sleeping sickness case management. To resolve these matters, she sets up a sleeping sickness task force composed of other medical experts and referents in Congo as well as other specialized staff in other countries with cultural similarities and international experts outside the organization. During her stay, Amy and her colleagues in the field office use documents provided and updated by the HQ, relevant to the situation on the ground and the mission activities.

[11] Sleeping Sickness in the Congo https://hub.maf.org/blog/sleeping-sickness-in-the-congo (last accessed 10 November 2016)

Nermeen would like to see how the knowledge available in the repository is being produced and consumed. For this purpose, he would like to have access to analytics that allows him to see popular and not popular content, active and not active users. Knowing the knowledge needs of his mission, he can better organize the production of material that can be useful to the mission staff. Moreover, based on the analytics, Nermeen can reward employees actively participating in the knowledge sharing and identify the inactive ones.

Nermeen is aware that enabling analytics requires recording of user traces. But he is also concerned that in some cases where he operates, such data is sensitive and can put mission members at risk. In such situations, Nermeen would prefer not to record traces, or at least pause their recording for some period until it becomes safe again.

When Amy is working in the field, she would like to be able to use relevant content about sleeping sickness created by her colleagues from other missions or even other humanitarian organizations. But such content can be challenging to find in a large knowledge repository and even more challenging to follow when new relevant content is being created. Amy would be happy if the knowledge sharing system would automatically identify relevant content and recommend it to her.

The Internet connection in the Zobia mission is unreliable, and connection drops are frequent. In such setting, Nermeen and Amy can not rely on cloud services for knowledge access. They need a solution that would enable fast relevant knowledge access even in situations when Internet connection is not available or limited. Ideally, that solution should be able to sustain short power outages as well.

After the mission. After coming back from the mission, Amy delivers her reports to Nermeen. Furthermore, she suggests updating the case management guidelines covering sleeping sickness to include updated knowledge she gathered through her exchanges with other experts and her field experience.

1.2.3 Discussion

From the support process perspective, both scenarios of educational and humanitarian knowledge sharing have common needs. In the educational scenario, teachers need to assemble relevant educational resources, afterwards, make them accessible to the students and monitor the student-resource and teacher-resource interaction to be able to improve the resources or the learning process if necessary. In the humanitarian setting, field doctors and HQ experts can collaboratively organize a workspace populating it with information relevant to the missions context. Afterwards, this information should be made accessible by a field worker so she can use it in her context, and finally knowledge managers need to be able to monitor the knowledge production and consumption to adjust the process when necessary.

Based on these common process support needs, in the next section, we formulate requirement for a knowledge sharing system. Both the teacher and the humanitarian worker would be

Figure 1.11 – A 3G mobile Internet stick mounted on the wall provides Internet access in the MSF mission in Kampala, Uganda.

happy to be able to benefit from content created by their peers. The teacher can reuse the educational materials by another teacher when designing his own course, and humanitarian workers may learn from materials created by others targeting the same issue (e.g., treating malaria). The similarities in the technological support needs of educational and humanitarian workers may be even more prominent, taking into account an ongoing convergence of knowledge management and technology-enhanced learning [21, 78].

1.3 Approach

In this section, we explain the methodologies we applied in this thesis.

1.3.1 Design and Implementation

In this thesis, we followed an iterative approach combining 1) participatory design ([56, 85, 56]) and 2) iterative agile software development ([67, 114]). In the process, we involved on each iteration the end-users of the developed solution. Starting from an early prototype, we arranged regular meetings with and feedback collection from the expected end-users of the developed solution including MSF knowledge managers and other staff members as well as the Go-Lab project participants. Such an approach allowed to collect suggestions and remarks continuously and adapt the development early on without waiting for a final version, where the cost of change is substantial.

The exposure to early prototypes and their development process starting from mockups and finishing by a working solution had another positive effect: the design participants became attached to the result since they treated it partially as their work [56]. In this way, the participants became early adopters and informed other members of their organization about the developed solution. Early adoption of an implemented solution is of particular importance in organizational environments, where members are often reluctant to the introduction of new technology.

Our research and development process was interdisciplinary in its core and when creating the solution it involved experts from different fields, including computer science researchers, knowledge management practitioners, information technology specialists, field experts and interface designers. Such an interdisciplinary approach is known to be beneficial when designing innovative solutions. The whole process was participation-driven, putting great emphasis on the needs and preferences of the end-user of the proposed solution.

When building the research prototypes, we followed the agile development methodology, where we iteratively developed solutions according to changing requirements emerging in the participatory design.

Evaluation of proposed approaches and, finally, developed solutions is an important part of the applied participatory methodology. When carrying out the investigation, we relied on an exploratory research approach, which included surveys and case studies conducted in the educational and the humanitarian domains. In the next subsection, we discuss the evaluation methodology and the evaluation tools used.

1.3.2 Validation and Evaluation

To validate the relevance of the proposed solutions, we evaluated the developed proof-of-concept prototypes in authentic educational and knowledge sharing settings. When evaluating developed solutions, especially in the field of human-computer interaction, it is often beneficial to combine several evaluation approaches, each with their pros and cons. Following participatory design allowed us to set up early on a channel of continuous feedback useful before as well as after deploying our solutions. When evaluating the solution, we employed two tools used in human-computer interaction research, namely surveys and case studies.

To obtain user opinions, we have used standard surveys including System Usability Scale (SUS) [17] that has known interpretations and allowed to make a comparison between systems [6]. The survey allows to get a high-level view of the perceived value of the developed solution, but not offering deep insight into the value for the end-user and potentially related issues.

Case studies are often a tool of choice when it is needed to observe and evaluate a system performing in an authentic setting. They allow observing how the system is employed in a target context and by the target user. Studying systems in authentic settings enables to obtain a realistic understanding of the system usage compared to high-level feedback obtained with surveys. Some researchers argue that evaluating developed tools in the context where they are expected to be used is the best way to get realistic evaluation data. The drawback of conducting case studies is that they are labor-intensive and require arrangement and physical presence of the researcher in the inspected environment.

1.4 Knowledge Sharing Systems and Requirements

Working simultaneously in the educational and humanitarian domains and targeting the knowledge sharing problem in both areas, we observed that, as a matter of fact, challenges and requirements for knowledge sharing in educational and humanitarian domains happened to have common features elicited through participatory design and detailed hereafter. In this section, we motivate and formulate the requirements for knowledge sharing systems with educational and humanitarian applications. We base our analysis on existing literature, as well as, in-depth field studies of MSF conducted by our colleague from MSF according to the participatory design practices described in Section 1.3 and presented in the previous section and studies done with teachers participating in the Go-Lab project.

Aggregation

Originally, the knowledge is present in the head of individuals that possess it. To be able to manage knowledge, such knowledge should be externalized. In the case of knowledge sharing systems, this implies that the internal knowledge should be transformed into a digital form using some representation.

A knowledge sharing system should provide means for aggregating knowledge from wherever it is already available (source-agnostic) and provide authoring tools for knowledge creation inside of the platform. The system should support aggregation and management of knowledge coming from heterogeneous sources in multiple formats, including uploaded files, web content, as well as from existing knowledge repositories. The system should also facilitate social interactions which are often used as means of knowledge sharing. Examples of such interactions are comments on content, dedicated topic-based discussions, and content ratings. Fulfilling these requirements allows bringing all relevant knowledge into the system. Since part of knowledge remains not externalized and not available in a digital representation, the system should support the aggregation of people that possess the knowledge. Such experts can be reached and can contribute when there is a need for the knowledge they possess. To facilitate the participation of experts in the system, it should support sharing of knowledge through authoring of documents or via forum discussions. Hence we can formulate the first requirement.

Req 1: Support Source-agnostic Knowledge Aggregation.

Organization

After the knowledge was acquired and is already present in the system, it needs to be organized to facilitate its access and update when needed. Often, knowledge sharing platforms are highly centralized and managed by an IT department with a limited number of administrators responsible for decisions regarding knowledge sharing. Such centralization may lead to knowledge sharing bottlenecks caused by, for instance, an inadequate permissions management.

Being able to share knowledge with the right audience is crucial in providing required motivation and balance between the privacy and the prestige of exposure [79]. The content and permissions management should be decentralized (delegated locally to responsible users) to address the issue of centralized bottlenecks and enable scalable processes. In the educational setting of Go-Lab, every teacher can create and publish her own inquiry learning space, in this way decentralizing the ILS production. MSF, as many NGOs operating in the humanitarian context, has a flat organizational structure. As highlighted by [79], the architecture and design principles underlying knowledge sharing systems supporting NGOs should reflect such organizational structure. We can hence formulate the second requirement.

Req 2: Enable Decentralized Knowledge Organization.

Analytics

For effective management of the knowledge sharing process, knowledge managers require an up-to-date information about the process including who are active contributors, which members are not contributing, how the content is consumed, what content is trending (reflecting information needs) and which content is not used at all. These are just a few examples of analytics needs expressed during our participatory design sessions with knowledge managers from MSF. Hence, proper analytics infrastructures are required to enable data-driven knowledge sharing management.

Business intelligence has been employed for a long time to measure the performance of different parts of the business, including among others revenue, sales, and effectiveness of marketing campaigns. When it comes to knowledge sharing, measuring the performance of individual members and the impact of the knowledge exchange on the performance of the whole organization can provide valuable insights. Still, for proper management, it is necessary to employ an analytics infrastructure that captures how individual members interact with knowledge present in the organization. Such an approach, among other benefits, can allow to identify and encourage contribution, find out knowledge needs inside of the organization to address them if necessary.

Proper analytics in humanitarian context can help to identify and potentially recognize those who actively participate in the knowledge sharing. Identifying such active members can be of particular importance for their motivation. This is due to the situation that those who share knowledge are often not valued since the sharing activity can be invisible to their managers. At the same time, the knowledge sharers can spend a considerable amount of time just helping others via a knowledge transfer, or connecting with the right people that have the right knowledge. Following this discussion, we can formulate the third requirement.

Req 3: Provide Knowledge Analytics.

Privacy

The deployment of analytics tools is often associated with privacy concerns. Part of such concerns comes from the legal framework in place in some countries or contexts (e.g., sensitive medical records or data of school students under specific age) and another part comes from the perception of people being tracked. Thus, proper privacy management is required in knowledge sharing systems.

When recording user activity traces, the system should be designed in a way that allows user to change the tracking policy as it was outlined in the user scenarios in Section 1.2 . The user should be in control of the recorded interaction data and the organization employing the knowledge sharing system should be able to decide where to persist the data. At the same time, the system should still meet the analytics requirement, and the analytics component should be able to function providing useful insight into the knowledge sharing process. Hence, we formulate the fourth requirement.

Req 4: Handle User Data Privacy Properly.

Navigation/Search/Discovery

By meeting the previously stated four requirements, a user of a knowledge sharing system can aggregate knowledge from wherever it is located, manage it in a decentralized way, rely on available analytics and be in control of her data privacy. Still, in the end, the key task of a knowledge sharing systems is to put knowledge to work by enabling access to knowledge by those who need it when they need it in their particular context. The following two requirements make sure that developers of knowledge sharing systems focus on making knowledge accessible online and offline.

When the Internet connection is available, knowledge should be accessible online without additional software installation effort. It has to be retrievable easily by navigating in well-structured content or by using a powerful search.

To be able to benefit from the knowledge available on a platform, first of all, the users should be able to find it. Hence, the system should employ a search engine capable of handling files and metadata of different formats. Also, the search engine should be able to search inside of discussions and comments created by the users. Since user-content interaction can be itself a part of knowledge (e.g., which content is useful and which is not), such interactions should be taken into account when the ranking of the search results is done. When providing such computationally-intensive feature as search, the system should be able to benefit from elastic computational resources offered by the cloud to provide a decent quality of service.

It is possible that existing knowledge sharing systems already employed by the organization contain useful knowledge. For instance, in educational settings, there are already numerous repositories with open educational resources. In the humanitarian domain, the case of knowledge scattered over several platforms is common as well. When we consider MSF, it relies on a total of 35 different web repositories and content management systems. In this case, an integrated or federated search can be put in place allowing to find knowledge located in such external systems. Implementation of the integrated search requires the system to index content available on external systems while the federated search relies on the search engines already available in such external systems.

Due to the size of current knowledge repositories containing tens of thousands or even millions items, it is not always feasible to find relevant content just by browsing or using a search engine. The mentioned issues make it necessary for the knowledge sharing system to assist the user in knowledge discovery by providing the user with content relevant to his or her interests. Thus, we formulate the fifth requirement.

Req 5: Facilitate Online Knowledge Access with Navigation, Search, and Discovery.

Delivery

Context-specific technical limitations can have an impact on knowledge access. Possible examples encountered in humanitarian context include crisis environments MSF often operates in, where Internet connectivity issues are common. Some MSF missions have permanently limited or absent Internet connection. At the same time, the mission employees are the ones who put organizational knowledge into practice, and they depend on timely access to up-to-date knowledge coming from the headquarters or other missions (combined with particularly limited time in conditions of urgency). The need for the offline knowledge access was reflected in the call for proposal for a novel knowledge sharing system for MSF.

When we take a look at the educational domain, the issue of limited connectivity becomes even more prominent. More than four billions of people do not have a connection to the

Internet, and hence can not learn by using content available online[12]. This situation leads to the formulation of the sixth requirement.

Req 6: Facilitate Offline Knowledge Access.

Requirements

Based on the analysis we formulated six essential requirements for a knowledge sharing system suitable for educational and humanitarian needs. These requirements can be structured according to three core parts of a knowledge sharing system shown in Figure 1.12: (1) Knowledge Aggregation; (2) Knowledge Organization and (3) Knowledge Access:

1. Aggregation
    - Req 1 – Support Source-agnostic Knowledge Aggregation (Aggregation)
2. Organization
    - Req 2 – Enable Decentralized Knowledge Organization (Structure/Permissions)
    - Req 3 – Provide Knowledge Analytics (Analytics)
    - Req 4 – Handle User Data Privacy Properly (Privacy)
3. Access
    - Req 5 – Facilitate Online Knowledge Access (Navigation/Search/Discovery)
    - Req 6 – Facilitate Offline Knowledge Access (Delivery)

Figure 1.12 demonstrates the relation between the three core parts of a knowledge sharing system and the six requirements we have formulated. Implementing the six enlisted requirements allows social media platforms to become a suitable solution for knowledge sharing.

1.5 Challenges and Contributions

Out of the six formulated requirements (Aggregation, Organization, Analytics, Privacy, Online Access, Offline Access), the knowledge Aggregation (Req 1) and decentralized Organization (Req 2) are the ones better supported by existing tools. For instance, Evernote[13] allows to aggregate content from the Cloud (enabling selection of specific parts of web pages) and the User Devices (e.g., by allowing files upload from computers and mobile devices), and

[12] Four Billion Without Internet https://mic.com/articles/125674/4-2-billion-people-in-the-world-still-don-t-have-internet-here-s-why-that-matters (last accessed 10 November 2016)

[13] Evernote https://evernote.com (last accessed 10 November 2016)

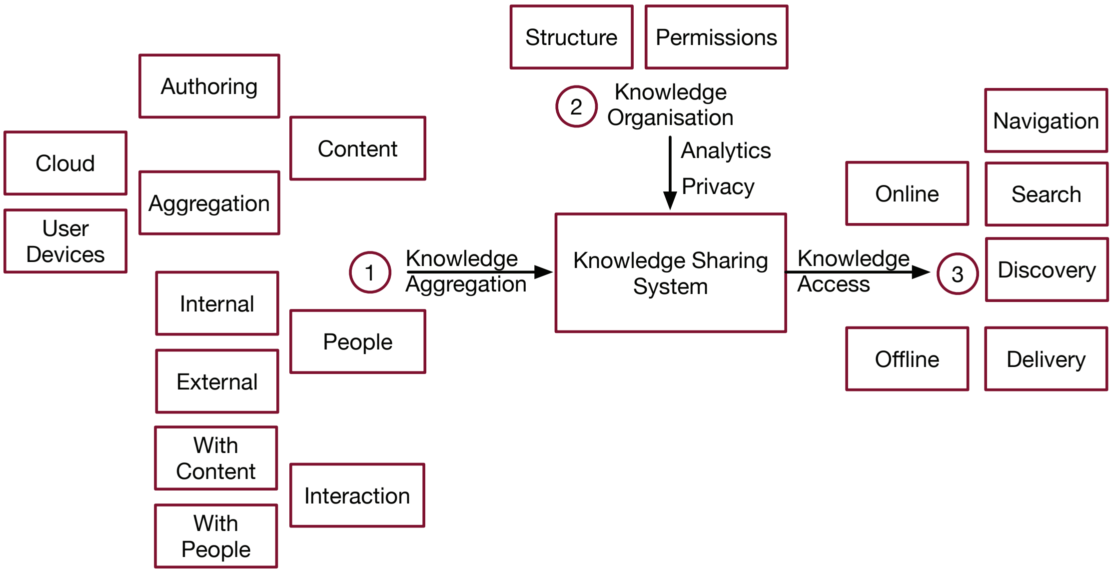


Figure 1.12 – A schematic representation of requirements for a knowledge sharing system, made of three core parts: (1) knowledge aggregation from heterogeneous sources; (2) knowledge organization with analytics and privacy; (3) knowledge access both online and offline.

facilitates discussions between the users. Google Drive[14] and Dropbox[15] have settings enabling decentralized permission management, where the possibility to invite new folder members is not anymore limited only to owners. At the same time, the available mainstream sharing tools insufficiently address requirements of Analytics, Privacy, Discovery and Delivery and hence they are the focus of this thesis.

This section provides a brief overview of the thesis. In each subsection below, we highlight the main research challenges addressed, formulate the research question, present in a nutshell our contributions regarding each question, and outline how the contributions are validated.

1.5.1 Analytics-related Contributions

The two scenarios presented in Section 1.2 share a common problem (**challenge 1**) of making users aware of and allowing them to reflect on interactions happening in a shared collaborative technology-mediated context. Gutwin and Greenberg ([45]) argue that for effective team cognition in a groupware, “the groupware systems must give team members a sense of workspace awareness”. They state that “maintaining this awareness has proven difficult in current real-time distributed systems where information resources are poor, and interaction mechanisms are foreign”. One of the purposes of analytics is to provide tools for supporting such awareness building.

[14] Google Drive https://www.google.com/drive (last accessed 10 November 2016)
[15] Dropbox http://dropbox.com (last accessed 10 November 2016)

Specifically, in the course of the Go-Lab project [42], covered in Section 1.2, online platforms are employed to support blended inquiry learning. In such a framework, where teachers share their knowledge with students in blended learning sessions through a combination of face-to-face interaction and online interaction in a learning environment, adequate guidance should be provided. This task is facilitated if teachers are made aware of the progress and the difficulties of their students. Learning analytics [104] based on the analysis and the visualization of the student online interaction traces is a common means. Such analytics allows the teacher to monitor how the students are consuming the shared knowledge and intervene during the sessions if necessary.

Similarly to the educational setting, when managing the knowledge sharing inside of organizations, knowledge managers need tools to monitor the knowledge sharing process to be able to improve it to make it effective. Similarly to how analytics applied in the cases of learning is called "learning analytics", we propose to refer to analytics related to knowledge sharing as "knowledge analytics". Similarly to learning dashboards, knowledge analytics dashboards can be used to visualize interactions happening on the platform and explore them to identify possible points of improvement in knowledge sharing. Hence, we can formulate the first research question.

RQ1 - Analytics: how to provide user-oriented analytics in knowledge sharing systems to support awareness and reflection?

By requiring the analytics to be "user-oriented", we highlight the need for the analytics to be designed for usage by the end-user of a knowledge sharing platform. Such user can rely on the analytics for awareness and reflection while accomplishing her own work duties, usually not related to data analysis. This approach differs from the one providing analytics that targets business intelligence analysts or researchers, whose primary goal is data analysis. Our goal with user-oriented analytics is that the user herself can benefit from the analytics for awareness and reflection without the need of analysts.

To address the presented issue, we propose a **model** of embedded contextual and personalized analytics (**contribution 1.a**) that extends social media platforms used for knowledge sharing. Embedded analytics means that analytics is embedded into an interaction context and can be accessed by the end users while interacting with the platform. In this way, the user does not need to leave the interaction context to observe and understand the interactions happening there, fostering in this way awareness. This approach is different from existing analytics approaches offering analytics either as an external system or as a standalone analytical component in an existing system. Contextual analytics means that the data consumed by the analytics system is taken from the context where the analytics is embedded making sure that relevant information is presented. In this way, the analytics engine has access to the pertinent information from the context (people, content, interactions), and can analyze and present it to the user right in the context. By relying on the embedded contextual analytics, the user can be aware of the interactions happening in her context and by using dashboards she can explore and analyze interactions that have occurred in the past. Also, finally, personalization enables the user to adapt the analytics to present information particularly relevant to her goals.

Analytical Apps. To address the problem of teacher awareness, we have proposed to use **analytical web apps** (**contribution 1.b**) relying on the embedded contextual model. Such applications can obtain data from an interaction context in an online platform via a set of standard-based APIs. Afterward, the real-time visualizations rendered by the apps allow the user to stay aware of the activities as they happen. By combining several of such apps on a single screen, a contextual dashboard can be constructed. Such a contextual learning dashboard allows the teacher to get an immediate overview of several metrics describing the knowledge sharing process happening in a particular context (typically, a session).

Backend Analytics Engine. One of the limitations of the analytics apps running in the browser is that the data analysis code is being executed inside of the browser. In the case when there is a big number of interaction traces, conducting such data analysis in a browser can become challenging or not feasible from the computational standpoint. In the case of blended learning sessions, apps are a suitable choice since such sessions are bounded in time, usually lasting under two hours, and frequently involve less than 50 students. To support knowledge managers and provide them with tools for interactive analysis of interactions taking place in their work context, we propose to use **backend analytics engine with interactive dashboards** (**contribution 1.c**) still in line with the embedded contextual model. Differently from the apps-based approach, such engine can scale and utilize resources of multiple computers.

In the educational setting, John, the physics teacher from Section 1.2, can use such analytical tools to monitor activities of students in a blended session being aware of their performance and assist the students if needed. After the session, the teacher may use the tools to reflect on the course to improve it in the future. In humanitarian scenarios, the knowledge manager can rely on the tools to understand how different parts of the knowledge sharing process perform and what requires improvement.

We have applied and demonstrated the suitability of the embedded contextual analytics approach in the Go-Lab setting (to support awareness and reflection in blended learning) and in the MSF setting (to monitor knowledge sharing in humanitarian organizations) presented earlier on in Section 1.2. To accomplish that, we have developed a set of apps that can be used by the teachers in the teacher and the student view of an ILS. In line with the participatory design methodology described in Section 1.3, to better understand the real needs of teachers, we have conducted an evaluation of the application mockups in a summer school with 32 teachers. The evaluation helped us to identify awareness needs that were first to be addressed. Afterward, a number of apps were developed based on the mockups and their usefulness and usability were evaluated. We conducted two case studies involving the apps implemented in the Graasp platform and demonstrated their value and ease of use as reported by the teachers (**validation 1.a**). Also, the usage data collected in the Go-Lab project confirms that the users adopt the analytics apps.

Following the participatory methodology, we have started by discussing with MSF knowledge managers embedded contextual dashboard mockups to understand their needs and suitable designs. Afterward, we have implemented the proposed dashboards in Graasp. When the implementation was finished, we have conducted user studies and conducted a survey validating the usefulness and the usability of the developed solution (**validation 1.b**).

Chapter 2 explains in more details the contributions, discusses the two architectures to integrate contextual analytics into social media and presents their validations.

1.5.2 Privacy-related Contributions

For a knowledge sharing platform to provide adequate activity analytics, it is necessary to record and process digital traces that platform users leave behind when interacting with the platform. Recording, retention, and analysis of such traces have privacy aspects that platform developers and users need to take into account. The developers need to make sure that the user is aware of the tracking being conducted and provide the user with an intuitive interface to adjust the tracking policy (**challenge 2**).

When considering specifically educational settings, learning analytics (LA) is often considered as a means to improve learning and learning environments by measuring student behavior, analyzing the tracked data and acting upon the results. At the same time, the student data privacy is regulated in different ways by national governments and often depends on the content and target of the data. As an example, in the case of the Go-Lab project, student data privacy is of particular importance, since student online activity tracking is subject to stronger privacy regulations, due to the age of school pupils.

In humanitarian settings, analytics regarding the knowledge sharing can be seen as beneficial for obtaining insights. Nevertheless, due to the context where the humanitarian workers operate, exposing their interaction patterns may be a source of concerns. The interaction patterns may, among others, expose person's location or time of day when the person is not working from home. Hence, in some contexts potential risks associated with data tracking may overcome the benefits derived from the analytics and it may be reasonable to disable tracking in such cases. In addition, specific regulations linked to sensitive documents, including medical records often encountered in humanitarian settings, may influence the user tracking policy employed by the platform. The provided examples show that the sensitivity of the tracked user data may vary from one interaction context to another, and hence it is necessary to provide the user with a way to adjust the privacy policy depending on the context. Following the motivation outlined above, we formulate the second research question.

RQ2 - Privacy: what privacy management interfaces and mechanisms are suitable for knowledge analytics and learning analytics?

To address this question, this thesis proposes an intuitive contextual privacy management model, which represents a system having access to user's data as an agent located in the interaction context (**contribution 2**). We call this approach AngeLA, inspired by an agent guarding over user privacy in Learning Analytics. In a nutshell, AngeLA serves two primary purposes: (1) to make the user aware regarding the ongoing tracking and (2) to provide an intuitive and contextual way to manage the activity tracking permissions. AngeLA mimics in a virtual space the privacy control mechanism that works well in a real room: if a person is present in a room, she can observe all activities happening in the room. It is also possible to think of AngeLA as an activity recording agent that, whenever present in the interaction context (space, room, chat group or even a virtual reality room), observes and records the interactions happening in the context.

Our approach allows establishing a data access policy per context visible to all members present in the context and putting them all into the same condition regarding their data exposure. This approach defining data privacy per shared context is different from the one employed by popular sharing platforms (including Google Drive, Dropbox, Facebook), where the privacy policy is defined per user. Also, with our approach, the tracking policy is visible to all of the members so they can be aware of the data being exposed. The users are aware of third parties accessing their data in the shared context in case some of the users invited a third-party agent into the context. Differently, in the majority of popular sharing platforms the fact of such data exposure would be only known and visible to the user that has provided access to the service and another user will not be aware of that.

In the educational setting John, the physics teacher, would like to benefit from learning analytics for awareness during his class and reflection afterward. For this reason, he invites AngeLA to his inquiry learning space in Graasp so that AngeLA aggregates the data for analysis.

Using our approach in the humanitarian context, Amy, the MSF doctor, can remove the tracking agent from her workspace, where she puts medical records of her patients. Contrary, the head of her mission Nermeen may invite the tracking agent into some of the shared spaces in Graasp without sensitive data, and where he would like to get analytics regarding the knowledge sharing ongoing in his mission, so he can improve the sharing.

We have implemented the approach in Graasp knowledge sharing platform and evaluated its usefulness and usability with the platform users, which were in this case teachers. We have conducted a survey-based evaluation with teachers that demonstrated that this approach has a high system usability score and is understandable and easy to use (**validation 2.a**). Additionally, we have looked into the AngeLA usage data in Graasp, indicating that the proposed approach is being adopted (**validation 2.b**). As another proof of the proposed approach validity, the approach was selected as a way to manage privacy and implemented by two European projects, namely Go-Lab and SiWay[16].

Although the idea of AngeLA was originally developed and evaluated in an educational domain, the proposed approach is general and can be used for managing privacy in interaction contexts

[16] SiWay Project, http://www.imaile.eu/public-tenders/phase-2-selection-results/ (last accessed 10 November 2016)

where there is a notion of membership or friendship. Such contexts can be virtual workspaces, chat rooms, discussion groups, and social networks of different nature. It can be applied as well to humanitarian knowledge sharing, where traces are being recorded to provide knowledge-related interaction analytics.

Our approach to managing privacy is detailed in Chapter 3, together with its proof-of-concept implementation and validation.

1.5.3 Knowledge Discovery-related Contributions

Finding relevant content is one of the core activities of users interacting with a content repository, be it knowledge workers using an organizational knowledge sharing system at a workplace or self-regulated learners collaborating in a learning environment. Due to the number of content items stored in such repositories potentially reaching millions or more, and quickly increasing, for the user it can be challenging to find relevant content by browsing or relying on the available search engine (**challenge 3**). However, even in the case of not so big repositories, it may still be challenging to find content relevant to the user interests since the user may be unaware of its existence or specific keywords to use when searching ([96]).

In an educational setting, such content repositories are usually learning environments, where both students and teachers regularly interact with the content located there. The teachers interact when preparing a course by assembling relevant resources while the students interact when following a course by working with the materials provided by the teacher or found in the learning environment. When constructing a course in a learning environment, teachers enrich the system with relevant materials including text files, web links, videos, audio recordings coming from their device or the cloud. Other teachers can benefit from content already available in the platform when preparing their own courses. Moreover, it may also be beneficial for the students to have access to the content that is relevant to their interests, but which the teacher did not directly include into her course [117]. In the case of learning environments with a vast number of content items, it may be hard for the user to find content items corresponding to her interests.

In the context of humanitarian organizations, internal knowledge sharing systems usually aggregate knowledge produced by the organization members. In this case, whenever knowledge workers need to get information relevant to their current tasks, they may access content produced by the organization members and located in a knowledge sharing system [59]. After deploying the Graasp knowledge sharing system in the MSF setting, we expected that knowledge workers would be able to find needed content quickly since the structure of the knowledge repository of MSF followed a well-defined taxonomy of topics and the sharing system had a powerful search engine. However, it became apparent that due to the vast number of content items located in the system and their growth pace it was not always possible for the users to find the relevant content. Hence, the challenge of knowledge findability needs to be addressed. We aim to address the described challenges by investigating the third research question formulated as follows.

RQ3 - Discovery: how to enable discovery of knowledge relevant to user interests?

To address the mentioned issues, we propose a knowledge discovery model employing an interactive recommender system that combines the content analytics, activity tracking, and information retrieval techniques. The system is employed to (1) build a user interests profile (**contribution 3.a**) and afterward (2) to suggest content relevant to the user and users with similar interests (**contribution 3.b**) enabling knowledge discovery. To perform the recommendation, first, for each item available in the content repository, we employ natural language processing techniques to identify a set of concepts related to the content in a similar way how humans would do it. Relying on high-level concepts instead of specific words present in the text when constructing user interests profile and afterward finding similar items, allows to identify the content that covers the same high-level concepts even if the specific words used in it are different. Next, we analyze the interactions of the users with the content items based on available user activity recordings and aggregate the concepts in the content that the user interacted with building in this way the user interest profile. Finally, we use information retrieval techniques to recommend to the user relevant content based on the similarity between the concepts in the content and concepts identified as user interests. In the same way, our approach allows finding relevant users based on the determined interests similarity. Our approach puts the user in control of her interests profile and allows to adjust the interests by removing some of them if necessary, as in the case when the user is not interested at the moment in some of the identified interests.

Such a knowledge discovery approach can be beneficial for John, the physics teacher from Section 1.2. For instance, when preparing a lesson about radioactivity, he can get recommendations of the content relevant to the topic that was uploaded earlier by his peer teachers. As well, the approach allows John to discover some new materials related to other topics he is interested in, so he can get acquainted with new educational resource and possible make use of them when needed.

Benefiting from the proposed approach, Amy, the doctor from MSF treating sleeping sickness, can discover content related to her mission that was created by her colleagues from other missions but was not explicitly included by the HQ into her workspace. Also, Amy may check the recommended content from time to time to see if any new information relevant to her interests appears at MSF.

To evaluate the usefulness of the proposed approach and the performance of the algorithm, we have integrated it into Graasp, a knowledge sharing system introduced earlier in Section 1.1. Afterward, we have conducted two evaluations: 1) a preliminary one with teachers that allowed us to identify potential areas for recommender system improvements early on (**validation 3.a**) and 2) an extended one with long-term Graasp users (**validation 3.b**). The evaluations have demonstrated that our approach allows identifying user's interests and provides the users with relevant recommendations. In addition, we have recorded and analyzed user-recommender interactions in Graasp (**validation 3.c**). The obtained data demonstrates that the recommender is being adopted.

While we draw our experience and motivation from the educational and humanitarian settings, our knowledge discovery contributions have a broad impact and apply to content repositories, whenever it is possible to obtain content analytics and track activities performed by the users (e.g., Google Drive and Dropbox).

1.5.4 Knowledge Delivery-related Contributions

With the growing popularity of the cloud, more and more educational platforms become cloud-based and require Internet access to be able to interact with them. At the same time, current educational resources incorporate an increasing amount of more bandwidth-heavy types of multimedia content, including videos, high-resolutions images and interactive books.

From our experience with European schools in the framework of the Go-Lab project, in the case when the Internet connection is available at a school, the content loading time can increase dramatically when multimedia content is being consumed by a whole class using multiple computers. In schools with limited or slow Internet access, such educational multimedia resources become unusable due to their loading time. Moreover, when the Internet access is not available or interrupted, the educational resources located in the cloud can not be accessed at all. Such cloud-centric approach to education delivery is of particular concern when we consider that at the moment of writing in 2016, at least four billion people still do not have reliable Internet access, so they cannot benefit from the knowledge shared online, including online educational resources. This disparity in online education access possibilities may lead to increase in the educational gap between the developed countries with a good Internet connection and less developed ones with limited Internet access.

If we consider the situation from a humanitarian perspective, the places where humanitarian action is needed are often those where Internet connectivity is limited or not available at all due to poor or damaged infrastructure. Recent prominent examples of such connectivity situations include the Nepal Earthquake in 2015 and the Ebola outbreak in 2013-2015. At the same time, timely access to critical information is of particular importance for any humanitarian mission operating in such situations of emergency (**challenge 4**). As a specific example, in the missions where MSF operates the Internet connection is absent or limited. Still, to be efficient in these settings MSF employees need to have access to relevant and up-to-date knowledge. This need makes essential the deployment of information systems matching requirements of a specific context where humanitarian organizations operate. Given the described situation, we formulate the fourth research question:

RQ4 - Delivery: how to facilitate knowledge delivery into underconnected settings where Internet connectivity is limited or absent?

Two ongoing trends in technology development make new approaches to knowledge delivery possible. The first trend is the affordability and hence ubiquity of mobile device (phones and tablets). Mobile devices can already be found in rural areas, even those without an Internet connection and their proliferation continues as prices drop. The second trend is the rapid development and affordability of small general-purpose computing devices, primarily single-board computers.

This thesis builds on these two trends and proposes a data delivery model that integrates social media with a peer-to-peer middleware and relies on low-cost computers for local knowledge replication (**contribution 4**). Thanks to a peer-to-peer protocol, a peer can find other local peers and synchronize data directly with them without the need of the central server, potentially benefiting from a faster local network. In this scenario, the data does not need to go through the central server (usually located remotely with limited bandwidth) as it is done in other mainstream cloud-based platforms. To minimize the bandwidth consumption and reduce latency, we propose to use local replicating peers running on compact inexpensive (and low-energy-consuming) devices allowing to ensure local data availability even when the Internet connection is absent. Such devices, when connected to a battery, can provide an autonomous repository with the knowledge wherever it can be needed. By using their mobile devices, users can log in to the local peer via WiFi. Because of the local peer availability, the data can be served from it even when the central server is not reachable. Software built following our model can enable knowledge delivery into challenging environments.

Following the proposed approach, it is possible to make learning resources available in a school located in a remote area with limited Internet access or relevant humanitarian knowledge in a mission with an absent connection. This way, John, the physics teacher from our scenario in Section 1.2 can take such a device with him into a rural area together with mobile devices (if needed) and conduct teaching sessions even without an Internet connection. In another case, when John teachers at his school with limited bandwidth, he can use the device to reduce the content loading time for his students, thus avoiding waiting during his classes. Whenever he has a good Internet connection and updates content on the online platform, the content gets automatically synchronized to local devices as soon as they get again connected to the Internet.

The approach as well allows an HQ of a humanitarian organization to deliver and keep up to date knowledge in their missions. In the proposed way, the knowledge can be made accessible by field workers even when the mission does not have a persistent Internet connection or when the work is needed to be done in an area located remotely from the mission office. In the scenario from Section 1.2, a field doctor Amy can have access in the field to all of the documents relevant to her mission that were prepared by the headquarters. By accessing the content locally, Amy can quickly go through multimedia content (for instance, videos) and at the same time having minimal (possibly even zero) bandwidth footprint.

We have integrated the proposed model into the knowledge sharing platform Graasp. Afterward, we have conducted synthetic tests to observe how the system performs in typical scenarios explained earlier and verified that the approach fulfills the delivery requirements (**validation 4.a**). Later on, we have evaluated the system in an authentic setting by enabling local content access in MSF missions in Maputo, Mozambique and Kampala, Uganda (**validation 4.b**). While we implemented the architecture in Graasp, other social media platforms can as well benefit from integrating the architecture in order to provide offline access to the content.

In Chapter 5 we discuss the contribution, present its implementation and validation.

1.6 Summary and Organization of the Thesis

Summing up, the primary goal of this research is to investigate how to enhance social media to support better knowledge sharing in two particular domains, namely the educational and humanitarian ones. Targeting this topic, we focused on the following four research questions:

- **RQ1 - Analytics**: how to provide user-oriented analytics in knowledge sharing systems to support awareness and reflection?
- **RQ2 - Privacy**: what privacy management interfaces and mechanisms are suitable for knowledge analytics and learning analytics?
- **RQ3 - Discovery**: how to enable discovery of knowledge relevant to user interests?
- **RQ4 - Delivery**: how to facilitate knowledge delivery into settings where Internet connectivity is limited or absent?

The structure of this thesis is schematically presented in Figure 1.1. First of all, in Chapter 2, we introduce the embedded contextual analytics model and present two general architectures implementing the model. Chapter 3 proposes an agent-based privacy management model and demonstrates our implementation of the model called AngeLA. Chapter 4 focuses on knowledge discovery and we introduce our approach relying on an interactive recommender system based on user interests. Chapter 5 proposes a novel content delivery model employing a peer-to-peer technology and low-cost computers and covers our implementation of the model called GraaspBox. Finally, Chapter 6 concludes with the summary of the contributions presented in this thesis, and possible directions for future work.

[17] CC BY 4.0 http://creativecommons.org/licenses/by/4.0/ (last accessed 10 November 2016)

How to enhance social media platforms to support knowledge sharing in educational and humanitarian domains?						
Knowledge Sharing System Structure	**Knowledge Sharing System Requirements**	**Research Questions Targeted**	**Contributions**	**Validations**	**Primary Setting**	**Chapter Number**
1 Knowledge Aggregation	**Req 1**: Support Source-agnostic Knowledge Aggregation	Not in focus of this thesis				
2 Knowledge Organization	**Req 2**: Enable Decentralized Knowledge Organization	Not in focus of this thesis				
	Req 3: Provide Knowledge Analytics	**RQ 1**: How to provide user-oriented analytics in knowledge sharing systems to support awareness and reflection?	**Contribution 1**: An embedded contextual analytics model and two general analytics architectures materializing the model.	**Validation 1**: Surveys with Graasp users, usage data, participatory design.	Educational Humanitarian	**Chapter 2**
	Req 4: Handle User Data Privacy Properly	**RQ 2**: What privacy management interfaces and mechanisms are suitable for knowledge analytics and learning analytics?	**Contribution 2**: An agent-based privacy management model.	**Validation 2**: A survey with Graasp users, usage data, participatory design.	Educational	**Chapter 3**
3 Knowledge Access	**Req 5**: Facilitate Online Knowledge Access	**RQ 3**: How to enable discovery of knowledge relevant to user interests?	**Contribution 3**: A knowledge discovery model based on interactive interests-based recommender system.	**Validation 3**: Two surveys with Graasp users, usage data.	Educational	**Chapter 4**
	Req 6: Facilitate Offline Knowledge Access	**RQ 4**: How to facilitate knowledge delivery into settings where Internet connection is limited or absent?	**Contribution 4**: A novel mobile knowledge delivery model employing peer-to-peer middleware and low-cost computers.	**Validation 4**: Tests in the laboratory, field studies.	Humanitarian	**Chapter 5**

Table 1.1 – A schematic structure of the thesis showing the relation between the research questions, contributions, validations and chapter numbers.

2 Embedded Contextual Analytics for Awareness and Reflection

One way to promote effective knowledge sharing on social media platforms is to provide the involved parties with relevant and timely analytics concerning the sharing process. Such analytics allows the users to be aware, and reflect on the knowledge sharing process, and consequently revise and adapt their participation if needed. To meet this need, in this Chapter, we propose the concept of embedded contextual analytics where the analytics is delivered (embedded) into the interaction context and presents (contextual) information relevant to the end-user in that particular context. Such analytics allows the user to stay informed about activities happening in a particular context without the need to leave it. Also, we propose two general architectures to materialize this concept, the first one using analytical web applications and the second relying on backend analytics engine and interactive dashboards. The two approaches allow integrating embedded contextual analytics into social media platforms. In this Chapter, we as well describe how we integrated these architectures into Graasp. Finally, we present an evaluation with the target users of the embedded contextual analytics tools following the proposed approach. This chapter targets the first research question:

RQ1 - Analytics: how to provide user-oriented analytics in knowledge sharing systems to support awareness and reflection?

The content of this chapter was partially published in [113, 110]:

A. Vozniuk, M. J. Rodriguez Triana, A. C. Holzer, S. Govaerts and D. Sandoz et al. Contextual Learning Analytics Apps to Create Awareness in Blended Inquiry Learning. 14th International Conference on Information Technology Based Higher Education and Training, Lisbon, Portugal, June 11-13, 2015.

A. Vozniuk, S. Govaerts and D. Gillet. Towards portable learning analytics dashboards. International Conference on Advanced Learning Technologies (ICALT), Beijing, China, July 15-18, 2013.

2.1 Introduction

Nowadays we observe a change in how knowledge is shared for learning purposes, as education becomes more distributed and flexible [36, 40]. Participating in courses remotely by using massive open online courses (MOOCs) is a common practice, whether still being physically present in an educational institution or studying remotely from home. At the same time, when learners study remotely, they tend to get less involved in the process and eventually loose motivation. Consequently, MOOCs experience a much higher dropout rate, compared to regular classes [62].

Another approach to learning that became more popular recently is blended learning, where different types of learning (online and offline, inside and outside of the classroom) are combined (blended) together to provide a richer and potentially more engaging learning experience. While numerous types of blended learning modalities exist [94], in this Chapter we focus on blended learning, where face-to-face and computer-mediated types of learning are combined. One type of such blended learning scheme relying on online labs to enhance classroom activities is considered as a promising approach to increase the skills and the interest of students in science, technology, engineering, and mathematics [25, 27].

The emerging distributed and flexible nature of the learning process creates new challenges. It becomes much harder for the teachers and the students to be aware, reflect on, and, in the case of teachers, to orchestrate the learning process (**challenge 1**). For instance, in the MOOCs, where thousands of students are simultaneously taking an online class, it is not feasible for a teacher to consider individual competencies and preferences of each learner. Similarly, to enable teachers to orchestrate a blended learning session (like the ones in Go-Lab explained in Section 1.1) and to provide adequate guidance to their students, it is critical that the teachers are aware of the progress and difficulties of their students. This task is not trivial for both large and small classes, as in such a blended learning framework part of the work is done online making it hard for the teacher to observe. Since the remote and blended learning part of the learning activities are computer-mediated, it is possible to record detailed student activity traces and to provide the teacher with dedicated learning analytics to improve awareness.

Learning analytics (LA) aims to address the highlighted issues by collecting and using traces of learners, including those that they leave online [30]. According to Siemens et al. [105], LA is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs". Learning analytics based on the real-time analysis of the student online interaction traces and their real-time visualization are considered an adequate mean to facilitate awareness [80]. Teachers and learners can benefit differently from LA. For instance, a teacher could observe live statistics about what and how learners are doing to coach them. For example, an LA-based notification system can predict which students are about to drop out [49] and notify the teacher. Then again, motivation to learn might improve if students are able to compare their actual individual progress with the expected progress [54] or the progress of their peers. In general, LA aims to help us to understand better how we learn and improve our learning [30].

Various online learning management systems (LMSs) and personalized learning environments (PLEs) [38] exist that support learning. Examples include Blackboard [16], Moodle [28] and Graasp [9]. Such systems often provide LA dashboards as a visual way to deliver LA to the user. Generally speaking, the dashboard is a visual display of the most important information needed to achieve one or more learning objectives, consolidated and arranged on a single screen so the information can be monitored at a glance [35]. Specifically, a learning dashboard is a single display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualizations [100]. Duval et al. proposed to use LA dashboards for representing important information about learners [30]. Support for such dashboards could, for instance, increase the use of LMSs and PLEs for MOOCs to provide an overview to learners and teachers.

When delivering learning dashboards to their target users, an open question remains how this delivery should be done. Existing LA solutions can be divided into two broad groups when it comes to how they are delivered to the end-user. The first group consists of completely standalone solutions that are not integrated into the learning platform but just use data recorded by it. The second group is the LA solutions integrated in some way into the platform where the user is working. When the dashboard is not integrated into the learning environment and is a standalone tool or a platform, the user needs to switch between the learning platform and the tool to see analytics. When the integration is done as a dedicated analytics section of the environment, the user still needs to switch between the context where she works and where analytics is located. This may deteriorate awareness and complicate the orchestration process and reduce the motivation to use the tools. Similarly, if the dashboard is not focused on presenting information relevant to the particular situation (context) the teacher or the student is dealing with, it could be hard to gain insights while looking at the dashboard.

Targeting the highlighted issues, we propose a model to deliver user-oriented analytics by embedding it into the learning context and presenting relevant contextual information. Analytics following this idea can be accessed directly in the interaction context and present relevant information scoped by this context and the learning objectives. As an example, the tools can show how a student progresses compared to other students or to the expected time, so she knows if she should complete the tasks faster. It can show how much time students spend on individual parts of the task so the teacher can see if the planned allocated time and the actual required time are aligned. By being aware of the situation and being able to analyze it, the users can adapt their behavior accordingly.

This chapter is structured as follows. First, Section 2.2 discusses related work and then Section 2.3 introduces the embedded contextual analytics model. Next, Section 2.4 presents a general architecture to build and integrate such analytics into social media platforms using analytical web applications. Section 2.5 discusses the evaluation of this proposal and presents the case studies done in a classroom involving the apps. Similarly to the previous two sections, Section 2.6 describes another architecture materializing our model by using backend analytics engine, and Section 2.7 describes its validation. Section 2.8 compares the two proposed architectures highlighting advantages and limitations of each of them. Finally, Section 2.10 wraps up with conclusions and future work.

2.2 Related Work

In order for a knowledge sharing system to provide analytics, the system needs to capture and store user-platform interaction traces and to allow analytics tools to access the recorded traces. In this section, we start by reviewing some of the existing approaches to capturing user activities on the platform. Afterward, we take a look at some representative examples of learning analytics solutions for awareness and reflection focusing on the way they are integrated with a learning environment and on the information they present.

2.2.1 Capturing Activities

The massive uptake of social media platforms has fostered the adoption of social media features in learning environments, e.g. social networks for collaboration and blogs are integrated into many online learning environments. Such social features provide a rich data source for analytics (e.g. interaction analysis). According to object-oriented modeling, the data that can be used by analytical tools represents dynamic (behavioral) and static (structural) views of the platform [50]. The dynamic view focuses on the dynamics of the processes ongoing in the platform including the collaboration between users, tools and learning resources and changes of their internal states. On the other hand, the structural view describes the static structure of the learning platform including the users, tools and learning resources and their relations [50]. If appropriate interfaces to access data are available, analytics can exploit both views.

Capturing data about user actions is a common problem for various social platforms (including learning environments). Such platforms track user actions for various reasons, e.g. (i) to increase user awareness of their activities and those of peers on the platform, or (ii) to enable personalization techniques (e.g. personalize search result ranking or advertisements). User actions on multiple platforms could be aggregated to provide a full picture of the user's activities. To achieve this, all platforms should exchange user action data.

User activities tracked by a learning platform is a common data source in the field of learning analytics [94, 97]. Usually, a learning management system or a learning environment have a logging infrastructure in place that records how the user interacts with the platform [94]. The more modern educational platforms support a structured representation of user activities using well-defined formats including ActivityStreams used in [110], xAPI employed in [60] or IMS Caliper outlined in [97]. On a high-level, all these three formats record user-platform interactions in the format of the actor-verb-object triplet capturing who did what with what on the platform. However, on a more detailed level, each format captures additional aspects of the interaction. In the triplet, the verb indicates the type of interaction, for instance, the verb "accessed" would indicate that the user viewed content, "downloaded" - downloaded the content and so on. Having a common set of verbs with a well-defined meaning is critical for being able to benefit from user interactions captured by several platforms [60]. Another format worth noticing is Contextualized Attention Metadata (CAM) [19], but it is being used less these days. Below, we provide more details about the formats.

RSS. Currently, many social media platforms still provide a data feed of activities of the user and her social network in RSS format. For instance, every Facebook page has an associated RSS feed[1]. The RSS format was developed to publish frequently updated media, such as blog entries, comments or videos. As such it is not suitable for capturing all rich metadata related to user activities.

CAM. One of the specifications used for educational purposes is CAM [115]. Contextualized Attention Metadata (CAM) aims to capture contextual information regarding the activity of a user [115]. CAM is very flexible in what it can track, which also limits its portability. CAM has no common vocabulary to describe actions and allows each platform to define its custom vocabulary, which makes it difficult to develop fully portable dashboards on top of the CAM specification. Besides, the CAM specification has not been widely adopted outside research so far.

ActivityStreams. In contrast to CAM, Activity Streams has a defined set of verbs with a corresponding meaning to describe actions[2]. This verb definition enables better interoperability across platforms, which leads to better portability. The ActivityStreams specification[3] is designed specifically to model user interactions and enables platforms to share detailed information on user activities. An activity stream is a sequence of actions made by a user. Such a stream in the ActivityStreams specification models the story of a person performing an action on or with an object. Technically in the ActivityStreams specification, an action is represented as a 5-tuple: (*Published*, *Actor*, *Verb*, *Object*, *Target*). *Published* represents the time at which the activity was published. *Actor* defines the actor that performs the action. *Verb* describes which action is performed. The specification defines an extensible set of verbs. *Target* is intended to describe the consequences of the action. Consider this example action: "Alice added a book by Tom Sawyer to her 'Favorite Books' collection". In this example, "Alice" is an *Actor*, "added" a *Verb*, "a book by Tom Sawyer" an *Object* and "'Favorite Books' collection" is a *Target* of the action.

ActivityStreams had a large uptake and was supported by most social media platforms (e.g. Google+[4] and Facebook[5]). Many organizations have contributed to the development of

[1] Facebook RSS feed http://ahrengot.com/tutorials/facebook-rss-feed/ (last accessed 10 November 2016)
[2] http://activitystrea.ms/specs/json/schema/activity-schema.html (last accessed 10 November 2016)
[3] http://activitystrea.ms/specs/json/1.0/ (last accessed 10 November 2016)
[4] https://developers.google.com/+/api/latest/activities (last accessed 10 November 2016)
[5] http://developers.facebook.com/blog/post/225/ (last accessed 10 November 2016)

the specification[6] and made it an open standard through the Open Web Foundation Final Specification Agreement[7] that allows the use, extension, and sale to anyone.

Due to the broad definition and extensibility[8] of the ActivityStreams specification, it can be applied to the learning domain [77]. ActivityStreams can be used to represent interactions of learners and teachers with the learning platform and is thus an appropriate specification to represent activity data for LA. Man et. al. also propose ActivityStreams as a standard for exchanging user activities among learning platforms [77]. Because of the described properties, ActivityStream is a suitable format we utilize to represent user activities in Graasp.

xAPI & IMS Caiper. Despite the flexibility of ActivityStreams two noticeable standards have emerged in the learning community recently tailored to represent learner's activities, namely xAPI[9] (or Experience API) and IMS Caliper[10]. At the moment of writing, xAPI becomes a de facto standard for capturing learner's activities. On the abstract level, the two standards record interactions in the format of the actor-verb-object triplet similarly to the ActivityStreams format. Each of the standards captures additional parameters of the learning process allowing to record a more detailed picture[11]. ActivityStreams is used in Graasp for historical reasons since the decision regarding the format was made at the beginning of the Go-Lab project in 2012 when xAPI was not mainstream yet. Currently, Graasp offers a mapping for activities stored in ActivityStreams format so they can be retrieved in the xAPI format if necessary.

2.2.2 Delivering Analytics

The research literature provides a number of examples of learning analytics tools developed to improve awareness of teachers in a classroom [108]. In our previous works we have done systematic reviews of the state-of-the-art of learning dashboards (in [100]) and tools for monitoring, awareness, and reflection (in [94]). In terms of how the tools are delivered to the user in an online learning environment, existing solutions can be split into two broad categories: (1) standalone solutions that get data from the learning platform, but are in fact dedicated separate systems and (2) integrated solutions that are integrated into the learning platform. The integrated solutions can be in turn arranged into three groups by the extent to which they integrate with the learning platform: (2.1) dedicated plugins or modules; (2.2) analytics embedded into the interaction context and (2.3) analytics embedded into the interaction context and presenting contextual information. Below, we review some of the existing solutions from each of these classes.

[6] http://wiki.activitystrea.ms/w/page/24500522/Implementors (last accessed 10 November 2016)
[7] http://openwebfoundation.org/legal/the-owf-1-0-agreements/owfa-1-0 (last accessed 10 November 2016)
[8] http://activitystrea.ms/specs/json/1.0/
[9] Experience API https://en.wikipedia.org/wiki/Experience_API_(Tin_Can_API) (last accessed 10 November 2016)
[10] IMS Caliper https://www.imsglobal.org/activity/caliperram (last accessed 10 November 2016)
[11] xAPI Caliper Comparison https://www.imsglobal.org/initial-xapicaliper-comparison (last accessed 10 November 2016)

Standalone Analytics Solutions

The standalone analytics solutions employed in practice can be divided into two broad groups: (i) general web analytics services (e.g. Google Analytics[12] and Woopra[13]) and (ii) services developed specifically for learning analytics (e.g. the CAM web service [86]).

One representative example of the second group is work by Charleer et al. that proposed to employ a learning analytics dashboard as a course-long awareness tool for teachers and students [20]. During the course, the student's interact using Twitter and blogging platforms and their activity traces are being recorded. A set of badges is used as an indicator of the student progress towards the course goals and the level of student activity in the course. The dashboard allows to track the progress of individual students and compare the progress among the students. Since teachers can be overwhelmed by manually observing each of the student activities, such a dashboard can help them to remain aware of the student progress just by checking if the students have received all expected badges at a given time. If a more detailed view of the progress is needed, teachers can look into individual actions of the students.

Web Analytics. Another option of building the tools could be to use general-purpose web analytics platforms (e.g. Google Analytics[14] or KISSMetrics[15]). They are easy to deploy and maintain, but they come with a number of shortcomings. First of all, storing student activity traces on third-party services raises privacy concerns since privacy-sensitive tracking data of learners are stored on third-party servers. Some platforms such as Google Analytics explicitly forbid in their service license agreement to send to the platform any data explicitly revealing the user identity. And due to privacy policies and laws, data of learners can often not leave the institution or the country. When no identity information is stored, it is not possible to track the performance of individual students. Additionally, web analytics tools such as Google Analytics or Piwik[16] usually aggregate and display the data for the whole web platform and provide information which is too generic. Also, such analytics tools are generally not accessible to the platform end-users and are often limited to the IT staff. In the case of the Go-Lab teachers or MSF knowledge managers, they are more interested to see analytics for their individual spaces or documents with the possibility of viewing data about their individual users, students or knowledge workers respectfully. Retrieving such information from web analytics tools is not always possible and when possible often requires writing custom queries. Finally, web analytics tools usually do not provide expressive semantics for the learning or knowledge context, i.e. they are able to identify web page visits made by students, but not if a student has changed an ILS phase or submitted an assignment. Another limitation of web analytics services is related to their general scope, where domain-specific information is not taken into account. For instance, just tracking that a learner visited a page, is not sufficient to deduce the activity she was involved in, e.g. reading, watching a video or playing a game. Such web

[12] http://www.google.com/analytics/ (last accessed 10 November 2016)
[13] http://www.woopra.com/ (last accessed 10 November 2016)
[14] http://www.google.com/analytics/ (last accessed 10 November 2016)
[15] https://www.kissmetrics.com/ (last accessed 10 November 2016)
[16] Piwik https://piwik.org (last accessed 10 November 2016)

analytics tools are not able to track information specific to learning.

Standalone LA solutions developed specifically with LA in mind also face issues. They are strongly coupled with a specific learning environment and thus lack re-usability and flexibility [48, 95].

Integrated Analytics

Dedicated Analytics Modules. Pluggable LA solutions are integrated into a platform via a plugin interface (e.g. in Moodle). Such a plugin interface is often not compatible across platforms. If a user switches to another platform, then she needs an LA plugin for that platform. Through the use of plugins, learning dashboards are often integrated into learning environments that lack LA. Moodog is such a plugin [118] for Moodle and other LA plugins exist to track learners using blogs [34]. Also, general-purpose analytics tools are often used, although they are not tailored towards learning. For instance, Google Analytics [7] is used to observe resource usage by students. Most existing LA solutions are either not specifically designed for learning (e.g. Google Analytics) or are tightly coupled to a learning platform (e.g. Moodog). It usually requires a considerable effort, most often involving software engineering, to adapt a learning dashboard or deploy it on another platform. This limits the adoption of LA tools.

Another example of integrated analytics is work by Maldonado et al. that presented in [76] a set of real-time visualizations that help teachers to be aware of the interaction patterns in tabletop-based group work. The visualizations are based on tabletop touch data and verbal data recorded by microphones. The authors suggest visualizing presented interaction on two levels: (1) class level, aggregating data from interactions around several tabletops in the classroom (useful to identify anomalies during a classroom session) and (2) detailed group level helping to detect possible abnormalities in a single group (useful for after-class analysis). A user study showed that such visualization tools are useful for the teacher and help her to identify the groups and individual students that require attention and intervene when necessary. The study also demonstrated that the teachers were able to detect the main anomalies in group interactions based on the class level while the detailed group level was used to confirm the assumed issues.

2.2.3 Discussion

Our goal is to answer RQ1 by providing analytics to the platform user in a way that integrates into her overall experience of interacting with the platform. The aspects of the delivery of analytics tools to the user get limited attention in existing literature. The research papers often stop at the point where a tool is developed and demonstrated to be working for specific scenario or experiment, but not many proposals are mentioning how analytics should be integrated into the overall user-platform interaction experience. This can also be caused by the fact that the majority of the learning analytics solutions were not evaluated in the authentic settings [94]. Another limitation of the reviewed proposals is the lack of general analytics architectures that can be used when implementing analytics on a particular platform.

Regarding what is being analyzed and visualized by analytics tools, existing solutions often present by default analysis of all of the interactions going on the platform. Such approach can be useful when targeting researchers or analysts that want to have a general overview of the activity and obtain high-level metrics. At the same time for the teacher and the students, it can make more sense to have analytical tools or even dashboards representing activities happening in particular scope, for instance, a class or a lab session. We aim to address the mentioned issues with our proposal presented in the next section.

2.3 Embedded Contextual Analytics

To answer RQ1 and provide user-oriented analytics supporting awareness and reflection of the platform users, we propose the model of embedded contextual analytics (**contribution 1.a**) schematically represented in Figure 2.1. We call the analytics "embedded" since such analytics is embedded into the context where the user interacts when completing her tasks. And we use the term "contextual" since analytics is aware of the context it is embedded into. Hence it analyzes and visualizes the data scoped by this specific context. For instance, such analytics may display information for a session scoped in time by its duration or for a particular team scoped by a shared space.

> *Embedded Analytics.* We call analytics *embedded* into the space s if and only if it is part of the space presentation. In case the analytics is constructed with analytical apps i_a, that implies $i_a \in I(s)$.
>
> *Contextual Analytics.* We call analytics s *contextual* in the space s if all of the activities that the analytics retrieves (denoted by A_r) belong to space s. Formally, this means $\forall a : a \in A_r \Rightarrow a \in A(s)$.

The analytics following our model is user-oriented since the primary objective of its design is to support the platform user. Since the analytics is located in the context where the user interacts, the user can check analytics while completing her tasks, hence awareness is facilitated. Also, since the user does not need to leave that context in order to see analytics, thus avoiding context switching. In case the context contains multiple members, they can have a consistent picture of the interaction by using the same analytics tools embedded in the shared context. Since the analytics is contextual, it immediately presents information relevant to that particular context where interaction happens enabling reflection, without the need of information prefiltering. Such prefiltering is usually required to remove irrelevant information when analytics solutions are targeting the whole platform.

As mentioned, the embedded contextual learning analytics tools and dashboards retrieve

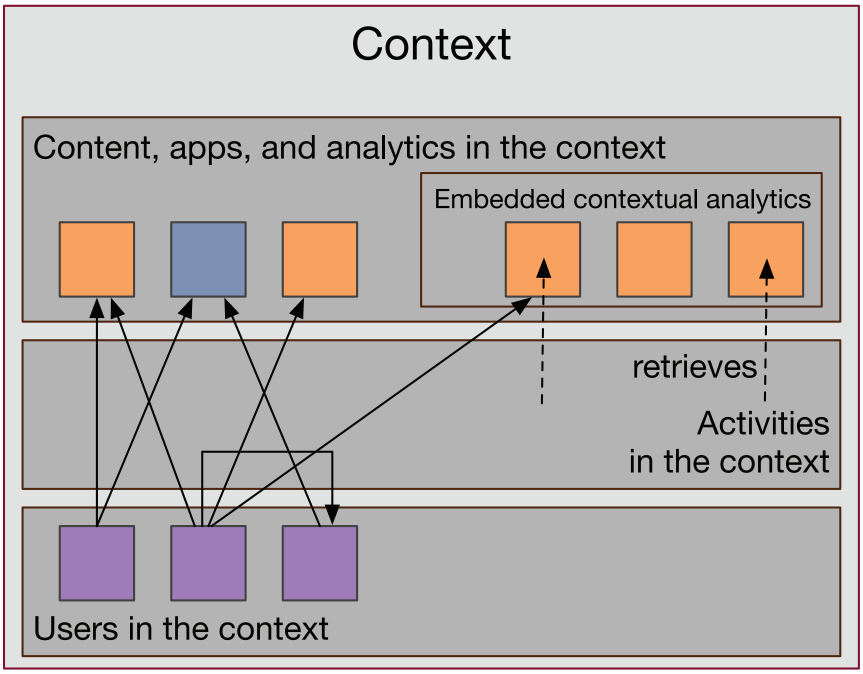


Figure 2.1 – The proposed embedded contextual analytics model. The analytics is embedded and accessible in the context where the user interacts with content and apps when completing her tasks. The analytics retrieves and analyzes the activity traces from the context.

data for analysis and visualization depending on the current context. To give an example, if the interaction context is a space about English learning, the traces from this space will be visualized. If the space concerns a blended learning session about sustainable development, corresponding student actions from this session will be visualized. This allows the analytics tools or dashboards to be developed once and be used in multiple contexts to obtain relevant context-specific information.

Embedded contextual analytics as well allows taking a look at a different granularity of activities in case the context are organized hierarchically. For instance, in case an interaction space for a team is part of the interaction space of a department that is part of the interaction space of organization. In such a case, the same contextual dashboard embedded in these interaction spaces will get data specific to the context that is a part of the organizational hierarchy.

In the following sections, we demonstrate how the general model of embedded contextual analytics can be materialized in online learning environments and propose two general analytics architectures to achieve that.

2.4 Embedded Contextual Analytics With Web Apps

In this section, we first describe a proposal of an apps-based analytical architecture and then show how this architecture was used to build tools supporting teachers in blended learning sessions.

2.4.1 Architecture & Components

We propose as **contribution 1.b** an architecture that relies on analytical web apps to provide embedded contextual analytics to the platform end-users in the form of dashboards. The proposed architecture is presented in Figure 2.2) and is composed of:

1. *Front-end* consists of the knowledge sharing platform UI and analytical apps. The apps retrieve, analyze and visualize activities and allow the user to view and interact with the analysis results.
2. *Back-end* provides access to the contextual data stored on the platform via a set of Standardized APIs. There are two aspects of the data representing the system model: the 'structural' (often referred to as static) and 'behavioral' (often referred to as dynamic, or timing) aspects of the model [50]. It is important that the APIs provide allow to retrieve data for both of these aspects since they can be used for analysis.
3. *Analytics Infrastructure* stores in the Data Storage the activity and social data that the apps retrieve and use in analysis.

Using analytical apps for analytical dashboards has several benefits. First, dashboards can be constructed using diverse apps that display visualizations and metrics. These apps can be arranged on a single screen to provide all information at a glance [35]. An example of such sample dashboard built with eight apps is schematically presented in Figure 2.3. Second, as such apps are programmed using just Javascript and HTML, the apps simply run in a browser, which provides an operating system-independent solution. Javascript and HTML have good support on mobile devices as no additional plugins are required in contrast to Flash or Java solutions. Additionally, the platform can also control access to the data that is accessible through a set of standard-based APIs with attached privacy policies. For instance, teachers can observe trends over the whole class, while students might only see and reflect on their own actions.

When we organize several analytical as a dashboard, such an assembled dashboard could be packaged as a web app in itself. In this case, a dashboard is a web app containing other web apps [8]. If a dashboard is bundled in this way, the whole assembled set of web apps become portable allowing teachers and learners to share assembled analytical dashboards across platforms.

2.4.2 Properties

Below, we discuss some of the properties concerning awareness and reflection enabled by analytical web apps.

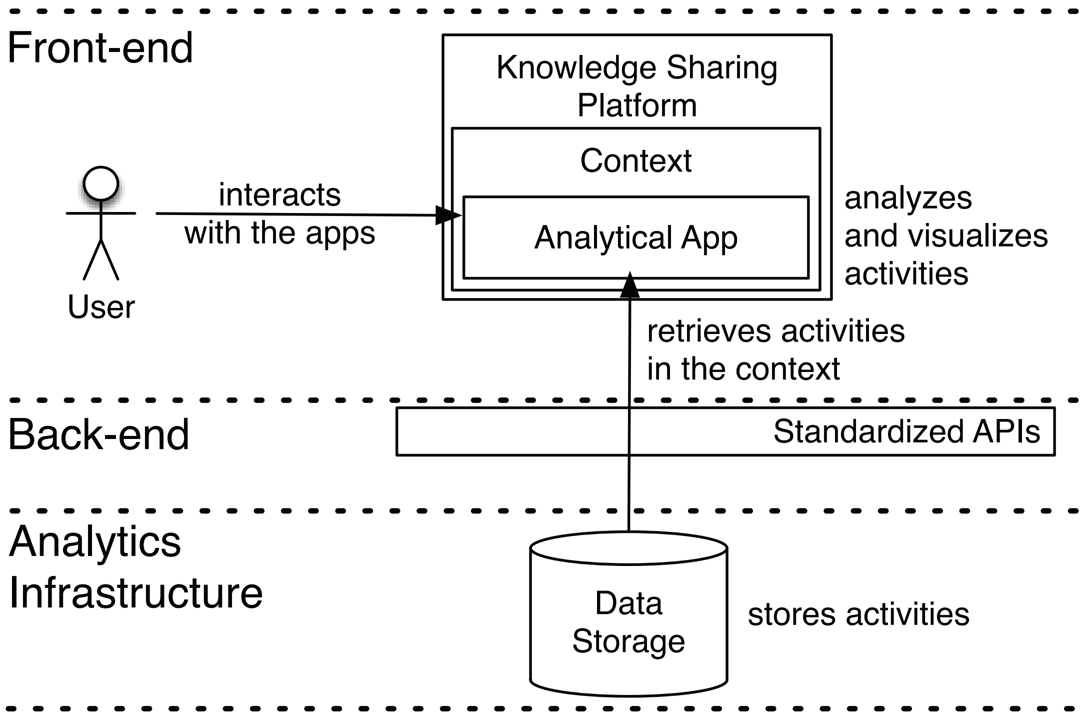


Figure 2.2 – The proposed architecture for the analytics web apps integration with the platform. The analytical apps are presented to the user in the context of her activities and retrieve contextual activities from the platform via a set of standardized APIs.

Real-Time Updates. When analytical apps load, they request the activities and the data available at that point of time. Afterward, the updates can be pushed to the apps through as they happen. This makes it possible for the apps to provide real-time visualization of the interaction and make users aware of the current state.

Personalization. Different users have different awareness and reflection needs, so personalization of analytical tools and dashboards is essential. Many of available learning analytics solutions provide a predefined hard-coded dashboard while in different pedagogical scenarios different tools may be needed. Thanks to its modularity, our approach with the analytical web apps enables personalization. A teacher can select from the repository specific analytics apps she finds to be useful in her session and build a dashboard from these apps. In this way, every teacher in every session has a possibility to personalize the dashboard for her needs in the session.

Interactiveness. The apps-based approach allows developing interactive analytical applications using web technologies such as d3.js[17]. Inter-widget communication can be used for letting widgets know on the status of each another to support cross-filtering of the visualized data [12].

[17]Data-Driven Documents https://d3js.org/ (last accessed 10 November 2016)

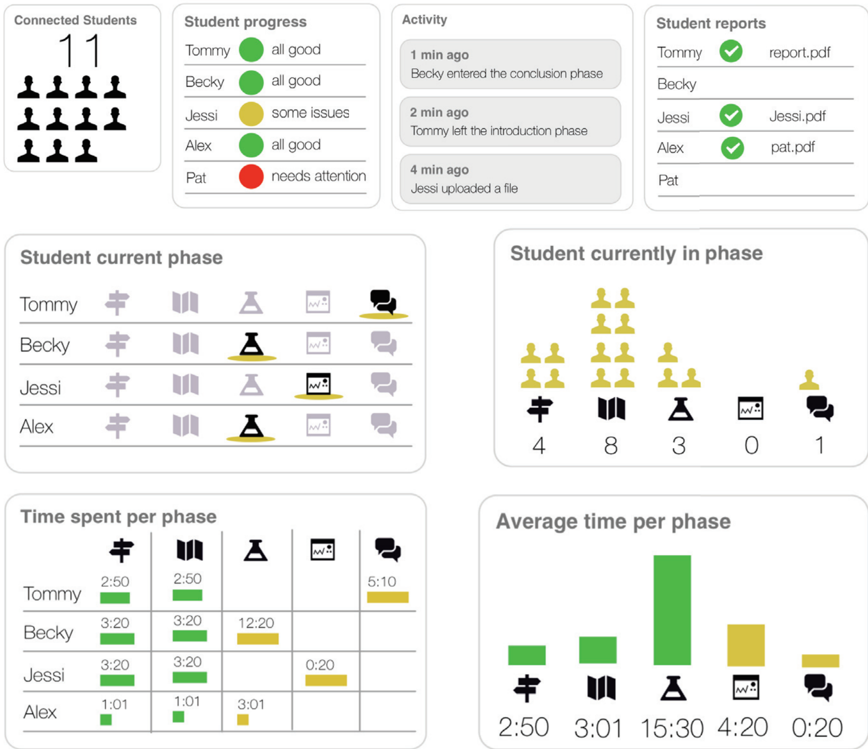


Figure 2.3 – A schematic representation of a dashboard constructed using eight analytical apps.

Portability. Software portability is the ability to run the same software on different platforms with no or little effort [83]. In the previous section, we discussed the interface specifications to exchange data between platforms and applications. To enable portability of analytics dashboards, it should be possible to simply take analytical tools from one platform and run them on another platform. Some configuration of analytical tools on a new platform might still be required, but its cost is considerably lower than developing the tools from scratch. An LA component interacts with a learning platform through a collection of interfaces. To facilitate portability, all interfaces for an analytics component have to be made identical across platforms [83]. Ideally, we would like such common interfaces to be already implemented by all learning platforms. Otherwise, an adaptation of the interfaces or dashboards will be required during the porting process [83]. Our approach with analytical web apps allows achieving portability by implementing learning dashboards as external pluggable components and using well-defined interfaces with a learning platform. The specification of these interfaces should be based on existing open standards. In this way, we can ensure that our solution is license-free and extensible. Also, using open standards increases compatibility with existing and future platforms and tools, because open standards often have a large user base that enables wider adoption and future support and extensions.

Data Storage and Privacy. With the apps-based approach, the data is stored on the platform and is accessible by the apps. In this case, the data does not need to leave the platform. This approach is suitable for the situation when the platform is considered trustworthy.

Reusing and Repurposing the Apps. Since the apps are developed using web technologies, their code is open and can be easily accessed and modified when needed. To encourage the sharing and reuse of the analytics apps, a repository can be developed. It can be used to publish new analytical apps and search for existing apps. Since awareness needs of the teachers can differ in different contexts (ILSs), they can use existing apps[18] to assemble dashboards to personalize them according to their specific needs in each of the contexts.

Apps Repository. For easier distribution, LA dashboards and apps can be added to a repository where teachers and learners can find and easily add them to their learning environment. We have presented several apps built following our proposed architecture. We presented several analytical apps, but more apps are available in the apps repository developed by third parties. For instance, in the case of the Go-Lab project such apps can be found on the Golabz repository[19].

2.5 Validation of Embedded Contextual Analytics With Web Apps

In this section, we show how the general architecture with analytical web apps described in the previous section was integrated into social media. We as well describe how it was deployed and used to enable awareness and reflection in blended learning sessions. After each of the sessions, we have conducted a survey to get a quantified feedback from the users about the usefulness and usability of the proposed apps. Also, we report the usage data showing how the apps are used in the Go-Lab project.

2.5.1 Integration into Graasp

To evaluate our approach, we have integrated the proposed architecture into Graasp as shown in Figure 2.4. This integration allows teachers that use Graasp to construct embedded contextual LA dashboards from analytical apps. Figure 2.5 shows an example of LA dashboard in Graasp. In Graasp, the context is defined by space (introduced in Section 1.1), hence our goal is to enable Graasp users to add and see analytical apps and dashboards embedded into the space presenting information from the space. For instance, in a space for learning English, only the data of English learners would be shown in the LA dashboard. Through spaces, we can track learners and their context, and we can provide detailed LA scoped by that context only.

To ease distribution and search of LA dashboards and analytical apps, Graasp has a transparent integration with the Golabz repository. It is possible to search and add analytic apps from the repository into a space in Graasp without leaving the platform.

[18]Go-Lab Apps https://github.com/go-lab/apps (last accessed 10 November 2016)
[19]Golabz learning analytics apps http://golabz.eu/apps?f[0]=field_app_category:7485 (last accessed 10 Novem-

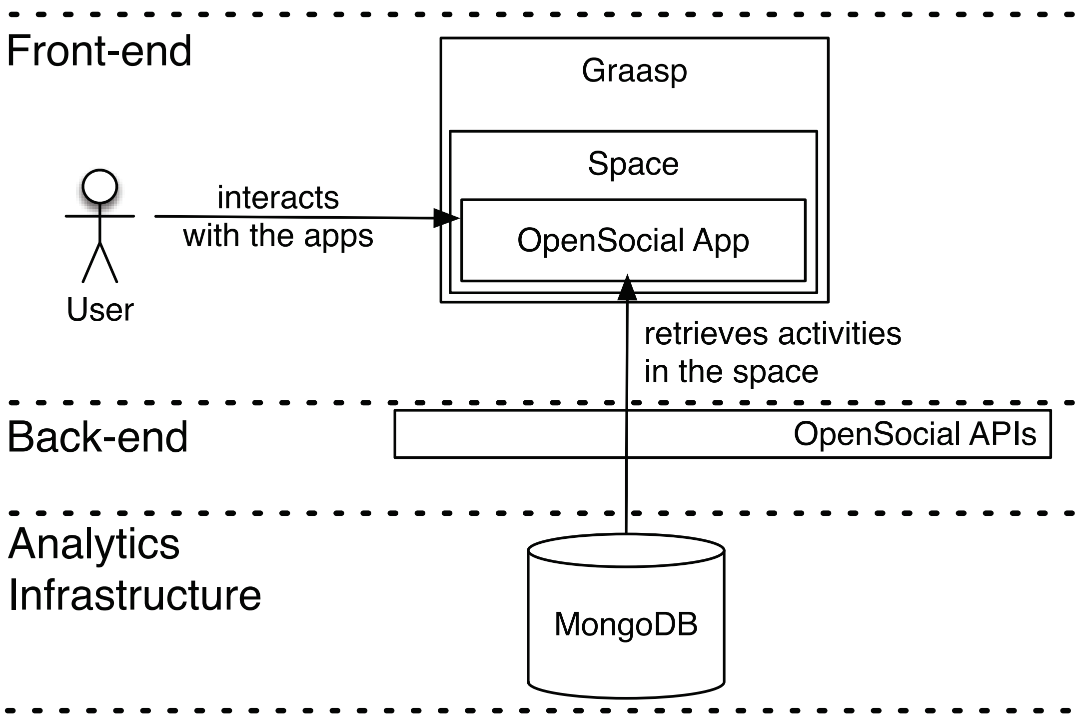


Figure 2.4 – The proposed apps-based architecture integrated into Graasp. The analytical apps are OpenSocial apps located in the space, retrieve activities for the space via OpenSocial APIs.

Capturing Activities

Each user (student and teacher) using the learning platform produces a stream of activities reflecting their interaction with the platform. Figure 2.6 (1) shows an example stream of user activities in a social media platform. Due to its expressive format, we employ the ActivityStreams specification to represent each of the actions. Figure 2.6 (2) demonstrates an example of the user activity represented in the ActivityStreams json format.

To capture activities, we implemented, in accordance with the ActivityStreams specification, recording of multiple interaction types in Graasp related to users, content, applications and spaces, including: *add*, *update*, *remove*, *invite*, *join*, *leave*, *favorite*, *comment* and *tag*. Context, where the interaction happens, can be captured through the *Target* field while tracking user activities.

Retrieving Data via OpenSocial APIs

The analytical apps need to access the contextual information to function. There are many ways to implement the APIs to provide this data access. We decided to go with OpenSocial,

ber 2016)

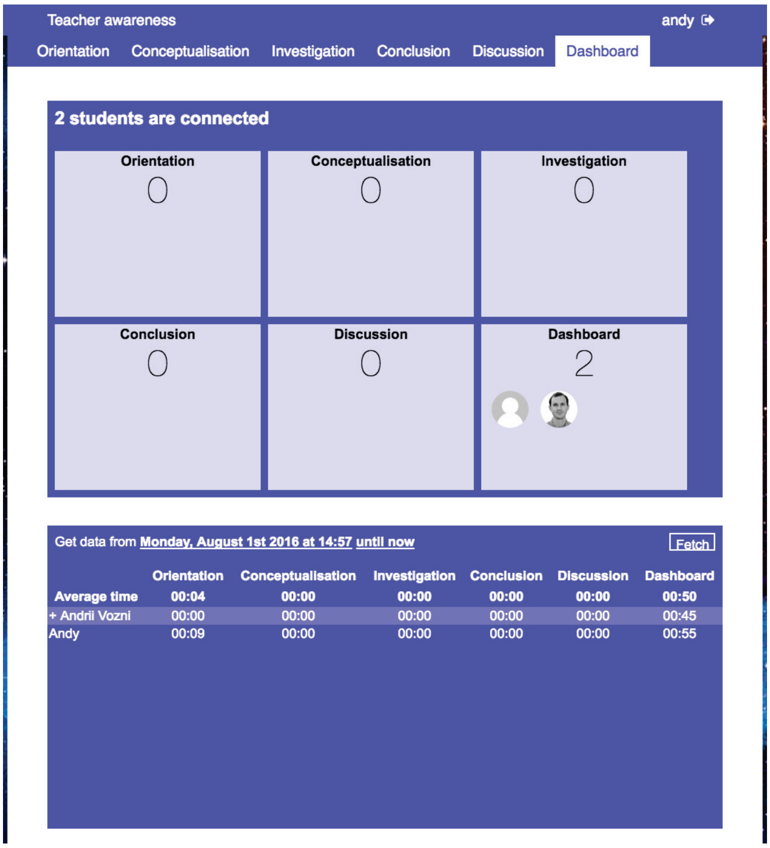


Figure 2.5 – A sample embedded contextual learning dashboard in Graasp containing the Active Users app (on the top) and the Time Spent app (in the bottom).

as it is a standard used to provide access to data in social media platforms. Relying on a well-defined data access standard has benefits of making developed tools portable across the platforms implementing the standard.

The OpenSocial specification[20] aims to provide a standard method for storing and accessing social data. Additionally, OpenSocial defines a component hosting environment (a container) and a set of common APIs for web applications. The standardization of the interfaces to store and retrieve data enables the development of components, which are portable across systems. The OpenSocial 2.0 specification incorporates an API for accessing user activities using the ActivityStreams format. OpenSocial 2.0 provides APIs for accessing both the dynamic and structural views of a social or learning platform. At some point, OpenSocial was adopted by many popular social media platforms, including Ning[21], MySpace and Orkut[22]. Originally OpenSocial was focused on social network applications, but thanks to educational interest in social media, OpenSocial became more popular in technology enhanced learning (TEL).

[20] http://docs.opensocial.org/display/OSD/Specs (last accessed 10 November 2016)
[21] developer.ning.com/docs/opensocial/1.0/index.html (last accessed 10 November 2016)
[22] https://developers.google.com/orkut/articles/tutorial/tutorial (last accessed 10 November 2016)

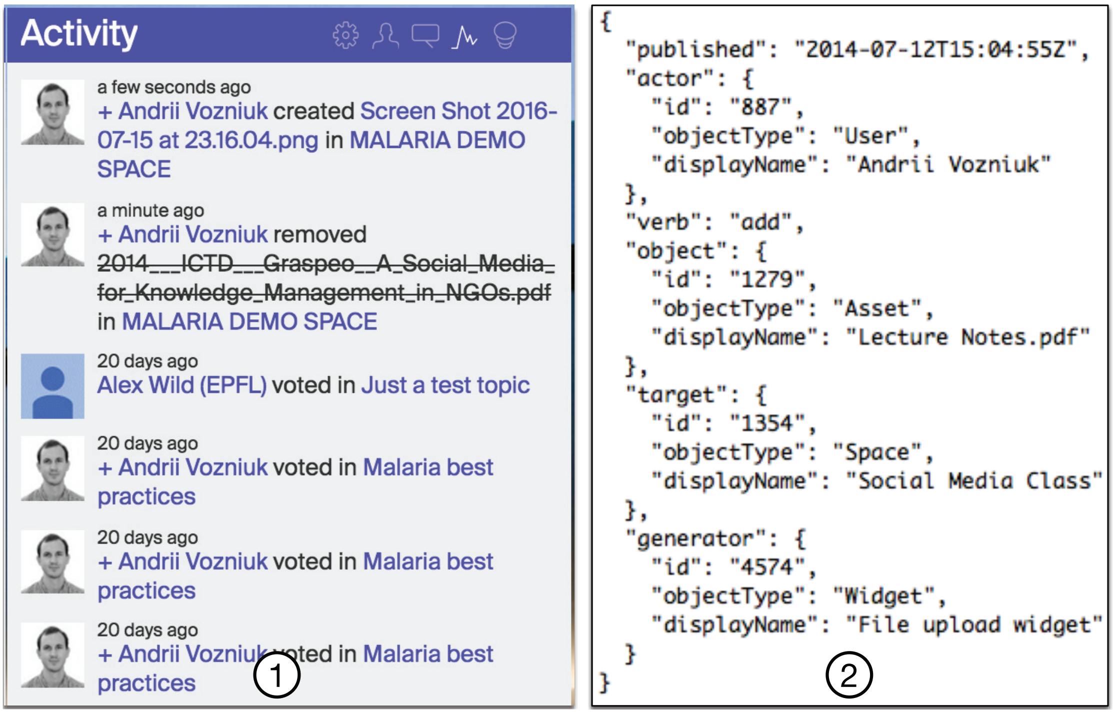


Figure 2.6 – (1) A sample stream of user activities in a social media platform. (2) A sample user activity represented in the ActivityStreams json format.

Notable TEL examples include Sciverse[23], Liferay[24], and ROLE [43].

Through the Space extension[25] of OpenSocial that is implemented in Graasp, LA can be made context-aware. Thus, a space can provide LA for a specific context. The analytical apps present in such a space are shared among the members of the space [11] (see Figure 2.7). By being context-aware (i.e. related activities are carried out in a given online workplace), roles and privacy settings can be enforced that enable control over the selected LA data in the dashboard. Analytics dashboards assembled with OpenSocial apps can run on any platform that implements the OpenSocial specification.

Graasp supports OpenSocial apps through the Apache Shindig container[26]. Shindig is a free open-source widget container with a reference implementation of the OpenSocial and ActivityStreams specification. By integrating Shindig into the Graasp platform, apps have read and write access to Graasp data through the OpenSocial APIs (e.g. user information or activity streams). This enables the creation of LA dashboards since apps can access user activities and social data for analysis or visualization through these APIs.

[23] http://surfnet.nl/en/nieuws/Pages/ConstructionstartsofOpenSocialenvironment (last accessed 10 November 2016)

[24] https://dev.liferay.com/develop/tutorials/-/knowledge_base/6-1/creating-and-integrating-with-opensocial-gadgets (last accessed 10 November 2016)

[25] https://opensocial.atlassian.net/wiki/display/OSD/Space+Proposal (last accessed 10 November 2016)

[26] http://shindig.apache.org (last accessed 10 November 2016)

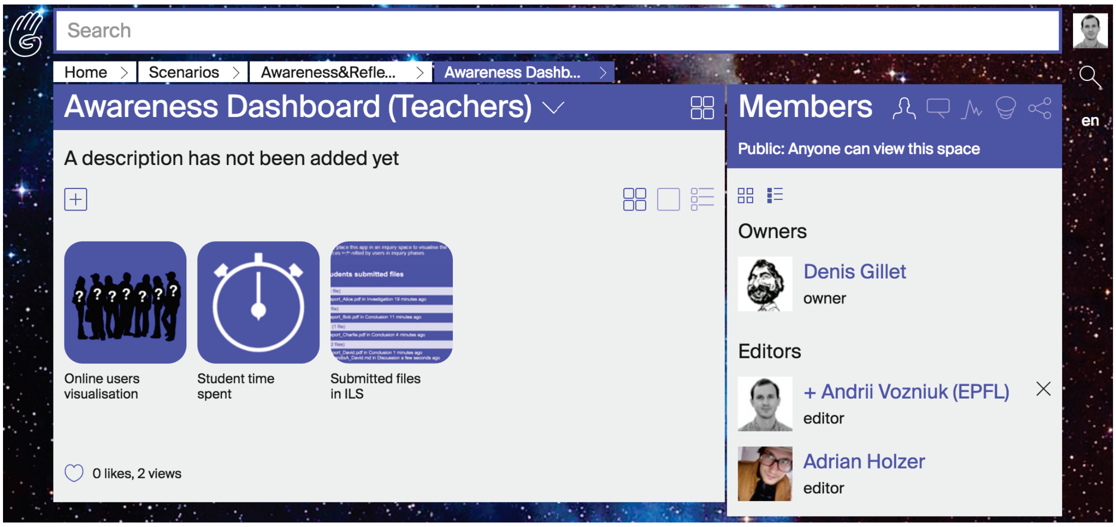


Figure 2.7 – A space in Graasp containing three analytical apps.

2.5.2 Developed Analytical Apps

In this section, we present the embedded contextual analytical apps we have built using the apps-based architecture to support blended learning sessions. We also verify if such tools are useful for the teachers to improve their awareness. To start with, we tried to understand teacher needs for such apps as discussed below.

Understanding Analytics Needs of Teachers

To understand the awareness information needs of the teachers, we interviewed 32 teachers during the 2014 Go-Lab Summer School asking them which kind of learning analytics application would be helpful during their blended learning sessions. Teachers were asked to what extent they agreed with the following statement: “I would use this app in class”. Also, we asked the teachers about the relevance of having an overview screen (dashboard). We used 7-point Likert scale ranging from -3 (strongly disagree) to +3 (strongly agree) to collect the responses. Afterward, we calculated the average (mean) of the score per app (presented in Table 2.1); the box plot visualization the survey data is shown in Figure 2.8.

According to the results, the required apps that received the highest score were: 1) submission by the student, 2) average time per phase, 3) connected students and 4) connected students per phase. The figure also demonstrates that the teachers were eager to have and use a single overview screen (dashboard) containing several apps at a time.

Table 2.1 – Results of the awareness needs survey.

2*App	Total No. answers	No. of answers per point -3	-2	-1	0	1	2	3	Average score
Submission by student	29	0	0	0	0	0	13	16	**2.55**
Average time per phase	29	0	0	0	2	0	10	17	**2.45**
Connected students	32	0	0	1	0	1	11	19	**2.45**
Students by phase	31	0	0	0	0	2	14	15	**2.40**
Phase for each student	30	2	0	0	1	2	11	14	2.01
Activity log	31	2	1	0	1	3	12	12	1.81
Phase changes per student	27	2	6	1	4	3	6	5	0.43
Overview screen	30	0	1	0	0	1	7	21	**2.53**

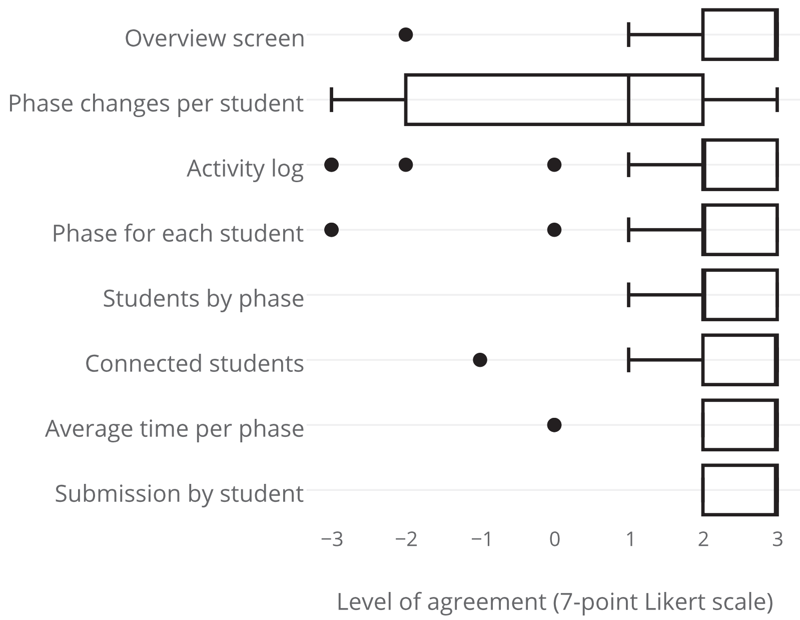


Figure 2.8 – A boxplot representing the results of the awareness needs survey.

Developed Apps

Below we discuss in more detail three contextual analytics apps we implemented to address the identified awareness needs. Along with the apps, we present a typical scenario in which the apps can be used in a blended session.

Active Users App. Figure 2.9 shows the *Demo* ILS with the *Active Users* app displaying the students currently active in each phase of the ILS. It is possible to see in the figure that the majority of the students are in the Investigation and the Conceptualisation phases (16 and 9 respectively) of an inquiry learning space, there are 2 students active in the Orientation phase and 1 student in the Conclusion phase. The app updates in real time as students switch the phases, so the teacher knows where the students are working right now.

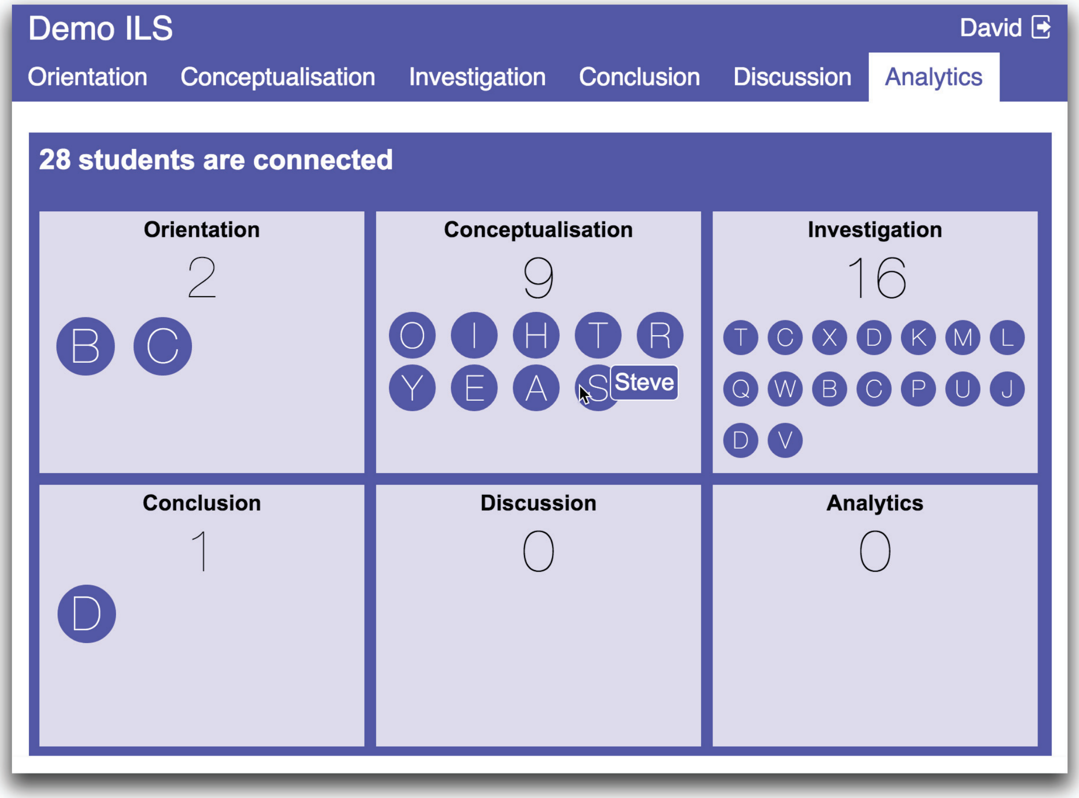


Figure 2.9 – The *Active Users* app displaying the students active in the ILS.

Time Spent App. Figure 2.10 shows the *Demo* ILS with the *Time Spent* app displaying individual and average time spent by students in each ILS phase. When a student is active in one specific phase, the teacher can see the timer running for this student in that phase in real time. Additionally, the average time spent by students in each phase is given, so it is possible to compare individual performance of students to average performance of the group. There is also the possibility to analyze a specific period of time, for instance, if the teacher wants to perform a post-analysis of the session.

Submitted Reports App. Figure 2.11 shows the *Demo* ILS with the *Submitted Reports* app displaying the files submitted by students in the ILS and the time when it was done. It is updated in real time as well. Therefore, every time a new report file is uploaded in any phase of the ILS, a new line appears in the interface. This app is particularly suitable to support blended learning scenarios, where the teacher wants to be aware of the submitted reports as the submissions happen.

Dashboard Templates. The Golabz repository provides the teacher with a set of apps that can be used inside ILS. When the teacher wants to view several visualizations at a time, Graasp allows her to organize several LA apps on a single analytics dashboard. An example of such dashboard containing two apps is presented in Figure 2.5. To facilitate the integration of LA

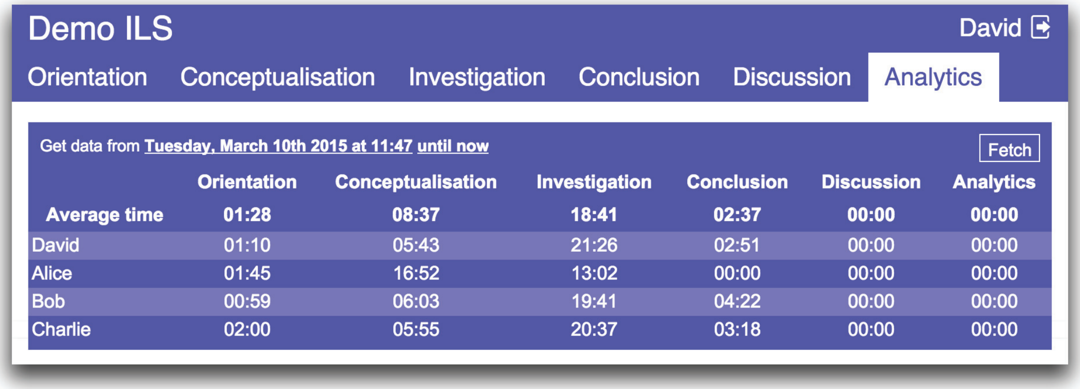


Figure 2.10 – The *Time Spent* displaying the time spent by students in the ILS.

Figure 2.11 – The *Submitted Reports* app displaying the submitted reports.

apps into the teacher workflow, some of the ILS already contain preconfigured dashboard templates with the tools that may be useful for a particular scenario. Still, the teacher is in control and can add other apps and remove the provided apps in case it is needed.

Scenario

Below we describe a typical scenario how the apps can be used before, during, and after a blended learning session.

Before the session the teacher adds the *Active Users*, *Time Spent* app *Submitted Reports* into the ILS.

At the beginning of the session she looks at the *Active Users* app and is able to see the students appear in the Orientation phase as they open the ILS on their computers. *During the session*, as the students complete the Orientation phase, they switch to the Conceptualisation phase, and the teacher is able to see it in the *Active Users* app shown in the example in Figure 2.9. With this app, the teacher spots the students that are still in the Orientation phase. The teacher then goes to the students and helps them accordingly. The teacher is using the *Time Spent* app and observes that almost all the students have spent between 4 and 6 minutes in the Conceptualisation phase as expected. *In the end of the session* the students submit a report in the Conclusion phase. As the submissions happen, the reports appear in the *Submitted Reports* app, which gives an overview of who has submitted as demonstrated in Figure 2.11. There are few minutes left till the end of the session, and the app indicated that 26 students have already submitted files. The teacher identifies who are the 2 students who still need to submit their files and goes to them directly to see if there is an issue.

After the session, the teacher uses the *Time Spent* app to evaluate how much time was spent on average by students in each phase. She notices that in average the students have spent almost 19 minutes in the Investigation phase, which is too much. This led the students to rush through the Discussion phase. The teacher will, therefore, make the necessary changes in the description of the Investigation phase so the next time students will complete it in less time.

2.5.3 Case Study with Analytical Apps

We have conducted an evaluation of the usability and the usefulness of the implemented apps for teachers (**validation 1.a**). Two case studies using the *Active Users*, the *Time Spent* and the *Submitted Reports* apps were carried out in Geneva between January and March 2015. The first study took place at the Ecole de Commerce Nicolas-Bouvier[27] involving 1 teacher (T1) and 11 students (between 18 and 20 years old). The second study was conducted at the College Sismondi[28] where 1 teacher (T2) and 17 students (between 15 and 16 years old) participated in the experiment. During these studies, we had two main objectives: (1) to check if such contextual real-time visualizations improve teacher's awareness and (2) to examine if the apps are understandable and easy to use.

To achieve these objectives we used three data sources: (1) first, we interviewed the teachers before the experiments to evaluate whether the information provided by the apps could be relevant for them and for which purpose; (2) then, one researcher attended the classroom to observe how the apps were used during the lessons and collected comments of students about the apps; (3) finally, we gathered additional remarks from the teachers in another interview once they had used the apps.

In the first interview, the teachers were asked about the extent to which they find each of the awareness apps useful for teachers and students. The main outcomes of the interview are summarized in Figure 2.12, showing the teachers' point of view using a 7-point Likert scale

[27] Ecole de Commerce Nicolas-Bouvier http://icp.ge.ch/po/nicolas-bouvier (last accessed 10 November 2016)
[28] College Sismondi http://www.sismondi.ch/ (last accessed 10 November 2016)

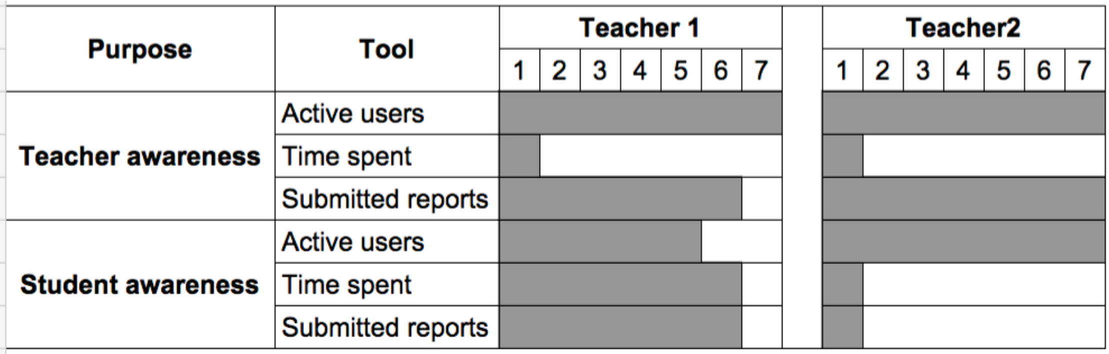


Figure 2.12 – The relevance of the apps according to the teacher interviews. The 7-point Likert scale was used ranging from 1 (strongly disagree) to 7 (strongly agree). The gray area indicates the median response value.

ranging from 1 (strongly disagree) to 7 (strongly agree). Dealing with the *Active Users* app, both T1 and T2 agreed on the usefulness of the app for awareness of the learning progress for teachers and students. The *Time Spent* app was not seen as significantly useful for awareness. However, both teachers agreed that it could be relevant for better understanding the students' progress. Concerning the potential convenience for students, both teachers came up to the conclusion that the app could help learners to be aware of their peers' progress. Regarding the *Submitted Reports* app, the teachers considered that they would use it for their own awareness, but they diverged about the benefits that the app could bring to the students.

In both studies, the teachers designed an ILS to be used by the students during a 90-minute session. In the first study, the students were allowed to work individually or in groups, and the teacher was walking around and answering the questions that emerged during the learning activity. Since the teacher was not near her computer, she decided to display the apps using the projector, so teachers and students could see the visualizations as presented in Figure 2.13. After answering the questions of the students, the teacher had a look at the student's distribution across inquiry phases and used that information to choose the next group to visit. At the same time, the students periodically observed the apps to compare their own progress with that of their peers.

In the second study, the students worked in groups of 2 to 3 people sharing one computer, while the teacher controlled the situation from his desk, going to the students just when he had doubts (see Figure 2.14). T2 used mainly the *Active Users* app to monitor whether the students were using the ILS or not, and to be aware of the current phase they were working on. Since, in this case, the information was not shown to the students, at the end of the lesson they were asked about whether they would like to have access to the apps or not. The responses were heterogeneous: 7 students did not have a clear preference, 4 were highly interested in the apps, while another 6 students were reluctant. For those who were interested in the apps, the main benefit was to have a reference regarding their progress. The main disadvantages identified by the students were 1) the distraction the apps may cause and 2) the stress caused by being compared among themselves. Nevertheless, they did not pose any concern about the teacher accessing the information.

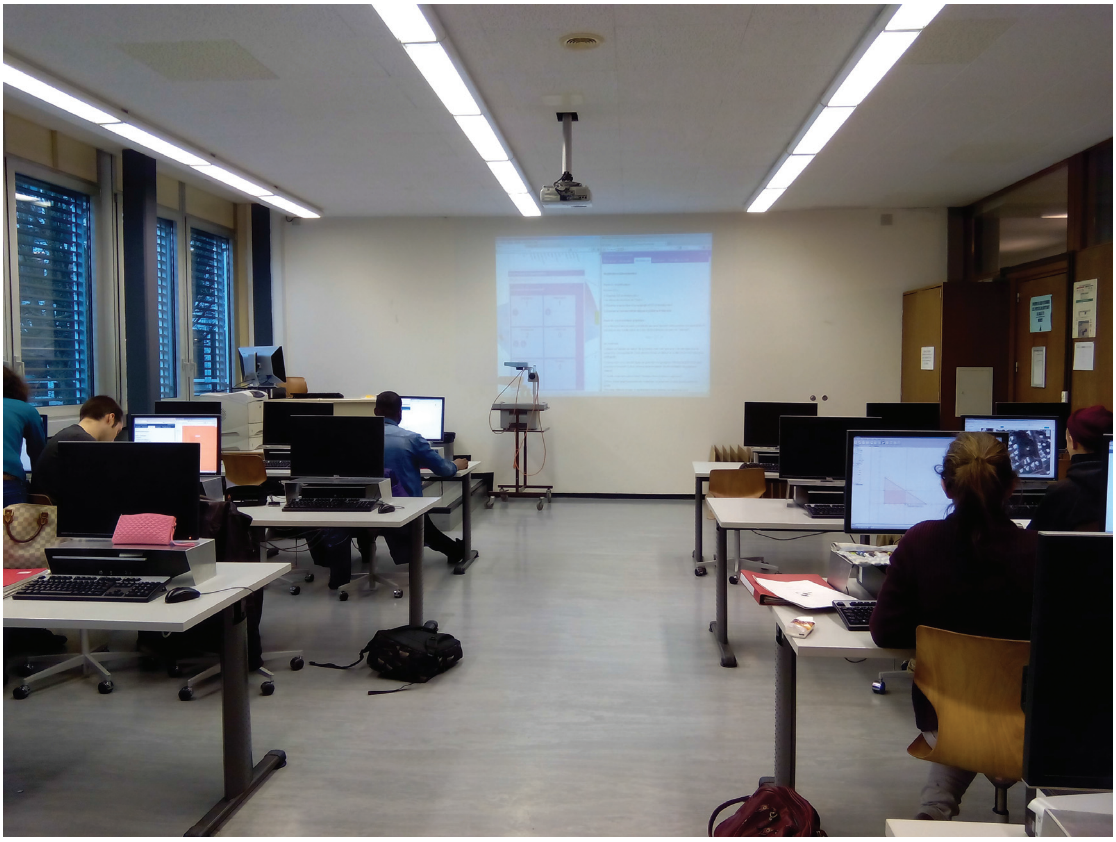


Figure 2.13 – An example of the *Active Users* app used by the teacher and the students for awareness. The *Online Students* app is projected on the wall so the teacher can see at which stage the students are working.

According to the feedback provided by the teachers, in general terms, it was straightforward to understand the visualization provided by the apps. Just one aspect required additional explanation for the teachers about the *Active Users* app: the colors used to show the time spent. For T1, the only inconvenience that she detected was related to the way the users were represented. Since she was not near her computer, the mapping between the student initials (used in the avatar) and the nicknames was not obvious. Besides, the different colors employed to show the time spent were not distinguished when projected on the wall. Nevertheless, both issues could be simply solved by using a legend explaining the colors and the correspondence between initial and nickname and choosing different colors, shapes or effects. Regarding the *Time Spent* app, both teachers agreed that it could be more useful for post-session reflection, especially since during the session normally they do not have enough time for these details.

In summary, the usability and applicability of the real-time awareness tools were well received by the teachers. They stated that the tools helped them to monitor the progress of the students in the classroom, and they could be as well used to have evidence of the work done at home.

Figure 2.14 – Second study: use of the *Active Users* app for teacher awareness.

2.5.4 Usage Data

We wanted to understand if the developed apps are employed by teachers to support awareness and reflection in practice. For this reason, we checked the Graasp database to see how many of the created inquiry learning spaces (ILSs) contain the developed apps. It is worth noting that ILSs can implement various learning scenarios not necessarily requiring these particular apps but still the numbers can indicate the general usefulness of the apps.

We have run the queries in the Graasp database on September 7, 2016. On this date, there were in total 13575 ILs in Graasp. Out of them, *Time Spent* app was present in 989 (7%) ILSs, the *Active Users* app was in 802 (6%) ILSs, and the *Submitted Reports* app was found in 129 (1%) ILSs. These numbers show that the teachers are eager to include the apps into their ILSs and indicate that the apps are useful to support their awareness and reflection.

2.5.5 Discussion

In this section, we demonstrated how our general app-based analytics architecture proposed in Section 2.4 was able to answer RQ1 for Graasp users when integrated into the platform. With its help, we facilitated teacher and student awareness in blended learning sessions. The teachers were able to personalize analytical dashboards by assembling the apps that matched their learning scenarios.

The blended learning sessions in Go-Lab are usually short-term, typically under two hours. Also, they involve a modest number of students that can fit into a single classroom; that is in practice under two hundred students. Under these conditions, the amount of traces generated during the session is moderate representing rather “small data”. Also, simple types of analysis and visualization employed by the teachers and students can be handled by the web apps running in a browser.

At the same time, the educational data continues to grow in volume, velocity, and variety. Making sense of and exploring patterns in the educational data in such conditions requires deployment and usage of appropriate scalable, interactive, real-time processing tools supporting interactive data analysis. In the case when monitoring needs to be done during longer periods of time with a larger number of learners (as in the case of MOOCs) and potentially capturing more details or requiring a more sophisticated data analysis, the scalability limitations of the previous approach become apparent. In the next section, we propose an architecture that meets the additional scalability and interactivity requirements.

2.6 Embedded Contextual Analytics with Analytics Engine and Interactive Dashboards

The scalability limitations of the apps-based approach come from the fact that the analysis is executed in the browser and so it is constrained by the browser capabilities. To enable scalable analytics, the analysis needs to be executed on the back-end, where it is possible to leverage the computational capabilities of a server or the cloud.

To answer RQ1 and to enable user awareness and interactive reflection in settings where large datasets need to be analyzed, we propose an analytics architecture composed of a scalable analytics engine and an interactive dashboard (**contribution 1.c**). The architecture is presented in Figure 2.15. In this architecture, the user accesses the dashboard that allows interactive exploration of the data. The dashboard is embedded into the interaction context, but differently, from the apps-based approach, it does not execute the analysis. Instead, it retrieves the analysis results from the back-end. When such request arrives at the back-end of the platform, a permission check is performed whether the user can access analytics. If it is the case, the request is forwarded (proxied) to the Analytics Engine that serves as a data storage and performs analysis according to the request from the dashboard. After the analysis is executed, the result is sent back to the dashboard where it is displayed to the user.

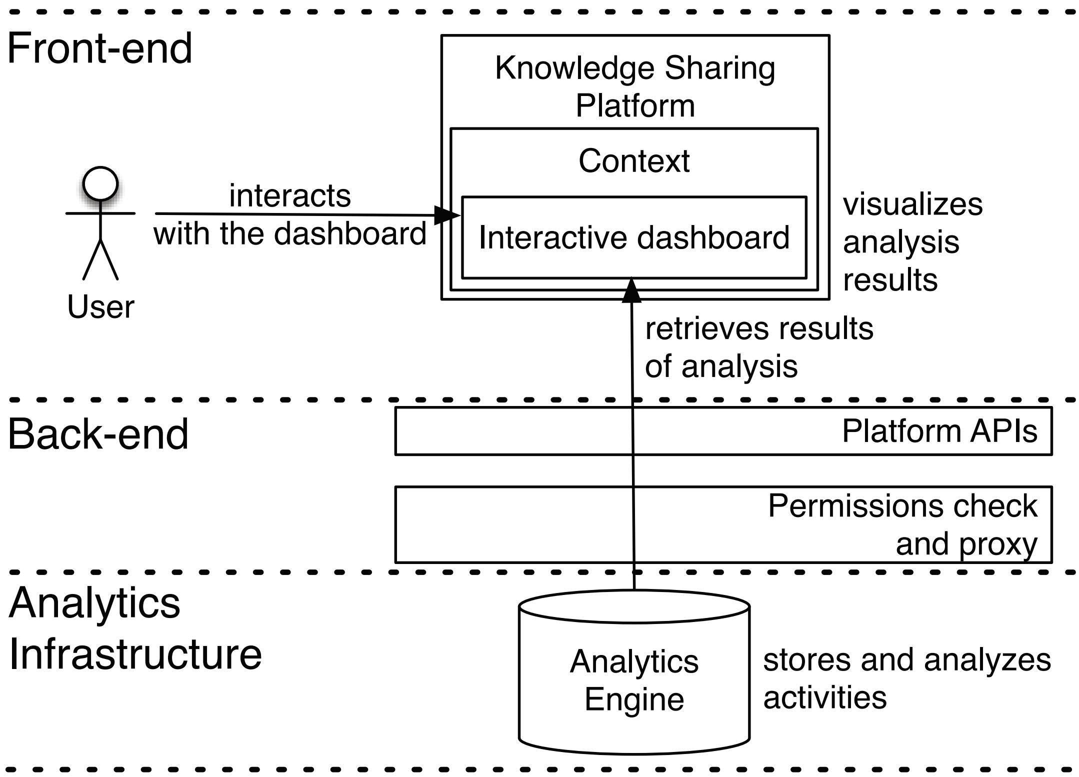


Figure 2.15 – The proposed backend analytics engine architecture. The dashboard is embedded into the context where the content and apps are located s o the user can interact with the analytics in the context. The dashboard obtains the analysis results for the context from the analytics engine located on the back-end.

2.6.1 Architecture & Components

To implement the proposed architecture when deploying in existing platforms, we propose a novel way to use Elasticsearch[29] as an analytics engine and Kibana[30] for interactive dashboards allowing to provide embedded contextual analytics.

Elasticsearch is one of the popular open-source search and analytics engines meeting the scalability requirements. It was initially envisioned as a search engine capable of operating at scale and in real-time, Elasticsearch is used by organizations such as Wikimedia[31] and Github[32], which deal with big data on a daily basis. Until recently, the exploitation of Elasticsearch for interactive analysis purposes by practitioners was hindered by a high entrance barrier due to the complexity of the query language and the query particularities. Now, such analysis is facilitated by Kibana, an open-source tool that allows to conduct analysis and build interactive dashboards allowing exploration of Elasticsearch data through a graphical user interface.

[29] Elasticsearch https://www.elastic.co/products/elasticsearch (last accessed 10 November 2016)
[30] Kibana https://www.elastic.co/products/kibana (last accessed 10 November 2016)
[31] Wikimedia https://www.wikimedia.org/ (last accessed 10 November 2016)
[32] Github https://github.com/ (last accessed 10 November 2016)

Kibana does not require the user to dive into technical details of the queries (although it is still possible) and hence makes big educational data visualizations accessible to regular users. The additional value of Kibana comes in play whenever several visualizations are combined on a single dashboard, enabling to use multiple coordinated views for an interactive exploratory analysis via cross-filtering.

Both Elasticsearch and Kibana, together with Logstash[33] are part of an analytics stack often referred to as ELK[34]. Logstash supports data acquisition from multiple sources (including Twitter, RSS, event logs) thanks to its rich set of available connectors. Custom connectors can be developed for case-specific sources. In addition to the mentioned values, ELK enables building analytics infrastructures decoupled from the learning platform, i.e., it allows to host separately the learning environment (with the analytics functionalities) and the data storage without affecting the end-user experience. In the next part, we explain how such architecture can be constructed.

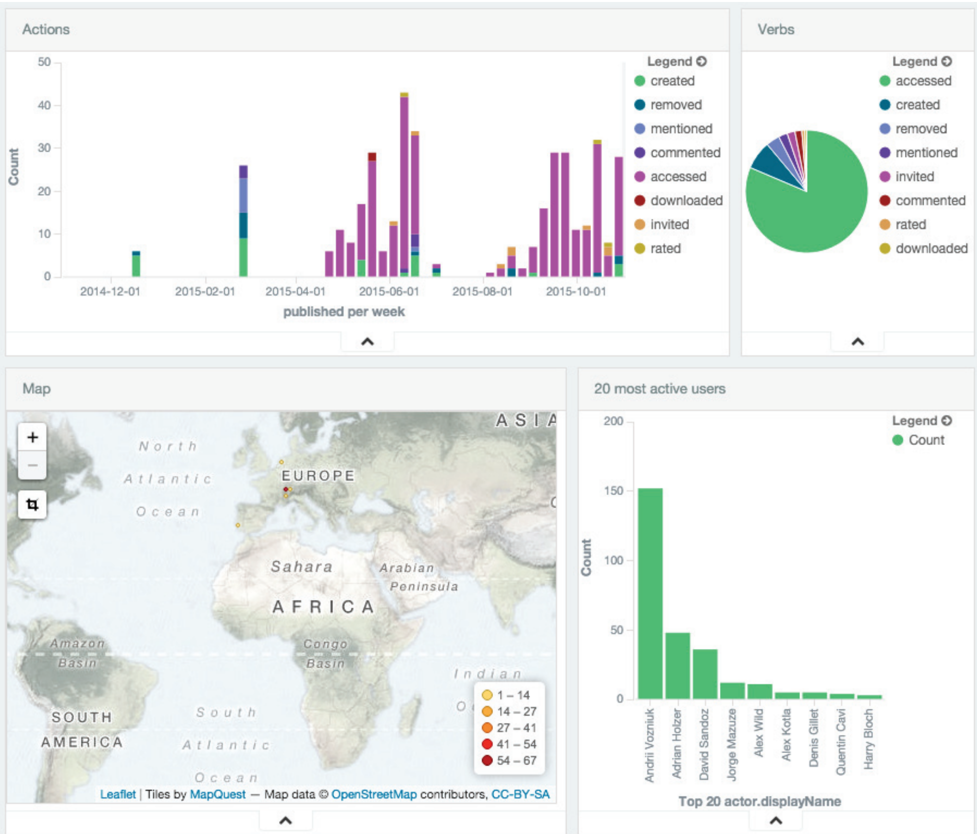


Figure 2.16 – A sample Kibana dashboard built from four visualization widgets.

[33]Logstash https://www.elastic.co/products/logstash (last accessed 10 November 2016)
[34]The ELK Stack https://www.elastic.co/webinars/introduction-elk-stack (last accessed 10 November 2016)

2.6.2 Properties

Below, we discuss some of the properties of the proposed architecture using backend analytics engine implemented with Elasticsearch and Kibana.

Real-Time Updates. Elasticsearch is a near real-time (NRT) engine[35], which in practice means that the updates happening in the interaction context are reflected on the Kibana dashboard according to the refresh interval set by the user. This interval can be under one second which can be perceived as real-time by the user.

Data Storage and Privacy. With the proposed architecture, an organization can setup their private Elasticsearch instance on premises and the activity traces can be forwarded and stored there for analysis. When a user interacts with a Kibana dashboard embedded into the platform, the dashboard retrieves and visualizes the data from that private Elasticsearch instance. That makes it possible to provide analytical services to the user without storing data on the platform in case this is a preferred option from the trust or legal point of view. The permissions to access the dashboard are managed on the platform with proper permission management limiting access to the dashboard to only the right users, protecting in this way the user privacy.

Interactiveness. The Kibana dashboards are interactive, and they support cross-filtering with coordinated multiple views that allows to filter data on multiple dimensions simultaneously. Moreover, Kibana provides options to disable and enable some of the filters when exploring the data. Filtering can be done by clicking on specific items on visualization widgets or by selecting ranges like in the case of the timeline widget or the map widget.

Personalization. It is possible to allow the user to choose from multiple dashboards available in the system in order to select the dashboard that matches her needs. This configuration can be as well done per context in case the platform provides such an option. But regarding the configuration of the dashboards, it can be done only by the user that has access to the Kibana administrative interfaces, normally an IT administrator. In this sense, the personalization capabilities by the end-user of the platform are fairly limited.

Portability. Kibana dashboards are embedded via an iframe and are not tightly coupled with the platform where embedding is done. This makes the dashboard portable allowing to embed the same dashboard into various platforms. But to make it work properly, two prerequisites need to be met: (1) the platform should supply the activity data to an Elasticsearch instance in a predefined format (in the case of this study that is ActivityStreams) and (2) the platform where the embedding is done needs to supply the context identifier to the dashboard, so the latter is aware of the context it is embedded into. Also, the platform needs to implement an API to proxy Kibana requests, so that a proper permission check can be done.

[35] Near Real-Time Search https://www.elastic.co/guide/en/elasticsearch/guide/current/near-real-time.html (last accessed 10 November 2016)

2.7 Validation of Analytics with a Backend Analytics Engine

As outlined in Section 1.1, Graasp is being used by MSF to share organizational knowledge. The process of knowledge sharing is managed by MSF knowledge managers that need to facilitate and adapt the sharing to the needs of the organization. During our participatory design sessions with knowledge managers from MSF, they expressed a strong need for obtaining key metrics regarding the process so they can better understand it and intervene if necessary. Some of the questions they needed to answer included who are the active contributors, active consumers of content, what are the most valuable documents, from which countries which content items are being accessed and downloaded the most. Who participates the most actively in discussion and who is not involved? Having answers to these questions can help to define a path towards improving the knowledge sharing. To provide MSF knowledge managers with answers to the mentioned questions, it was possible to query the database and to simply display numbers and graphs on the platform. But based on discussions with them during participatory design sessions, it became clear that they wanted to be able to dig further into the analytics and hence interactive visualizations were required.

2.7.1 Integration into Graasp

To validate our approach, we have integrated the proposed architecture into Graasp (**validation 1.a**) to allow knowledge managers and knowledge workers to be aware of and reflect on the interactions happening in their workspaces. The high-level architecture is shown in Figure 2.17.

As explained in Section 1.1, the back-end of Graasp is built with Node.js and uses MongoDB as a data storage. Mongoose.js[36] library is used for object-document mapping when interacting with the database. Since the activities of users are captured inside Graasp, they need to be transmitted to Elasticsearch for storage and processing. To achieve that, we use the mogoosastic plugin[37] for mongoose that makes sure that all of the changes in the actions stored in MongoDB are reflected in Elastisearch data, so it remains consistent (see Figure 2.18).

Kibana provides an embed code, allowing to integrate dashboard into another platform. We use the embed code to embed dashboard into Graasp web pages. Also, we supply the embedded dashboard with the context identified. A context in case of Graasp can be a space (on any level of nesting), a content item (pdf, Microsoft office document, etc.), an application or a discussion. Since the visualizations present on the dashboard have a context identifier, they visualize the data specific to this context.

The embedding is schematically shown in Figure 2.19. When a user opens the dashboard in the space, first the Graasp back-end checks if the user has the necessary permissions to view the dashboard in the current context. At the moment, only space owners have access to the dashboard. We decided to put such a check in place to allow limiting the dashboard access to

[36]Mongoose library http://mongoosejs.com/ (last accessed 10 November 2016)
[37]Mongoosastic plugin https://github.com/mongoosastic/mongoosastic (last accessed 10 November 2016)

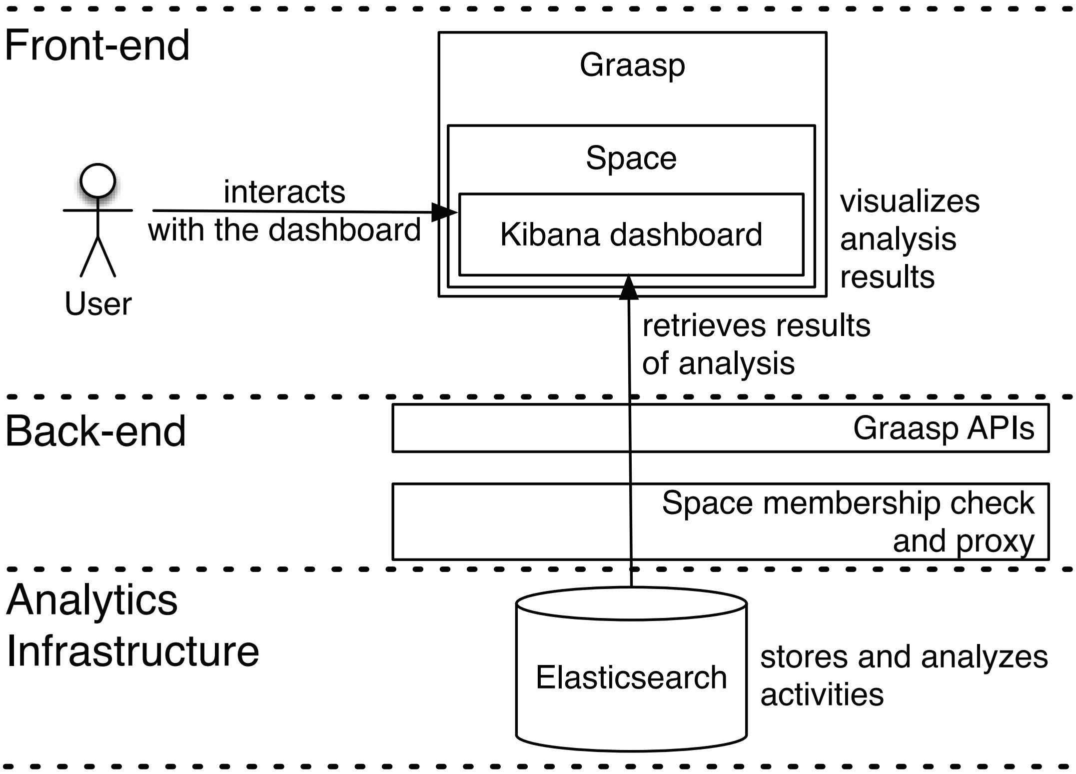


Figure 2.17 – The proposed backend analytics engine architecture integrated into Graasp. The user interacts with the Kibana dashboard located in the space that retrieves the analysis results from the Elasticsearch engine through platform APIs and a proxy on the back-end.

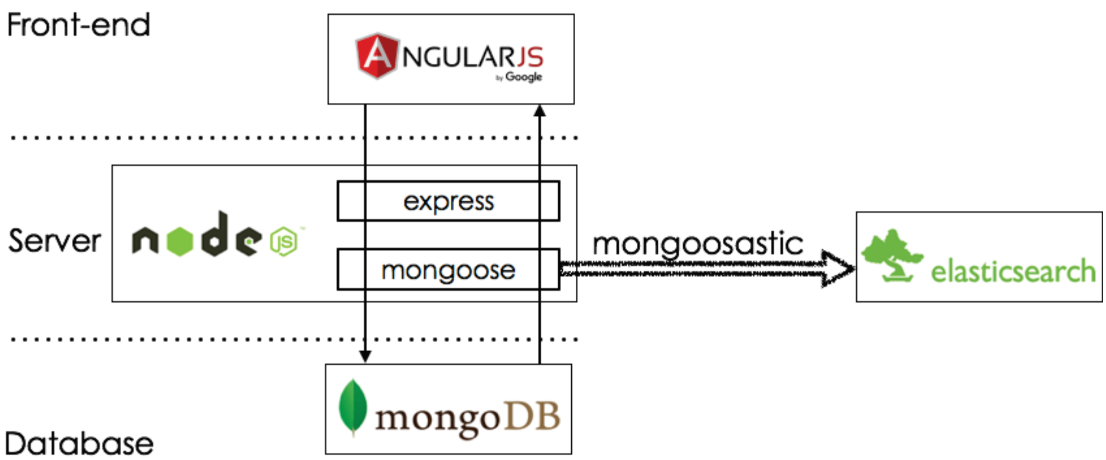


Figure 2.18 – Mongoosastic plugin for mongoose is used to send data to Elasticsearch.

relevant users since some of the users were concerned about their analytics being accessible by all of the space members. If the user has the access, the Graasp backend makes a request to Kibana for the dashboard iframe and sends it back to the Graasp front-end. Afterward, the iframe starts making requests to Graasp for javascript, CSS files, and other resources.

After these resources are loaded, requests are made to obtain the values of the metrics to be visualized. Graasp does the permission check on every request, and if they are sufficient, proxies the request to the Kibana server and then the Kibana response back to the client. If the permissions are not sufficient, the user can not access the dashboard and the user privacy is protected.

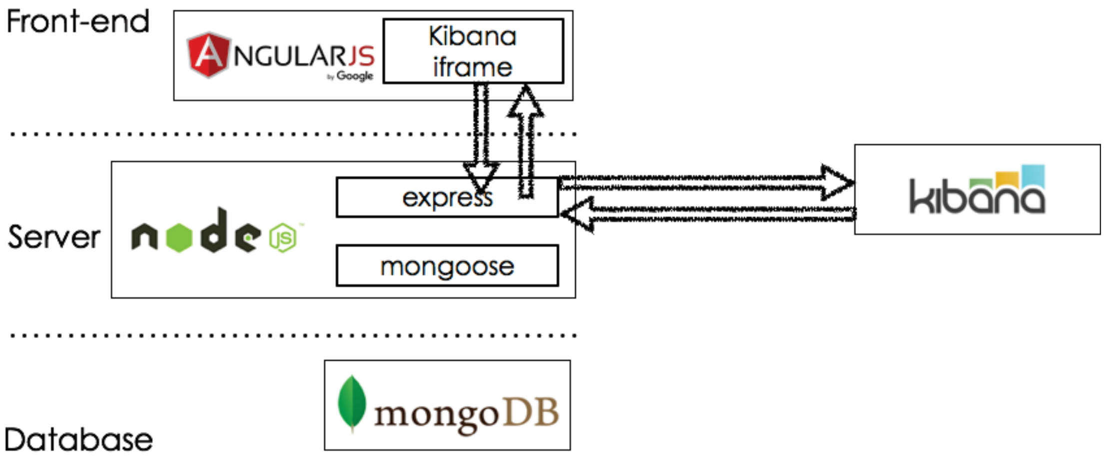


Figure 2.19 – A schematic explanation of the dashboard integration into a web-platform.

2.7.2 Developed Interactive Dashboards

To build a dashboard suitable for knowledge managers and knowledge workers interacting in Graasp, we have applied the participatory design methodology (see Section 1.3), involving MSF knowledge managers and knowledge workers from the very beginning and on every stage of conception. An early mockup of the proposed dashboard is presented in Figure 2.20.

Figure 2.21 presents the first version of the Kibana dashboard integrated into Graasp. Based on the feedback from the users, it became clear that more detailed explanations are necessary regarding the purpose of the dashboard and the individual widgets displayed on it. For this reason, we developed the second version adding widgets that contain descriptive information as shown in Figure 2.22.

2.7.3 Usage Data

To understand if the dashboard is used, we have set up the recording of users visiting the dashboard tab in Graasp. In the period from May to November 2016, 156 distinct users visited the tab 512 times in total. The users checked the dashboard for 307 different items in Graasp. These numbers indicate that the dashboard is adopted by the Graasp users.

Figure 2.20 – An early mockup of the analytics dashboard embedded into Graasp. Multiple types of visualizations describe the interactions in the Field Essentials space.

Figure 2.21 – The first version of the Kibana dashboard embedded into Graasp.

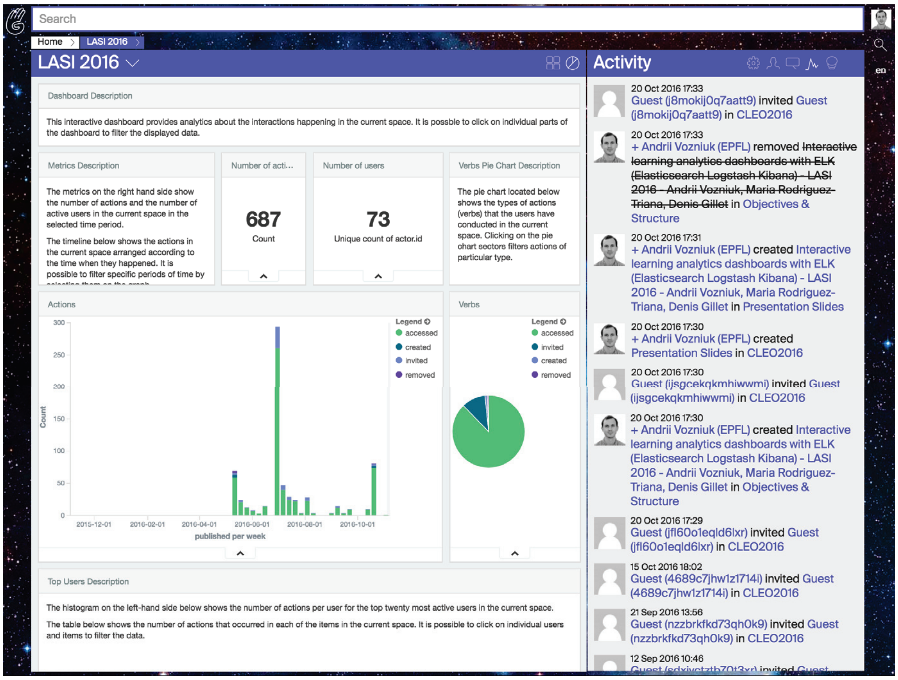


Figure 2.22 – The second version of the Kibana dashboard embedded into Graasp providing a detailed description of each visualization.

2.7.4 Discussion

In this section, we demonstrated how the proposed general architecture based on the backend analytics engine could be used to answer RQ1 in a specific knowledge sharing platform, in our case Graasp. We used Elasticsearch as an example of scalable analytics engine and Kibana for building interactive dashboards. Using these technologies, we constructed dashboards providing user-oriented analytics. This allowed to answer RQ1 for knowledge managers by supporting their reflection regarding the knowledge sharing process. The reported usage data indicated that the platform users adopted the constructed dashboard.

2.8 Comparison of the Approaches

In the previous sections, we have presented two general architectures to deploy embedded contextual analytics in web-based knowledge sharing platform. In this section, we highlight the differences between the architectures, advantages, and limitations of each of them as summarized in Figure 2.2.

	Data Storage	Computations	Scalable	Typical Usage	Construction	Customization	Extendability	Dashboards Construction	Portability Requirements
Analytical Web Apps	On the platform	Front-end, in browser	No	Short-term sessions, awareness	Developers	Virtually not limited, fits well scenarios	By building new apps	Teachers, Students	OpenSocial APIs
Analytics Engine and Interactive Dashboard	On the platform or on premises	Back-end, with analytics engine	Yes	Long-term, reflection	IT Admins, Analysts	Limited by Kibana, general vocabulary	By building Kibana plugins	IT Admins, Analysts	ElasticSearch & Kibana integration

Table 2.2 – Comparison of the analytical architectures based on analytical apps and backend analytics engine with interactive dashboards.

Data Storage. The proposed two architectures express two different data persistence policies. The apps-based approach assumes that the learning platform itself is a safe place to store the traces. So it gets the traces from the platform and then analyses them on the front-end. The backend analytics-based approach considers that it should be possible to store the data outside of the learning platform, for instance in a school or on student's personal hosting and still have the analytics presented inside of the learning platform.

Computations. In the apps-based approach, the computations over the data are made on the front-end. Hence, the computational capabilities are limited by the device where the browser runs. While the computing power of devices, particularly the mobile ones, keeps on increasing it is still limited by a single device. Differently, in the Elasticsearch-based approach, the computations are done on the back-end, and Elasticsearch makes it possible to use computations capabilities of multiple servers on-premises or in the cloud.

Scalability. The apps-based approach is limited to utilizing resources of a single computing device where the browser is running and thus is not scalable. The analytics backend approach can use multiple nodes and is scalable.

Portability. The two proposed architectures offer different approaches to portability. The apps-based architecture requires the platform to provide a standard set of APIs so the apps can access the data and to store the data in a standard format, so the apps know how to analyze and represent it. The backend analytics engine approach requires the platform to send the data in a standard format to an Elasticsearch instance so that configured Elasticsearch queries can run on the data.

Typical Usage. Due to the scalability limitations, the apps-based approach is suitable for short sessions where a moderate amount of interactions happen. The backend analytics engine offers scalability provided by using a set of servers on the backend and hence is suitable for settings where big educational or humanitarian data needs to be analyzed.

Personalization. Apps support personalization by the user allowing to configure personalized dashboard by adding and removing the apps, and the apps can be developed to fit the needs of the users. In contrast, it can be challenging to personalize the interactive dashboard with Kibana and may require an IT specialist to modify the dashboard by configuring visualization widgets.

In summary, based on the described properties both proposed architectures can be used for awareness and reflection. But due to real-time, personalization, not-scalable properties of the apps-based approach, in practice, it is more suitable to support awareness. At the same time, the backend analytics engine-based architecture due to the interactivity of the dashboards and ability to scale well and analyze big amounts of data is more suitable for reflection.

2.9 Implications for Educational and Humanitarian Settings

When managing knowledge sharing inside of organizations, knowledge managers need tools to monitor the knowledge sharing process in order to be able to improve it and make it effective. Similarly to how analytics applied in the cases of learning is called "learning analytics", we propose to refer to analytics related to knowledge sharing as "knowledge analytics". In an analogous way as learning dashboards can be used to understand and improve learning, knowledge analytics dashboards can be employed to visualize interactions happening on the platform and explore them to identify possible points of improvement in knowledge sharing.

The backend analytics-based architecture with Elasticsearch and Kibana was originally designed for enabling awareness and reflection in the knowledge sharing process. Then, it was demonstrated to Go-Lab project members. The project members thought that such dashboards could be useful to monitor overall project metrics by aggregating data from several

ILS located in the same space. As a result, at the moment of writing the analytics engine and interactive dashboard are applied in the learning context as well. We plan to validate this architecture with project members as well as teachers as a complement to apps.

2.10 Conclusion and Future Work

In this chapter, we aimed to answer RQ1 by providing the knowledge sharing platform users with user-oriented analytics that is specifically designed to support their awareness and reflection when interacting with the platform. Our main contribution is an end-user oriented embedded contextual analytics model answering RQ1. The model advocates delivery of analytics to the platform user in an embedded and contextualized manner by (1) integrating analytical tools and dashboards directly into the context where the user interacts, enabling in this way awareness and reflection and (2) by making the tools analyze and present the data from within this interaction context.

To be able to answer RQ1 in existing platforms and put the proposed model to work we made two additional contributions that are two general analytics architectures materializing the model. The first architecture relies on modular real-time analytical apps that can be embedded into the context and that analyze and visualize activities performed by users in this context as they happen. We have validated the first architecture by developing three sample analytical apps addressing the primary awareness needs of teachers in blended learning sessions. The second architecture is designed on top of a backend analytics engine facilitating scalable and interactive analysis. This architecture was validated by developing knowledge analytics dashboards to support awareness and reflection of knowledge managers on a knowledge sharing platform.

The two presented architectures are designed with different assumptions regarding the privacy of the collected data. The apps-based approach is designed assuming the platform itself is a safe place to store the traces. The backend analytics approach allows storing traces separately from the platform building on the assumption that organizations and users want to keep traces in their private location. In practice, the choice of the suitable architecture depends on specific requirements and tradeoffs among multiple properties, including real-time updates, interactiveness, scalability, personalization, and data privacy.

We would like to stress that the proposed model and corresponding architectures are more broadly applicable than just for learning or knowledge analytics. They can be employed for general user activity analysis (e.g., software development analysis) or as a part of even more general tools that make use of user activities (e.g., recommendation or gamification services).

It is also worth taking a look at possible directions for future work.

Student Needs. When proposing the apps-based analytics architecture, we targeted the teacher awareness needs primarily. But during the evaluation, the students expressed such needs as well, for instance, to make them aware of their progress compared to the progress of others. In the future, it is possible to identify particular awareness needs of students and build apps that address issues related specifically to student awareness. However, during the evaluation we also noticed that some students were not happy about others seeing their progress, we think that the questions of students data privacy and the granularity of the interactions being recorded are worth investigating.

Types of Indicators. While our primary goal in evaluations was to provide analytics about interactions happening in context, the concept of embedded contextual analytics is not limited to the interaction-based indicators. Other possible indicators included content-related (e.g., showing the main concepts present in the content) and user-related (e.g., indicating which activities individual students have completed in the past).

Teaching and Learning Outcomes. An open question remains regarding the impact of the proposed LA apps and dashboards on teaching and learning. While we obtained positive reported feedback from the teachers and students, it is still not clear if the teachers improve orchestration by relying on analytics and whether there is a measurable positive impact on the learning outcomes. Measuring changes in learning outcomes resulted from deploying technology is a known challenge in the learning analytics community [100].

Content Analytics. Last but not least, in this chapter we mainly focused on LA for formative assessment of activities and learner's progress. Teachers and schools are also requesting to use LA for normative assessment (grading). This challenge will be tackled in the Next-Lab project by providing embedded contextual analytics not only about the conducted activities but also regarding the artifacts (e.g., reports) produced during these activities in the context.

In this chapter, we focused on recording user interactions in a learning platform, their analysis, visualization and interactive exploration to facilitate awareness and reflection. During the case studies in the classroom and when deploying knowledge analytics, the users expressed privacy concerns related to recording, retention, and analysis of the interaction traces. In the next chapter, we are going to target privacy-related challenges.

3 An Agent-based Contextual Privacy Management

The storage and analysis of user interaction traces discussed in the previous chapter are often associated with privacy concerns. Part of such concerns are triggered by the legal framework in place in a particular country or context (e.g., sensitive medical records or data of school students under specific age), and another part is related to an established negative perception of being tracked due to possible data misuse. To address such concerns, in this chapter, we present a novel agent-based privacy management interface and the corresponding privacy enforcement mechanism. Also, we describe the implementation of the proposed privacy interface in a social media platform and discuss the evaluation results. This chapter targets the second research question:

RQ2 - Privacy: what privacy management interfaces and mechanisms are suitable for knowledge analytics and learning analytics?

The content of this chapter was partially published in [109]:

A. Vozniuk, S. Govaerts, L. Bollen, S. Manske and T. Hecking et al. AngeLA: Putting the Teacher in Control of Student Privacy in the Online Classroom. ITHET, York, England, September 11-13, 2014.

3.1 Introduction

Learning analytics (LA) is often considered as a means to improve learning and learning environments by measuring student behavior, analyzing the tracked data and acting upon the results. The use of LA tools implies recording and processing of student activities conducted on software platforms. Recently, personal data privacy has received global attention, due to the revelations of the NSA scandal[1]. These revelations and the scope of NSA programs made some users of online services much more concerned about their data privacy and established a more negative perception of being tracked. The users demand more transparency concerning

[1] The Guardian NSA Files, http://www.theguardian.com/world/the-nsa-files (last accessed 10 November 2016)

what personal information is collected, by whom, how it is processed, and for which purposes. Moreover, the users want to have more control over the data collection and processing policies.

To be able to control data privacy, it is important to understand the many aspects and definitions of data privacy. According to the state of the art analysis of data privacy done in the framework of the SPION project [2], two types of data privacy can be identified: (1) social privacy and (2) instrumental privacy. Raynes-Goldie [91] defines *social privacy* of users as "the control of information flow about how and when their personal information is shared with other people". As an example, social privacy can be achieved by introducing a trust-aware data sharing mechanism [73]. According to Boyd [14], *instrumental privacy* is defined as the control of data access by corporations and governments, for instance for data analysis or data mining. In this chapter, we propose a novel model for managing instrumental privacy permissions. This model defines when the user data can be stored by the platform and accessed by other third-party platforms.

Data privacy is regulated in different ways by many national governments and often depends on the target audience. For instance, in the case of schools and teachers, student data privacy is of particular importance, since student online activity tracking is subject to stronger legal privacy regulations, due to the age of school pupils. Apart from governments guarding over privacy, the software itself and its privacy policies have evidently a significant impact on data privacy. Boyd et al. [15] discuss how privacy can be enforced on the system architecture level and thus be an inherent part of the software design. Establishing privacy mechanisms on the system design as well as the user interface levels is also our goal in the Go-Lab European project (covered in more details in Chapter 1). We need to make sure that the user is aware of the tracking being conducted and provide the user with an intuitive interface to adjust the tracking policy (**challenge 2**).

This chapter focuses on the instrumental privacy, and we propose a flexible, contextual and intuitive way to provide the teacher with full control over student activity tracking in online collaborative environments. We call this approach AngeLA, the name is inspired by an agent guarding over student privacy in Learning Analytics. AngeLA, as an agent-based privacy management interface, mimics in a virtual space the privacy control mechanism that works well in a real room: if a person is present in a room, she can observe all activities happening in the room. AngeLA serves two primary purposes: (1) it increases the awareness of teachers about the activity tracking and (2) provides an intuitive way to manage the activity tracking permissions.

This chapter illustrates the design and implementation of a data privacy management system that allows teachers to control the activity data collection policy in an online learning environment (OLE). The chapter is structured as follows. First, Section 3.2 formulates the requirements for a tracking permissions management system in an OLE. Then, Section 3.3 considers existing approaches for privacy management user interfaces. Afterward, Section 3.4 explains the philosophy and idea behind the proposed solution (called AngeLA). Section 3.5

demonstrates how the widely applicable AngeLA interface is implemented in a specific context and platform and evaluation results with the platform users are presented in Section 3.6. Finally, the Section 3.8 discusses the conclusions.

3.2 Requirements

To identify requirements for a privacy management system, we arranged discussions with over ten experts from educational technology and pedagogy from Go-Lab. As result of these discussions and based on the Go-Lab project prerequisites, we have identified five key requirements for a privacy management system in an online learning environment, namely:

1. *fostering awareness.* The user interface of the privacy management system should make a teacher aware of the ongoing activity tracking.
2. *offering intuitive UI.* The privacy management UI should be built with the help of concepts and elements of UI already familiar to users.
3. *enforcing student privacy.* A teacher should be in control of the tracking and be able to enable or completely disable it when needed.
4. *providing flexibility.* It should be possible to adjust privacy policy depending on the context, for instance, for one group of students the tracking can be enabled and at the same time for another one it should remain disabled.
5. *enabling data aggregation.* The system should be able to aggregate relevant student activity data coming from all relevant parts of the learning platform.

3.3 Related Work

An intuitive way to increase user data privacy, implemented by web platforms such as Facebook[2] and Google+[3], is to have an extensive privacy policy settings. But even in the case when such a privacy control mechanisms are provided, it could be hard to use and understand because of the complicated menus and navigation hierarchy. A large number of options could as well make it hard for the users to make proper privacy decisions [99]. Furthermore, the default privacy policy could change quickly[4] and it could be hard for users to follow and adapt.

To simplify understanding of data sharing policy in a social network, Iannella et al. [53] suggest to use a set of icons in the user interface. While having proper icons in the GUI fosters privacy awareness, the approach targets social privacy, i.e. sharing the data between different users of the platform and seems not to influence the instrumental privacy.

[2] Facebook Data Use Policy https://www.facebook.com/about/privacy/ (last accessed 10 November 2016)
[3] Google+ https://plus.google.com/ (last accessed 10 November 2016)
[4] The Evolution of Privacy on Facebook http://mattmckeon.com/facebook-privacy/ (last accessed 10 November 2016)

Hull et al. [52] focus on the design of user interface that can increase user awareness about how the changes made in the user interface can influence user privacy. The authors argue that most of the privacy issues often found in online social networks can be explained by the difference in the context (contextual gap) between the online and offline social settings. Indeed, offline privacy is highly defined by the social context while online systems lack much of this context. Hull et al. showcase the privacy issues on Facebook as an example. They conclude that most of the issues could be resolved just by improving the user interface in a way that makes the information flows more transparent for users.

Another important aspect influencing the user privacy is the default privacy settings [66]. Some websites could benefit from understandable and restrictive privacy policy. For instance as showed by Tsai et al. [107], a more protective and transparent privacy policy of a shopping website increases the probability of users buying on the website. At the same time many sites profit from information disclosure made by users, for instance, Facebook has a default policy [44] promoting data disclosure. Online platforms could even employ so-called Dark Patterns[5] to make it harder for the user to change the default privacy settings into a more strict privacy policy.

3.4 The Analytics Tracking Agent

To answer RQ2, we propose a novel model for managing activity tracking permissions base on the virtual agent metaphor (**contribution 2**). The model relies on a virtual agent for handling privacy settings, acting as a data aggregator and transmitter. Our approach is motivated by the privacy mechanism and policy embedded into a physical classroom or any room in general. In a physical classroom the teacher is in control of the privacy of what the students do in the classroom, e.g. she can decide which student behavior she shares with the parents or with other stakeholders. If the teacher wants to have a private discussion, she can ask all unwanted parties to leave the room.

The agent mimics in a virtual space this privacy control mechanism that works well in a physical space. In a collaborative online space (e.g. an online learning space or a chat room), user management mechanisms exist to grant and revoke access to the space or resource. The proposed software agent can be invited into an online space together with other members. When the agent is a member of an online space, it can observe all activities taking place in this space, just like any other member of this space and like the teacher in the classroom. The agent can thus aggregate all activities of the space members and can *store* them in the platform or can *transmit* this information to third-party analytics infrastructures as shown in Figure 3.1. To let the users better understand the purpose of the agent, it has its own profile in the system detailed in Figure 3.2 describing the purpose of the agent and its privacy model.

[5] Dark Patterns, http://darkpatterns.org/ (last accessed 10 November 2016)

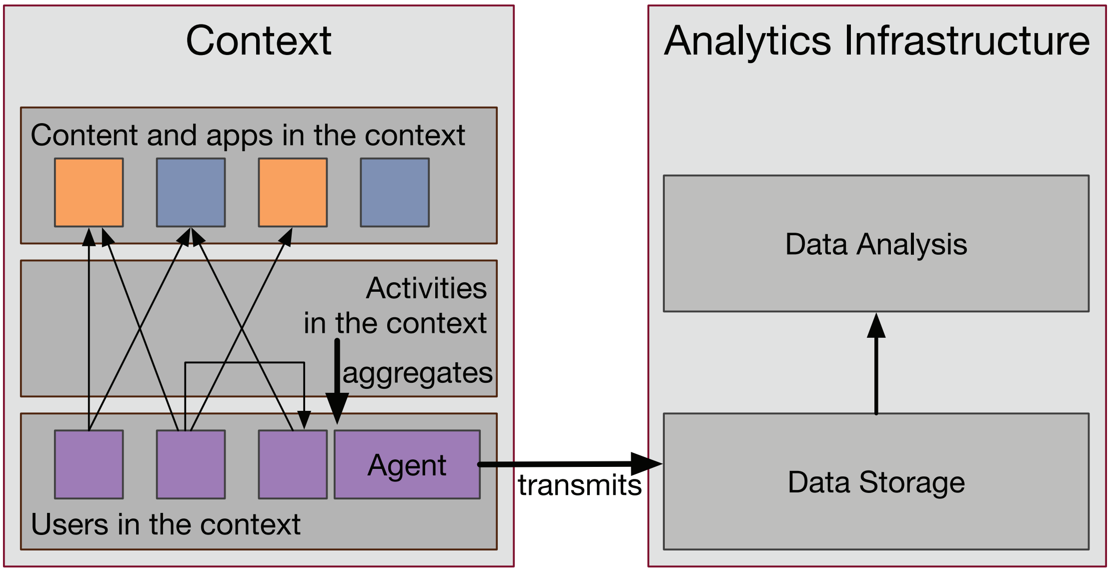


Figure 3.1 – A schematic representation of the agent-based privacy management. When the agent is a member of a context, it aggregates and transmits user activity data to analytics services.

Formally, for the activity a_t that occurred at point of time t in space s, the following check against the space members $M_t(s)$ at time t is used to decide if this action should be stored and transmitted to the analytics infrastructure:

$$\begin{cases} \forall a_t \in A(s) : TrackingAgent \in M_t(s) \Rightarrow Store(a_t)\&Transmit(a_t) \\ \forall a_t \notin A(s) : TrackingAgent \notin M_t(s) \Rightarrow \neg Store(a_t)\&\neg Transmit(a_t) \end{cases} \tag{3.1}$$

These rules do not depend on the action type and allow to decide at any point in time if the action should be stored and transmitted or not. The rules also do not depend on which of the users added or removed the tracking agent. They are defined by the agent's presence as a member of the space at the moment when the action occurred. Hence, the addition or removal of the agent does not have an impact on the previously observed actions. Hence, in practice, it is reasonable to apply the rules at the instance when the action is observed by the platform $a_{t=now}$ and a simpler version of the rules can be used at that moment:

$$\begin{cases} TrackingAgent \in M_{t=now}(s) \Rightarrow Store(a_{t=now})\&Transmit(a_{t=now}) \\ TrackingAgent \notin M_{t=now}(s) \Rightarrow \neg Store(a_{t=now})\&\neg Transmit(a_{t=now}) \end{cases} \tag{3.2}$$

Any of the online space owners, in our case a teacher, can revoke the agent's access to the space, after which the agent can no longer track any user activity in this space. The agent's permissions can be configured per space depending on the activity context in the same way as a person can be invited to be present in some room and at the same time not invited to other ones. By managing privacy via access control of the agent to an online space, the teacher is in full control of when the student activity is tracked and when not. This privacy control happens through already familiar user management functionality of inviting collaborators. Furthermore, by having the tracking agent as a visible space member, the teacher is aware of agent's presence and hence that the tracking is turned on. Having easy-to-use privacy control and being open about the tracking policy can also increase trust in the system [107]. With the agent we propose a soft paternalistic privacy policy, where the system does not force a user to make specific privacy decision, but rather makes her aware about the ongoing activity tracking and provide an easy and intuitive way to change it, hence answering in this way RQ2.

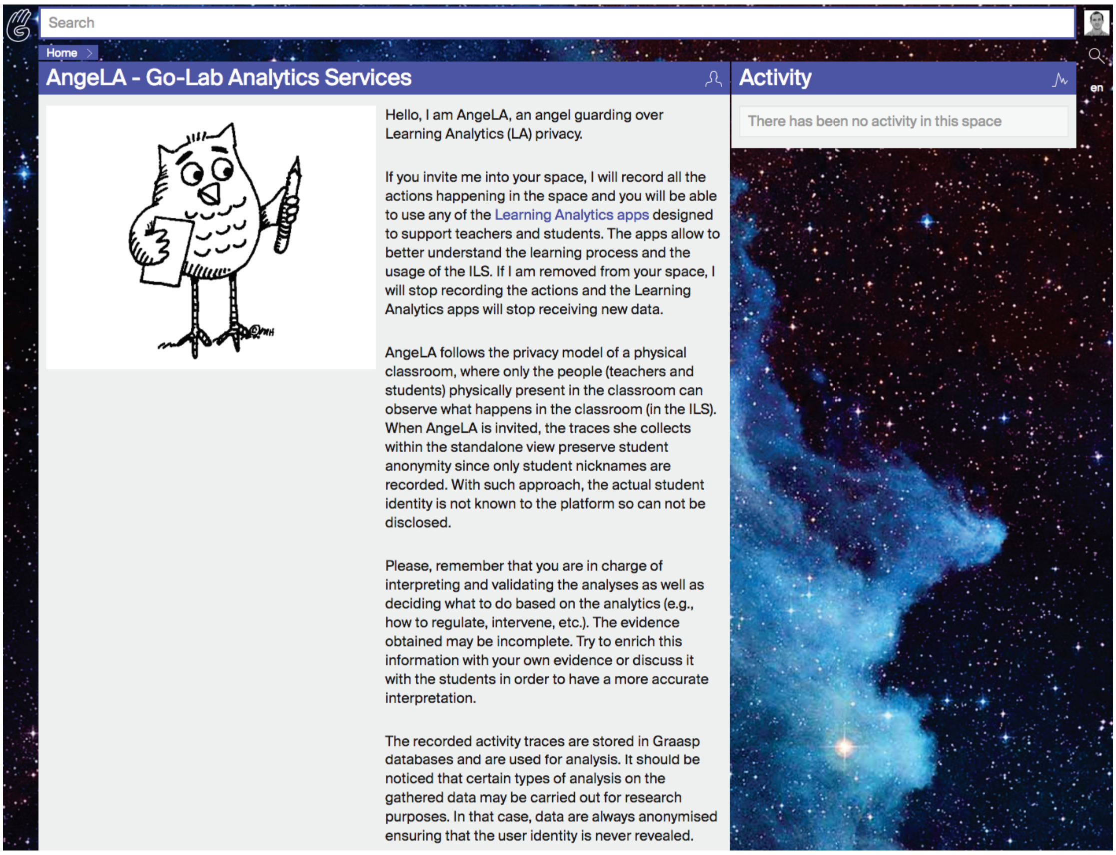


Figure 3.2 – The AngeLA agent profile in Graasp describing its data collection purpose.

European Union Privacy Compliance. It is worth taking a look at how the proposed approach using a tracking agent complies with the requirements of the European General Data Protection Regulation (GDPR) (EU) 2016/679[6]. Specifically, we consider the Articles 7, 11, and 12

[6]REGULATION (EU) 2016/679 http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679 (last accessed 10 November 2016)

defining relevant aspects of the user privacy management. According to Article 7 (Conditions for consent), "The data subject shall have the right to withdraw his or her consent at any time." and "It shall be as easy to withdraw as to give consent." The tracking agent approach fulfills these requirements since it is as easy to remove the agent as to add it and the user can remove the agent any time disabling the tracking. Article 11 says "Where ... the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible." According to Article 12 (Transparent information, communication and modalities for the exercise of the rights of the data subject), "The controller shall take appropriate measures to provide any information ... relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child." The agent can meet the requirements of Articles 11 and 12 by providing a clear description of its purpose, specifying what information is recorded and how it is processed, including handling of the user identity. In case of AngeLA, such description is provided in her profile as displayed in Figure 3.2.

3.5 Implementation in Graasp

We have implemented the proposed approach in Graasp (covered in Chapter 1) with the agent that we nicknamed AngeLA for this case. In the Go-Lab project, teachers use Graasp to construct Inquiry Learning Spaces [42]. The teacher can control the student privacy by adding or removing AngeLA as a space member. When removing AngeLA, Graasp displays a message informing the user about the consequences as shown in Figure 3.3.

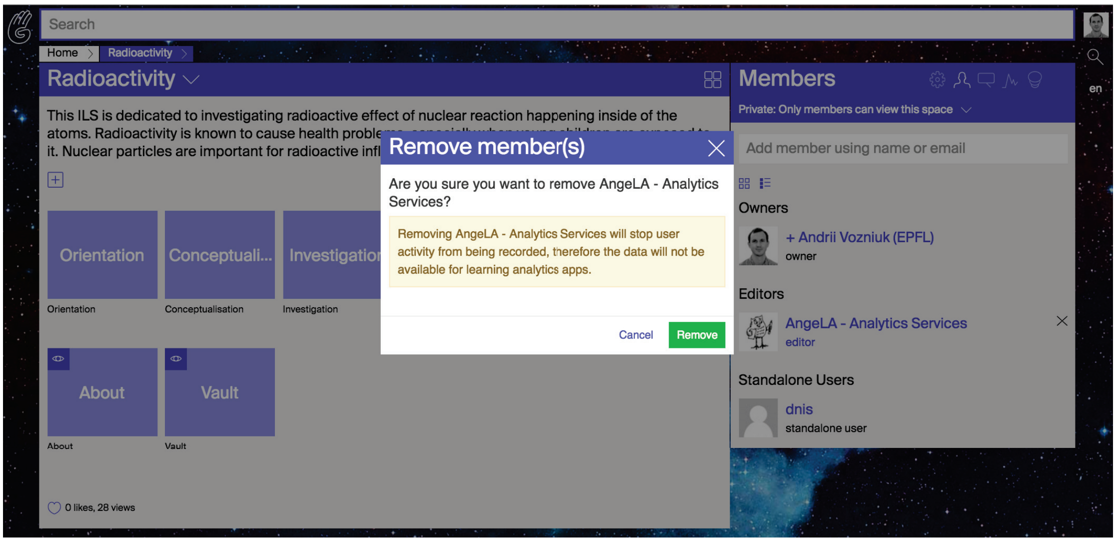


Figure 3.3 – A message in Graasp informing about the consequences of removing the tracking agent from a space.

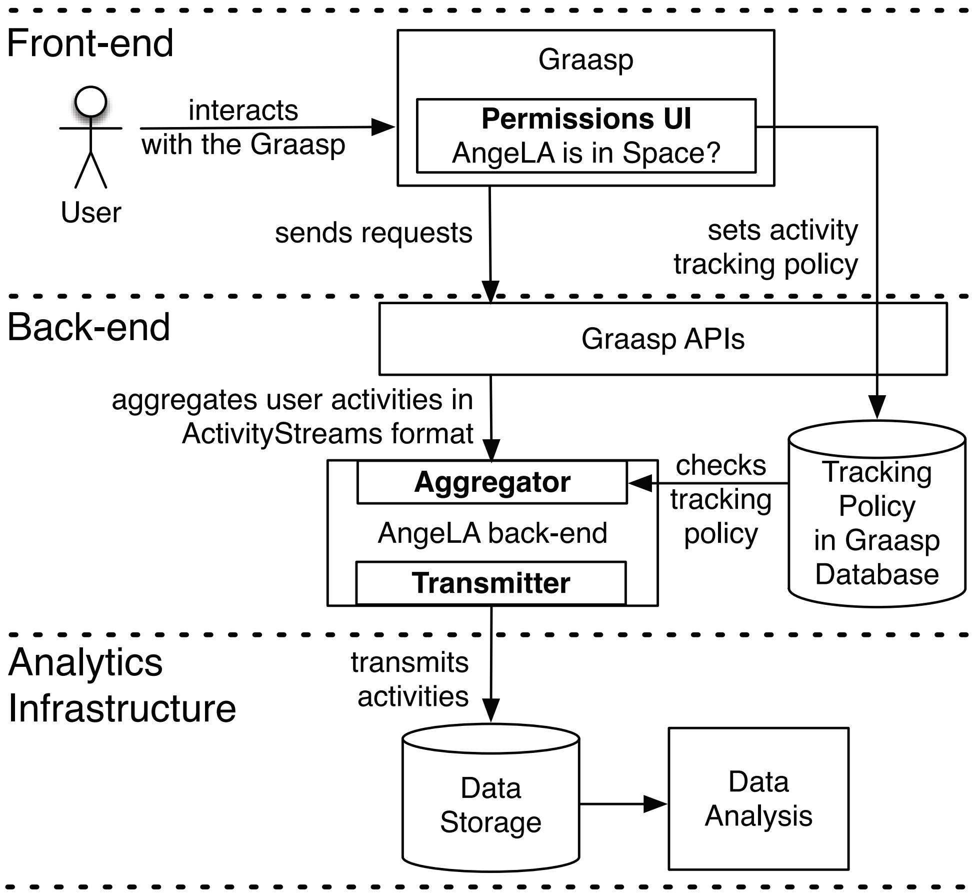


Figure 3.4 – An architecture of the agent-based interaction tracking in Graasp. Whenever the AngeLA user is present in the space, the AngeLA back-end aggregates the activities and transmits them to the analytics infrastructure.

The Learning Analytics Tracking Agent (AngeLA) architecture as presented in Figure 3.4 consists of the following three main components: (1) the Tracking Permissions UI that the user interacts with, (2) the User Activity Aggregator, and (3) the Activity Data Transmitter. The latter two components constitute the privacy management mechanism. Here, we take a look at each of the components.

Tracking Permissions UI. AngeLA provides an easy-to-use and familiar manner to manage privacy: i) to enable the activity tracking in a space the teacher just needs to invite AngeLA to the space (see Figure 3.5 (2)) and ii) to disable tracking the teacher can just remove AngeLA from the space. When AngeLA is present in the space (as in Figure 3.5 (3)), all the activities of space members will be collected, stored and sent to the LA back-end. This behavior is intuitive and familiar for the teacher since all space members are expected to be aware of the space activities.

User Activity Aggregator. AngeLA continuously aggregates activity streams of the Graasp

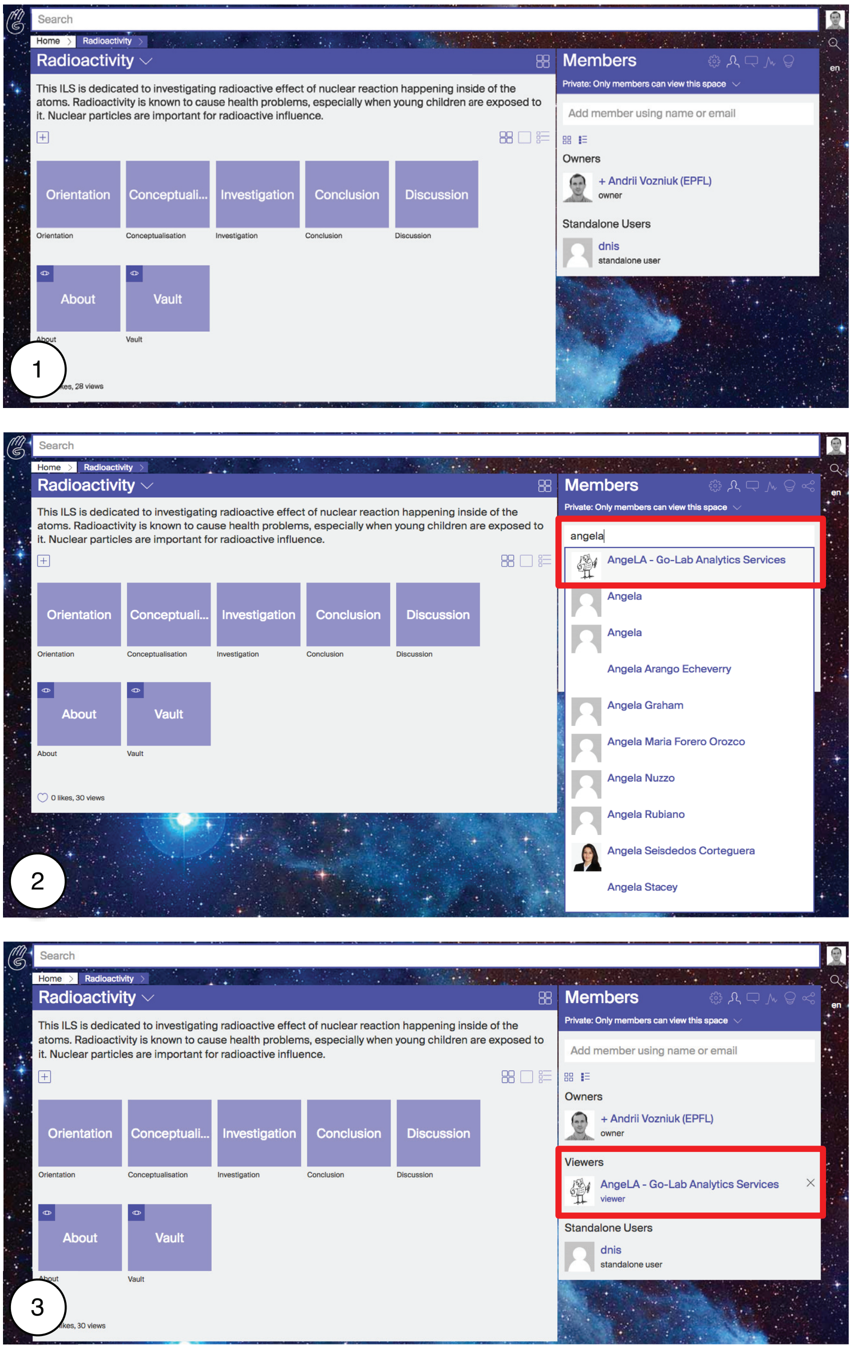


Figure 3.5 – AngeLA, an agent-based privacy interface. (1) AngeLA is not a member of the "Radioactivity" space in Graasp. (2) AngeLA is invited to become a member of the space after searching in the members list. (3) AngeLA is a member of the space.

users across the spaces where it is a member into a single activity stream. In the case of Go-Lab, the data is coming from Graasp (both the teacher and the student view) and the OpenSocial applications that the users interact with. In the case of the applications, they submit the captured activity data via the OpenSocial API as shown in Figure 3.4.

Activity Data Transmitter. All of the collected activity records are sent to the Graasp back-end for storage and further processing. The Activity Streams format is used to represent the actions during the transmission. As a mean to provide additional privacy in the Go-Lab project, as proposed by Li et al. [71], students use nicknames instead of real names to represent their identity in Graasp. In this case, only the teacher can do the mapping between the nickname and actual student name and hence knows a student's identity.

The nicknames approach indeed provides an opportunity for students to hide their real name from the platform. A teacher could ask students to use nicknames, but she can not guarantee that any of the students would not put a real name. Moreover, since the data collection is happening, it is possible that student identity could be revealed if proper data mining algorithms are applied. AngeLA approach aims to provide teachers with a mechanism to completely disable the data tracking, which guarantees that activity data is not collected.

To enforce privacy, it is necessary to establish a strict and clear default privacy policy [66]. In Graasp, AngeLA is a member of the inquiry learning spaces by default since this allows to record activities starting from the space creation before students start working within the space. The recorded activities enable learning analytics beneficiary for the teachers, but it is still possible to change the policy at any time by simply inviting or removing AngeLA from the space.

3.6 Evaluation

In the previous section, we have described the implementation of our approach in Graasp validating its feasibility. In this section, we focus on the evaluation of the approach using surveys with Graasp users and presenting a summary of actual usage.

3.6.1 User Survey

To understand if people comprehend the AngeLA metaphor, we conducted a usability study of the proposed approach implemented in Graasp with N=16 (7 were male, and 9 were female) Graasp platform users during two workshops covering inquiry-based learning (**validation 2.a**). The workshop participants performed activities within Graasp, where, as part of their task, they needed to define and adjust the data tracking policy for learning analytics using the AngeLA interface. After the workshop sessions, the participants were asked to fill in a SUS questionnaire [17] adapted to focus specifically on the evaluation of the privacy management component. The complete questionnaire we used is available in Appendix A.

Based on the responses, the final SUS score of the privacy management system was equal to 74, meaning that the system has a GOOD usability rating [6]. Figure 3.6 presents an overview of answers to individual questions of the SUS. The users reported that they were eager to use the AngeLA interface frequently ($\mu = 4.38, \sigma = 0.81$), they did not find it unnecessary complex ($\mu = 1.88, \sigma = 1.15$), they thought that it was easy to use ($\mu = 4.06, \sigma = 0.93$), they thought that they would not need the support of a technical person ($\mu = 2.06, \sigma = 1.06$) they found various functions to be well integrated ($\mu = 3.88, \sigma = 1.20$), they did not find too much inconsistency ($\mu = 2.31, \sigma = 1.14$) they would imaging most people learn to use it quickly ($\mu = 4.00, \sigma = 0.97$). It was not cumbersome to use ($\mu = 2.19, \sigma = 0.98$), they felt rather confident using it ($\mu = 3.94, \sigma = 2.06$) without the need to learn a lot of things before getting going ($\mu = 2.31, \sigma = 1.30$). In addition to these positive results based on the survey responses, we have obtained positive feedback during face-to-face communication with the workshop participants.

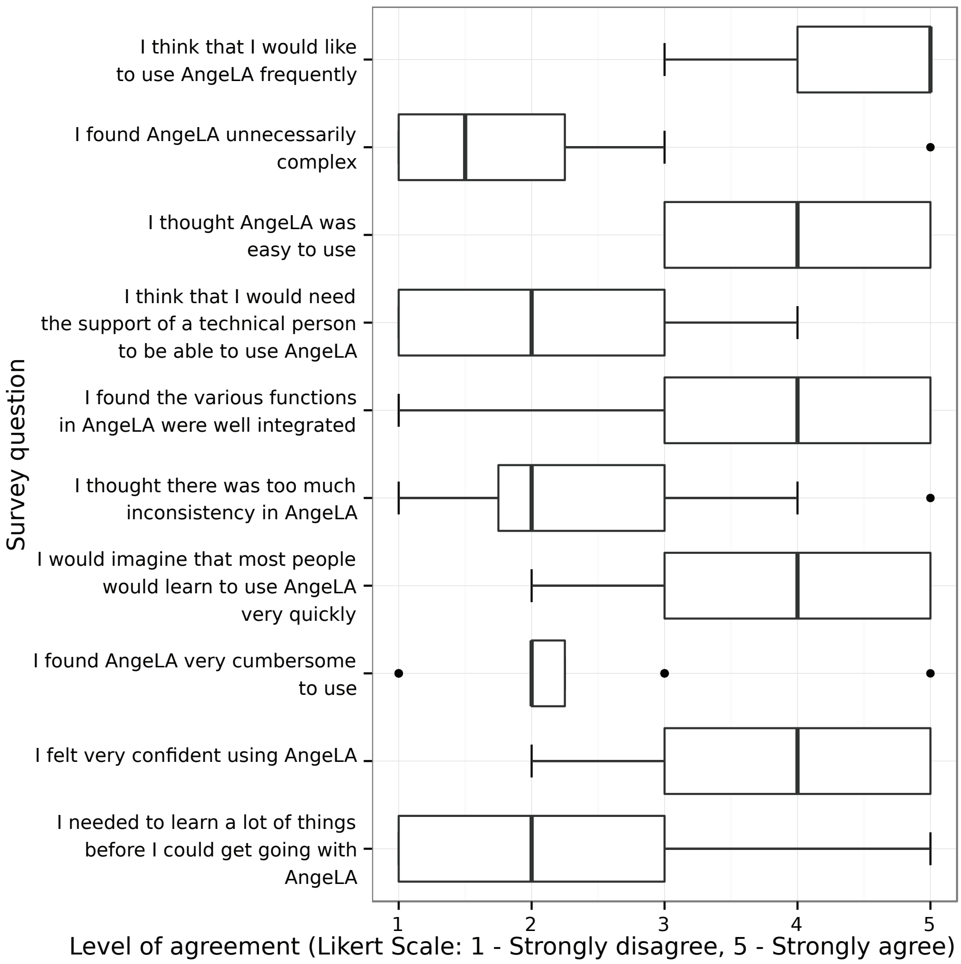


Figure 3.6 – System usability score survey results for the agent-based privacy management interface implemented in Graasp (AngeLA).

3.6.2 Actual Usage

In addition to evaluating the usability of the approach, we have looked into the AngeLA usage data to understand if the approach is used in practice (**validation 2.b**). As of August 2016, in Graasp there were in total 13163 inquiry learning spaces. Out of them, 11929 (91%) had AngeLA as a member and 1234 (9%) did not have AngeLA. The absence of AngeLA can be due to several reasons. The first reason is that an ILS was possibly created from a template that has had AngeLA removed by its creator, so all ILS created from the template did not have AngeLA. Since the recording of the ILS provenance was not available from the very beginning, it is not possible to get a precise number of such ILSs created from templates without AngeLA. Another reason is that AngeLA was initially present in the ILS, but was explicitly removed by the ILS owner, there are 168 (1%) such ILS in Graasp. This data suggests that most of the ILS owners prefer to keep AngeLA to benefit from learning analytics services, but still some owners prefer to have AngeLA removed to stop activity tracking.

In addition to the reported usage data, our proposed privacy management approach with AngeLA was selected by educational technology experts and is currently employed for privacy control by two EU projects, namely Go-Lab and SiWay further demonstrating its validity.

3.6.3 AngeLA & AngeLO

It is worth noticing that due to the requirements of the Go-Lab project, AngeLA aggregates activities in the ActivityStreams format and transmits them to the Go-Lab analytics infrastructure. Later, in the course of the SiWay project, similar requirements emerged, and we created another agent nicknamed AngeLO following the same privacy model. AngeLO aggregates the user activities using the xAPI format and transmits them to the SiWay analytics infrastructure implemented with a learning record store[7] (LRS). These two cases demonstrate how the general agent-based privacy model enables flexibility and can be applied in various context not depending on the activity format or the analytics infrastructure in place.

3.7 Implications for Humanitarian Settings

Recording, retention, and analysis of interaction traces have privacy aspects that platform developers and users need to take into account. The developers need to make sure that the user is aware of the tracking being conducted and provide the user with an intuitive interface to adjust the tracking policy.

In this chapter, we have evaluated our approach in the educational setting. Yet, the approach can be applied in a humanitarian setting as well. In humanitarian settings, analytics regarding the knowledge sharing can be seen as beneficial for obtaining insights. Nevertheless, due to

[7] Learning Record Store https://en.wikipedia.org/wiki/Learning_Record_Store (last accessed 10 November 2016)

the context where the humanitarian workers operate, exposing their interaction patterns may be a source of concerns [90]. The interaction patterns may, among others, expose person's location or time of day when the person is not working from home. Hence, in some contexts potential risks associated with data tracking may overcome the benefits derived from the analytics and it may be reasonable to disable tracking in such cases or limit what is tracked. In addition, specific regulations linked to sensitive documents, including medical records often encountered in humanitarian settings, may influence the user tracking policy employed by the platform. The provided examples show that the sensitivity of the tracked user data may vary from one interaction context to another, and hence it is necessary to provide the user with a way to adjust the privacy policy depending on the context.

Using our approach in the humanitarian context, Amy, the MSF doctor, can remove the tracking agent from her workspace, where she puts medical records of her patients. Contrary, the head of her mission Nermeen may invite the tracking agent into some of the shared spaces in Graasp without sensitive data, and where he would like to get analytics regarding the knowledge sharing ongoing in his mission, so he can improve the sharing.

3.8 Conclusion

In this chapter, our goal was to answer RQ2 and come up with a suitable interface and mechanism for privacy management. Following this research line, we proposed a privacy management model that is built on the metaphor of physical presence of an agent as a regular member of the interaction context. The proposed model answers RQ2 by providing an intuitive and contextual way to control privacy while enabling privacy awareness cues. By implementing the model in Graasp with AngeLA, we demonstrated how our model allows controlling privacy in knowledge sharing environments. Furthermore, the proposal was integrated into two European projects and evaluated with users, where we have obtained a GOOD usability rating for it. We expect that employing clear protective privacy policies in social media platforms would lead to higher level of trust and potentially adoption of the platforms.

Applications of the proposed privacy management model can be extended beyond the implementation discussed in this chapter to collaborative interaction platforms where there is a notion of context membership. Such membership can be available in virtual workspaces, chatrooms, discussion groups, virtual meeting rooms, etc. To illustrate the possible applications, Figure 3.7 and Figure 3.8 display mockups showing how the approach could potentially be realized respectively in Google Drive and Facebook. The model can be applied as well to humanitarian knowledge sharing, where traces are recorded to provide knowledge-related interaction analytics as explained in Chapter 2 and such recording may be not desirable in particular context.

Configurable Backend. Currently, the tracking agent settings are binary, the agent either sends the data to the analytics services or not. In the future, we plan to extend the agent by 1) allowing the configuration of the analytics services endpoint that the data is sent to,

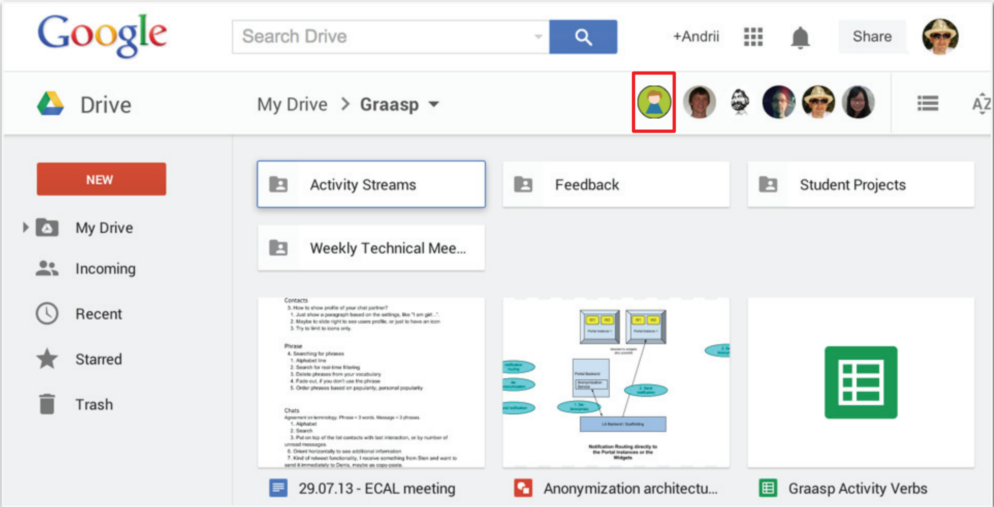


Figure 3.7 – A mockup of the proposed agent-based privacy management interface in Google Drive. Since AngeLA is a folder member, it can view the content and the activities there.

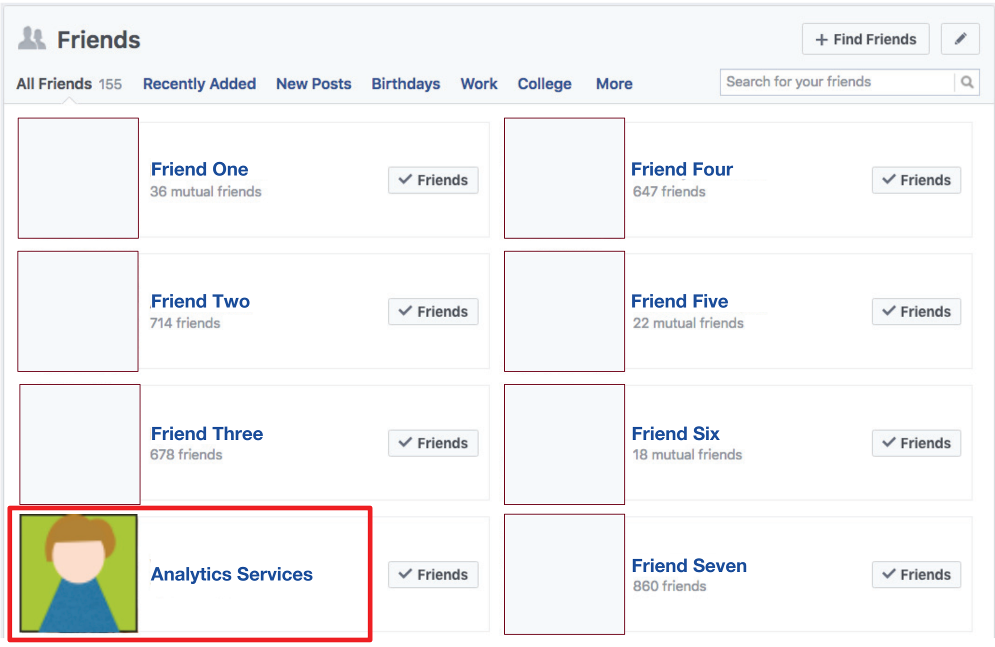


Figure 3.8 – A mockup of the proposed agent-based privacy management interface in Facebook. Since AngeLA is a friend, it can view user's activities and posted content.

2) enabling fine-grained user control over the types of data being sent, and 3) by providing options to allow user identity anonymization required in particular contexts. As example, using these settings the agent could be configured to send interaction traces from all inquiry spaces of a given school into the school's private learning record store.

Agent Conversational Interface. The settings highlighted in the previous paragraph can be provided to the user with a graphical user interface (GUI). Still, it may be promising to develop the metaphor of an agent further on by allowing the user to conduct conversations with the agent by sending messages to it and receiving textual responses. In this way, the user can ask the agent about the privacy policy currently in place and tell the agent how to change it. This approach is similar to how people talk in a room. Yet, it remains an open question if such a conversational way of defining privacy policy is beneficial compared to the traditional ones employing a GUI. Regarding the implementation in Graasp, since contextual chats are already supported by the platform, conversation with an agent can be implemented.

4 Knowledge Discovery with Interactive Interest-Based Recommender

Finding relevant content is one of the primary activities of users interacting with a content repository, be it knowledge workers using an organizational knowledge management system at a workplace or self-regulated learners collaborating in a learning environment. Due to the number of content items stored in such repositories potentially reaching millions or more, and quickly increasing, for the user it can be challenging to find relevant content by browsing or relying on the available search engine.

In this chapter, we address the problem by providing content and people recommendations based on user interests enabling relevant knowledge discovery. We propose an approach to building a user interests profile automatically by combining content analytics and activity tracking. The chapter also describes our implementation of the recommender system in Graasp and results of two evaluations. The evaluations demonstrated the ability of the approach to identify interests relevant to the user, and afterward to recommend relevant content. This chapter targets the third research question:

RQ3 - Discovery: how to enable discovery of knowledge relevant to user interests?

The content of this chapter was partially published in [112]:

A. Vozniuk, M. J. Rodriguez Triana, A. C. Holzer and D. Gillet. Combining Content Analytics and Activity Tracking to Identify User Interests and Enable Knowledge Discovery. 6th International Workshop on Personalization Approaches in Learning Environments (PALE), Halifax, Nova Scotia, Canada, July 13-16, 2016.

4.1 Introduction

Knowledge plays an essential role in value creation in the post-industrial economy. Knowledge is acquired and enriched by learning, which often takes place at a workplace or in an educational setting. While learning, people interact with content as a knowledge medium, located in various content repositories. In an educational setting, such content repositories are usually learning environments, where both students and teachers regularly interact with the content located there. The teachers interact when preparing a course by assembling relevant resources. The students interact when following a course by working with the materials provided by the teacher or found in the learning environment.

When working on a course in a learning environment, teachers should be able to benefit from the open content already available in the platform relevant to their teaching activities and potentially uploaded earlier by other teachers. Such content can include text files, web links, videos, and audio recordings coming from their device or the cloud. Moreover, it may also be beneficial for the students to have access to the content that is relevant to their interests, but which the teacher did not directly include into their course [117]. In the case of learning environments with a vast number of content items, it may be hard for the user to find content items corresponding to her interests (**challenge 3**).

To address the mentioned issues, we propose a recommender system that combines the content analytics, activity tracking, and information retrieval techniques to enable knowledge discovery. To evaluate the usefulness and the performance of the algorithm, we have implemented the approach in Graasp, a knowledge management system used in educational [10] and humanitarian [111] settings. Afterward, we have evaluated the approach with teachers, identifying their interests and providing them with recommendations.

This Chapter describes the algorithm used, the implementation details of the approach in Graasp and the evaluation of the approach with users. The structure of this Chapter is as follows. First, Section 4.2 reviews some of the relevant approaches to content analytics, activity tracking, and knowledge discovery. Afterward, Section 4.3 explains our approach to constructing user interests profile and demonstrates how we make recommendations based on the interests. Section 4.4 illustrates an implementation of the proposal, while Section 4.5 talks about the evaluation methodology and the results. Finally, Section 4.7 presents the conclusions and highlights directions for the future work.

4.2 Related Work

In this section, we review relevant work from the domains of content analytics, activity tracking, user interests mining and take a look at notable systems supporting knowledge discovery.

4.2.1 Content Analytics and Activity Tracking

Content Analytics. Content analytics allows the machine to gain an understanding of the content, similarly to how a human would do it by, among others, extracting the main topics, concepts, and entities present in the content. Kovanovic et. al. did an extensive overview of content analytics techniques since they are often employed in the domain of learning analytics [63]. For instance, in the line of our work, Bosnic et al. proposed to use automatic extraction of keywords from textual content as a foundation for content recommendations [13].

It is worth noting that existing papers focus mainly on analysis of textual content [63]. Recently, considerable progress was made in the understanding of multimedia formats, such as object recognition in images or videos [64], or automatic speech recognition. This allows broadening the scope of content analytics from purely textual information to supporting the various multimedia formats.

Understanding the content alone is not sufficient for understanding the learning since, according to Moore, learner-content interaction is a defining characteristic of education [84]. Moore argues that such learner-content interaction is necessary to happen for the education to take place since "it is the process of intellectually interacting with content that results in changes in the learner's understanding, the learner's perspective, or the cognitive structures of the learner's mind" [84]. Recognizing the importance of the interaction, below we consider approaches for capturing and persisting the interactions through activity tracking.

Activity Tracking. As covered in Chapter 2, there are multiple formats for structured representation of user activities including ActivityStreams used in [110], xAPI employed in [60] or IMS Caliper outlined in [97]. On a high-level, all these three formats record user-platform interactions in the format of the actor-verb-object triplet capturing who did what with what on the platform.

Combining Both. While there is a considerable number of papers employing content analytics or relying on interaction analysis, the number of studies combining both is still limited even though it was considered a promising direction [94, 63]. One noticeable recent proposal combining the both approaches is by Kim et. al. [58] where they use content analytics and recorded interaction data to understand better how students learn with video and eventually to improve their experience, for instance, by explaining better the identified confusing topics. Following this direction, we consider the combination of both content analytics and activity tracking as a core part of our proposal.

4.2.2 Mining User Interests

The obtained user interests can be used for different purposes, including recommendations and privacy awareness. As an example, Harkous et. al. proposed in [47] to employ a content analysis of the files located on Google Drive of a user to understand the topics, concepts, and entities relevant to the user. They used the obtained information with the goal to improve the user awareness through a new permissions model called Far-reaching Insights. This model informs the user about the insights that third-party applications can derive about her, based on the accessible Drive data given the requested permissions are granted. In our approach, we want to explore how identified interests can be used to provide the user with relevant content. In the following subsection, we review some of the systems enabling knowledge discovery with such recommendations.

4.2.3 Knowledge Discovery Systems

Klamma et al. have formulated a set of requirements for a collaborative adaptive learning platform [61]. One of the requirements is "Support for personalized learning resource delivery through an intelligent adaptive engine, being able to connect people to the right knowledge and deliver quality learning resources that are tailored to the learner's preferences and learning goals." [61]. Learning platforms often integrate such engine in the form of a recommender system. Drachsler et. al. have conducted an extensive review of 82 recommender systems used to support learning in [29]. Below, we review several proposals (some were not in the review), particularly relevant to our approach.

Zaldivar et. al. propose in [117] to collect the online resources visited by the students while completing their assignments that were not explicitly provided by instructors. In this way, the instructors can discover relevant learning resources that they possibly were not aware of, but that can be still beneficial for the students. In their approach, the authors record the web pages that students visit and perform a lexical analysis of the pages content. Afterward, they apply information retrieval techniques to identify the online content (web pages) that are the most similar to the content provided by the instructors as part of the course.

In [31] El Helou et. al. proposed a recommender system that considers user interactions with content items to construct a user-content associations graph. After the graph is built, the system applies a ranking algorithm to provide the user with personalized recommendations of relevant actors, activity spaces, and knowledge assets taking into account the context.

Motivated by the presented approaches, in the next section, we propose to design a recommender system that combines content analysis, activity tracking to identify user interests, and information retrieval techniques to suggest relevant content and people.

4.3 Interests-Based Recommender

To enable discovery of knowledge relevant to user interests and answer RQ3, we propose a recommender system that combines the content analytics, activity tracking, and information retrieval techniques to (1) build the user interests profile (**contribution 3.a**) and afterward (2) to suggest content relevant to the user as well as users with similar interests (**contribution 3.b**). A high-level idea of the approach is presented in Figure 4.1. To perform the recommendation, first, for each item available in the content repository, we employ natural language processing techniques to identify a set of concepts describing the content in a way similar to how humans would get a conceptual understanding of the text after reading it. Relying on high-level concepts instead of specific words present in the text when constructing user interests profile and afterward finding similar items, allows to identify the content that covers the same high-level concepts even if the specific words used in it are different.

Next, we analyze the interactions of the users with the content items based on available user activity recordings and aggregate the concepts in the content that the user interacted with building in this way the user interest profile. Finally, we use information retrieval techniques to recommend to the user relevant content based on the similarity between the concepts in the content and concepts identified as user interests. In the same way, our approach allows finding relevant users based on the determined interests similarity. Our approach puts the user in control of her interests profile and allows to adjust the interests by removing concepts if necessary, as in the case when the user is not interested anymore in some of the identified concepts. In this section, we explain in more details how each of the steps of our proposed approach works.

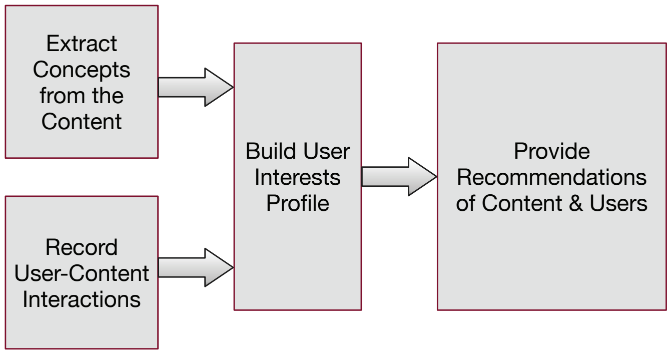


Figure 4.1 – A schematic representation of our knowledge discovery approach.

4.3.1 Identifying User Interests

To identify user interests, our system first needs to understand the concepts covered in the content. Second, it requires knowing how the user interacts with the content items through recorded user activities. Having both the identified concepts and the activities, the system can construct the user interests profile. Below, we explain each component of the approach.

Content Analytics. Content can be available in multiple formats; thus a data processing pipeline needs to be built for concepts extraction from multiple formats. Then, the extracted content and store them in an index for further use. A general representation of the key steps of the pipeline is shown in Figure 4.2. It is possible to see that for various content types (plain text, pdfs, Microsoft Word documents, images, etc.) should be processed in a different way to extract the concepts. For instance, we can immediately do the content analysis of the plain text files, while for binary text files such as pdfs, the textual content should be extracted first before the content analysis can be done. On the first step, textual content is extracted from the stored items, in the second step, the content analysis is performed leading to a set of concepts associated with the content. Finally, on the third step the extracted content and the concepts are tokenized and put into a searchable index so that they can be used in the recommendation step.

Activity Tracking. Our approach requires recording user-content interactions, namely the

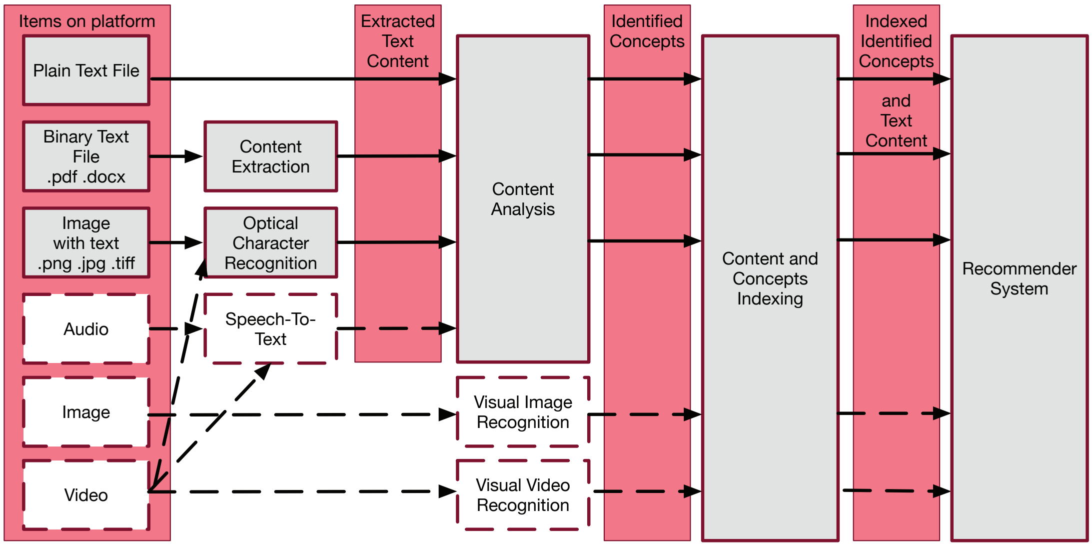


Figure 4.2 – A possible pipeline architecture to extract concepts from diverse content types. Dotted lines mark the parts yet to be implemented in Graasp.

triplet user-verb-object. We consider different types of interactions as a manifestation of different interest strength. For instance, intuitively when a user downloads the content it manifests a stronger interest in the content compared to just viewing it online. Our approach does not assume a specific activity recording technique or data format used, but it requires the approach to capture the user identifier, the verb indicating the type of interaction and, the identifier of the resource the user has interacted with.

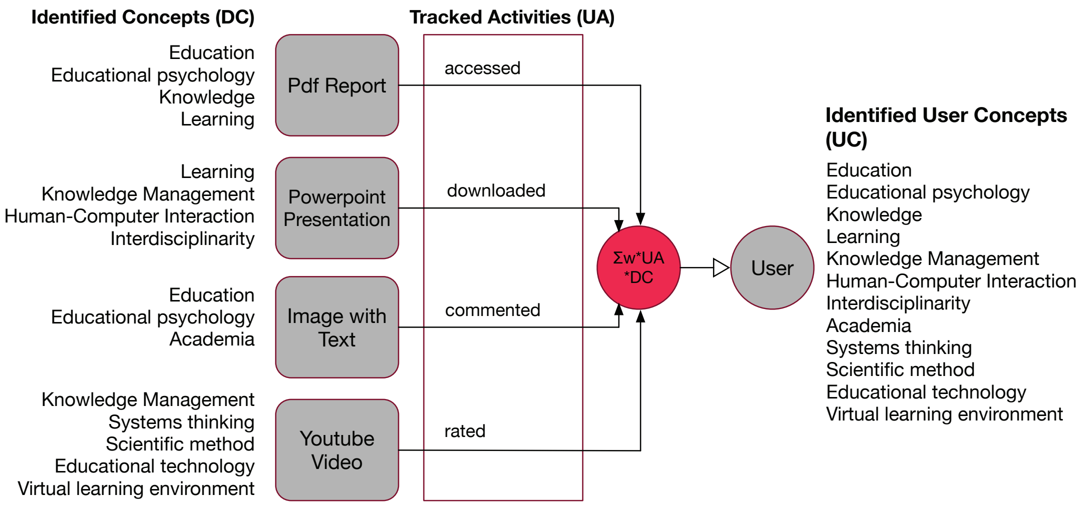


Figure 4.3 – A schematic representation of the proposed approach. The system aggregates the concepts from the content as the user interacts with the content.

Computing User Interests. As the user interacts with the content, the system aggregates the concepts identified in the content, weighting them according to the type of interaction as demonstrated in Figure 4.3. The aggregated concepts form the user interests profile shown in the right-hand side of Figure 4.3. Let's look into more details how the system can compute the user interest profiles at any point in time following the explained idea.

We denote by n the number of users on the platform, by p - the number of possible interaction types, by m - the number of content items on the platform and by k - the number of concepts identified in the content. Then at any point in time the user interests profiles can be computed in the following way:

$$UC_{n\times k} = \sum_{v=1}^{p} w_v UA^{v}_{n\times m} DC_{m\times k}, \tag{4.1}$$

where $UC_{n\times k}$ is the matrix of user concepts of interest, hence UC_{ij} is the relevance of the concepts c_j for the user u_i; w_v is the weight assigned to specific interaction type v indicating how strongly specific action of the user expresses her interest in the content; $UA^{v}_{n\times m}$ is the matrix capturing user-content interactions of type v so UA^{v}_{ij} is the number of times the user u_i has done interaction of type v with the content d_j; $DC_{m\times k}$ contains the concepts represented in the content so $DC_{f\times r}$ is the relevance of concept c_r to the content item d_f.

While the formula presented above is suitable for computing the profile first time when the recommender is deployed, the profile does not need to be recomputed from scratch and can be updated incrementally. On every user-content interaction, we update in real-time the user concepts of interest based on the ones that were found in the content as follows:

$$UC^{after}_{1\times k} = UC^{before}_{1\times k} + w_v UA^{v}_{n\times m} DC_{m\times 1}, \tag{4.2}$$

where $UC^{before}_{1\times k}$ is the vector of user concepts before the interaction and $UC^{after}_{1\times k}$ - after the interaction; $UA^{v}_{n\times m}$ is a matrix having 1 in position (i, j) if the user u_i had interaction of type v with the content item d_j, all other elements are 0; and $DC_{m\times 1}$ contains relevance values for the item concepts.

Once the profile constructed, in the next section we explain how it can be used for recommendations.

4.3.2 Recommending Relevant Content and Users

Connecting right people with right knowledge is a possible way to improve knowledge sharing. We aim to improve knowledge discovery by facilitating the creation of connections between knowledge sources and users in need of knowledge. Knowledge sources can be individual content items or other users with similar interests possessing the knowledge. We propose an approach that suggests 1) content relevant to users and 2) users with similar interest. Below, we present the two main steps of our approach.

Step 1. Computing term weights with TF-IDF. On the first step, we compute the relevance of specific terms (including concepts) for the content items by using a known information retrieval technique, namely term frequency - inverse document frequency (TF-IDF) as explained in [89]. In this way, for each content item we obtain a vector that contains weights of individual words or concepts cw_{ci} present in the content:

$$cw_{ci} = tf_{ci}idf_{ci}, \tag{4.3}$$

where tf_{ci} is the term frequency representing how often the term ci appears in the document and idf_{ci} is the inverse document frequency indicating how common is the term ci in all documents.

Step 2. Scoring relevant items with cosine similarity. To obtain for the user u suggested content items or relevant users, we compute the relevance score for the item d using the cosine similarity between the two vectors representing the user and the content:

$$S(u,d) = \frac{V(u)\cdot V(d)}{|V(u)||V(d)|}, \tag{4.4}$$

where $V(u)$ and $V(d)$ are the vectors containing weights of the user terms and the document terms computed in Step 1. $V(u)\cdot V(d)$ is a scalar product of the two vectors; $|V(u)|$ and $|V(d)|$ are Euclidean norms of the vectors.

4.3.3 Discussion

The recommendation algorithm explained in the previous subsection answers RQ3 by allowing to obtain ranked lists of content and users relevant to the user interests. Important advantages of our proposal compared to mainstream recommender systems include (1) transparency of the user model, (2) interpretability of the recommendations and (3) and interactivity of the system. The system can show the users what interests were identified and hence its model is transparent to the user. The recommendations are obtained based on the textual similarity of the interests and the concepts in the content and hence can be interpreted by the user. Finally, the system is interactive allowing the user to adjust the list of interests in her profile by removing the irrelevant ones to better match her preference at the particular moment.

4.4 Implementation

To validate the feasibility of the approach and further evaluate it, we have implemented it in Graasp. To enable our approach, we implemented extraction of text content from multiple file formats and built the activity logging infrastructure as explained earlier in Chapter 2. Also, we extended the platform to enable content analytics with concepts extraction, construction of the interests profile, and items recommendations with Elasticsearch[1]. Below we explain the architecture of the implemented solution.

4.4.1 Concept Extraction and Activity Tracking

Concept Extraction. The concepts extraction is done as soon as content is uploaded to Graasp. To extract the concepts, we have implemented the processing pipeline presented in Figure 4.2. First, the type of the content is identified, and Graasp tries to extract textual information when possible. For plain text files, it just reads the text content of the file. For binary text files including pdfs and Microsoft Office formats, we use the textract library. For images, Graasp tries to perform Optical Character Recognition, and read the text presented on the image using the tesseract[2] library. In the future, we foresee extracting text from Audio and Video files relying on Speech-To-Text technologies (shown with dotted lines in Figure 4.2) and obtaining concepts for images and videos with the help of visual recognition tools [64], for instance using clarifai[3]. Once the text is available, we analyze its content, identifying the concepts present there. For this purpose, we concatenate the item name, the item description, and the extracted content and employ AlchemyAPI[4] to obtain the concepts. After the system identifies the concepts, it indexes them in Elasticsearch together with the text content extracted before.

Activity Tracking. Graasp uses the ActivityStreams format for storing user activities on the platform. Some of the actions that the platform records include *access*, *download*, *add*, *remove*, *rate*, *comment*, *invite a member*, and *remove a member*.

4.4.2 Interests and Recommendations

Constructing Interests Profile. Graasp continuously updates interests profile of the users as they interact with the content. Users interests are displayed next to their profile information as

[1] Elasticsearch Open Source Engine https://github.com/elastic/elasticsearch (last accessed 10 November 2016)
[2] Tesseract Library https://github.com/tesseract-ocr (last accessed 10 November 2016)
[3] Clarifai https://www.clarifai.com/ (last accessed 10 November 2016)
[4] AlchemyAPI Concept Tagging http://www.alchemyapi.com/api/concept-tagging (last accessed 10 November 2016)

demonstrated in Figure 4.4. The user can adjust her profile by removing individual concepts by pressing the X button located next to the concept, and in this way influence in real-time the content and users suggested by the recommender.

Computing Recommendations. In Graasp, we rely on Elasticsearch for computing recommendations whenever the user wants to see them. Elasticsearch is built on the Lucene[5] text search engine that internally employs vector space model, TF-IDF, and cosine similarity when finding relevant items[6], similarly as in our proposed approach described in Section 4.3.2. We assemble all of the concepts and the terms from the user profile into a single search query. We run this query against the name, description and content fields assigning different boost weights for matches happening in different fields. The obtained results are presented to the user next to her profile as illustrated in Figure 4.4 (1).

4.5 Evaluation

Our goal was to understand opinions regarding the approach and its performance when put into practice. For this purpose, we have conducted two survey-based evaluations of the developed approach: 1) a preliminary one with pre-service teachers (student teachers that are about to start teaching) that allowed us to identify potential areas for recommender improvements early on (**validation 3.a**) and 2) an extended one with long-term Graasp users (**validation 3.b**). This section explains in more details the methodology used and the main outcomes.

4.5.1 Evaluation Methodology

Surveys are one of the common ways of evaluating recommender systems, allowing to collect opinions regarding the system from multiple users in a reasonable timeframe [32, 103]. Our goal was to validate if the approach is useful if its implementation in Graasp is usable, as well as if the system can actually identify relevant interests and recommend relevant items.

Survey Structure. For both surveys, we have used a questionnaire consisting of three parts. The complete questionnaire is available in Appendix B. In the first part, we have asked about general disposition regarding the interests identification and the interests-based recommender. The second part was the System Usability Scale (SUS) [17] evaluating the usability of the implemented system. We have chosen SUS among other usability surveys because of its well-understood interpretation and known robustness [6]. In the third part, we evaluated the quality of the identified interests and recommended content: how relevant it is to the user. Two types of questions formed the survey. The first type included questions to indicate the level of agreement with specific statements, where we followed the 5-point Likert scale ranging

[5] Apache Lucene https://lucene.apache.org (last accessed 10 November 2016)
[6] Theory Behind Relevance Scoring https://www.elastic.co/guide/en/elasticsearch/guide/current/scoring-theory.html (last accessed 10 November 2016)

Figure 4.4 – (1) The Interests sidebar displays the identified user interests next to his profile in Graasp. (2) The Suggested sidebar and (3) the Similar sidebar display recommended content and people based on the user interests.

from 1 - Strongly Disagree to 5 - Strongly Agree to obtain quantitative results [75]. The second type included open questions where we asked the responders to provide us with qualitative feedback regarding the approach and its implementation.

4.5.2 First Evaluation Results

The first survey was conducted with six participants of a workshop on inquiry-based learning for pre-service teachers in secondary education. During the workshop, the participants registered in Graasp and interacted with the platform to carry out a set of activities during 2 hours. At the end of the session, we asked the participants to fill in the survey. Below we discuss the results obtained for the three parts of the survey.

Approach. As summarized in Figure 4.5, the users valued positively the idea of using their interests to guide the recommendations (mean $\mu = 3.17$). They were as well still positive about the necessity of seeing their interests identified by the system ($\mu = 3.17$). Interestingly, at the same time the users wanted to be able to edit their interests ($\mu = 3.33$) validating the necessity to display the identified interests explicitly. When asked about the applications of the recommender, the users were positive about finding users by their interests ($\mu = 3.33$), but at the same time, they were rather negative about other users seeing their identified interests ($\mu = 2.67$). In addition to this quantitative analysis, during the workshop, the participants were keen on understanding how the interests were extracted and highlighted the novelty of the approach.

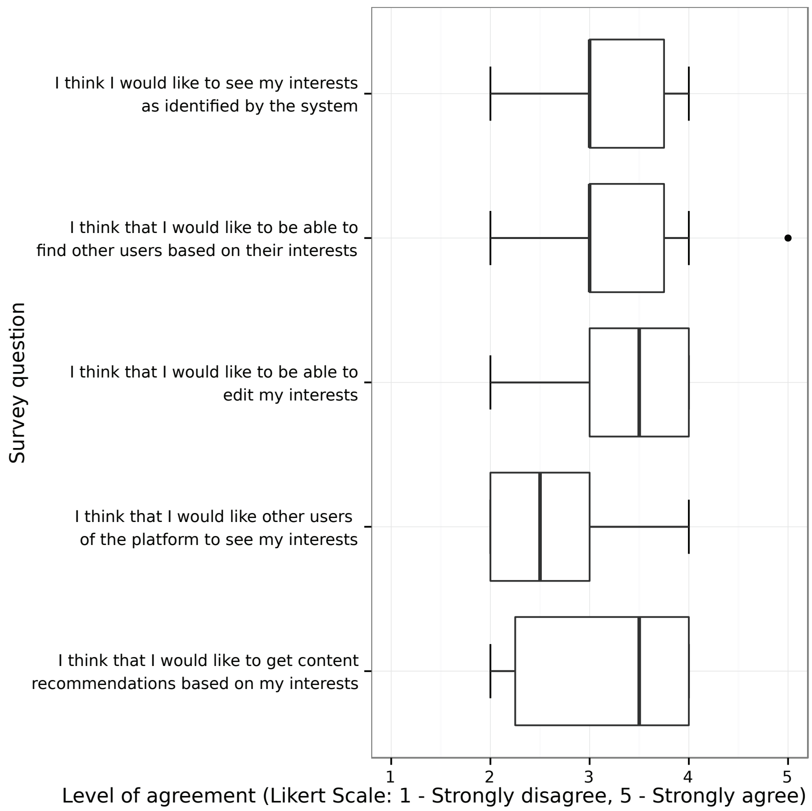


Figure 4.5 – A boxplot representing interests-based recommender perception in the first evaluation.

Usability. Based on the responses, the final SUS score of the recommender system in Graasp was equal to 63, meaning that the system has an OK usability rating [6]. The Figure 4.6 presents an overview of answers to individual questions of the SUS. The users reported that they were eager to use the recommender system frequently ($\mu = 3.33$), they did not find it unnecessarily complex ($\mu = 1.83$), they thought that it was easy to use ($\mu = 3.67$), they neither agreed nor disagreed if they would not need the support of a technical person ($\mu = 3.00$). Just one person responded explicitly that she would need technical support or previous background to use the recommender. According to the discussion continued with this person after the workshop, these answers seemed to be partially conditioned by her cognitive load. This is due to the short time available to get used to the platform itself and to integrate all of the ideas presented in the workshop. Furthermore, the participants found various functions to be rather well integrated ($\mu = 3.17$), they did not find too much inconsistency ($\mu = 2.00$) they would imagine most people learn to use it quickly ($\mu = 3.17$). It was not cumbersome to use ($\mu = 2.00$), they felt rather confident using it ($\mu = 3.17$) without the need to learn a lot of things before getting going ($\mu = 2.50$).

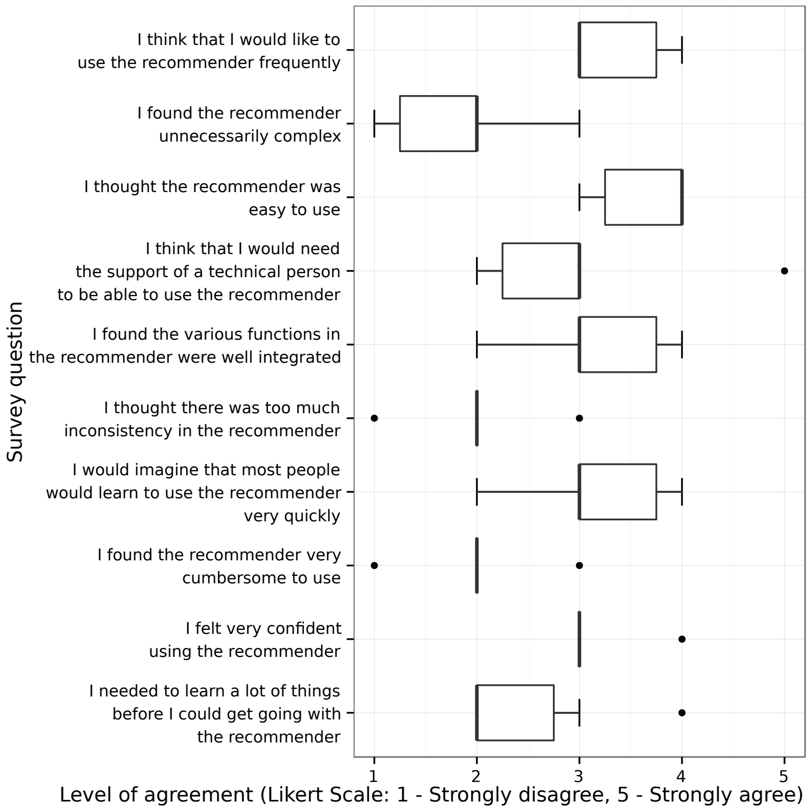


Figure 4.6 – A boxplot representing system usability scale questionnaire results for the first evaluation.

Accuracy. Despite the limited amount of traces collected due to the short time of interaction with the platform, the results outlined in Figure 4.7 point out that both the extracted interests and recommended items were relevant ($\mu = 3.17$ and $\mu = 3.67$). Although we did not target specifically the item diversity when designing the recommender, the participants found the recommended items to be rather diverse ($\mu = 3.50$). It is noteworthy that when we asked the users to check how many relevant interests and recommendations appeared in the top 10, we discovered two clear groups. While most of the users reported more that six relevant items, two users obtained less than two relevant items. In general, the participant considered that it was useful for them to be able to get recommendations based on interests as done with our approach ($\mu = 3.67$).

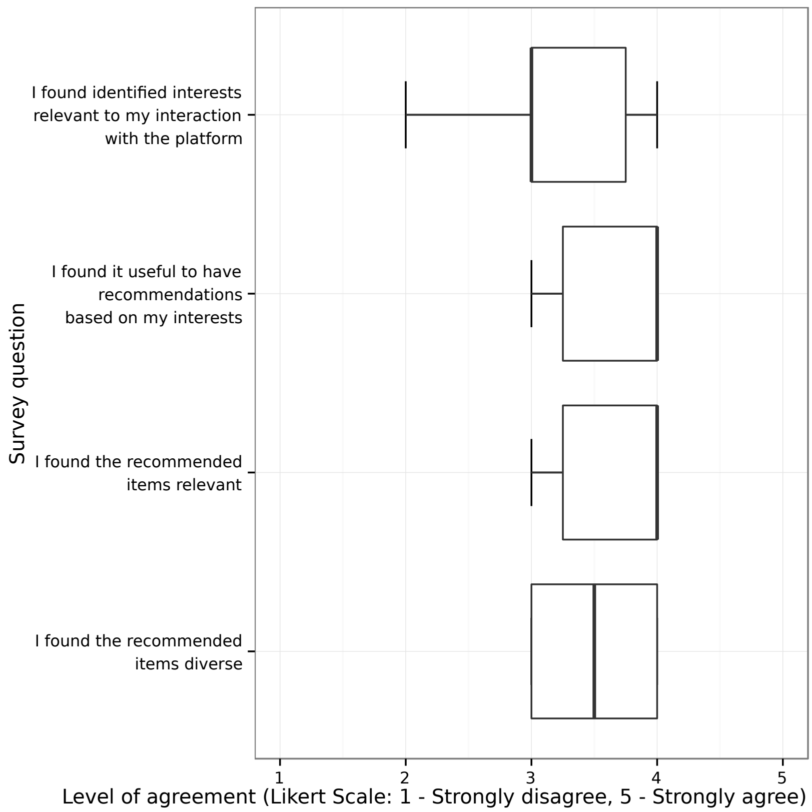


Figure 4.7 – A boxplot representing interests-based recommender accuracy in the first evaluation.

Privacy Implications. Right now, the interests are visible only to the user, but we consider putting in place a mechanism to allow users to make their validated interests visible to other users of the platform and to make it possible to find the user based on her interests as it was proposed in [88]. However, based on the comments from the participants, although some of them were eager to make their interests visible, others were reluctant. Thus, it will be necessary to allow users to configure the visibility of their interests to preserve their privacy, following the recommendations provided in codes of practice for Learning Analytics [101].

Sensitivity to Inaccurate Concepts. In the case when concepts were inaccurately identified for an item with which a user has frequently interacted, these concepts will appear at the top of the user interests list. To improve the quality of identified concepts and reduce the sensitivity to inaccurate concepts for the second evaluation we decided to put in place two conditions when aggregating concepts: (1) take into account only concepts with relevance more than 0.8 out of 1 and (2) consider only concepts that were encountered at least in three content items. The first condition allows to remove concepts with low relevancy and the second should reduce the influence of a single misidentified concept, by requiring that this concept appears at least in three items. We experimented with the number of required items by varying it from two to six. We did not observe improvements in the quality of remaining concepts when using values more than three, while the number of the concepts was decreasing with the increase of the number of required items.

After the corresponding changes had been made in the recommender implementation, we have conducted the second evaluation with more users that exploited Graasp for longer periods of time so that more activity traces were available.

4.5.3 Second Evaluation Results

The second evaluation aimed to collect feedback from long-term Graasp users for which the system has a considerable number of interactions recorded. Our expectation is that these users have a more established vision regarding the types of their activities in the system. As in the first evaluation, the survey consisted of three parts: (1) disposition regarding the recommender; (2) system usability scale; (3) quality of recommendations.

To collect the responses, we have sent an email with the survey to all of the Graasp users that had more than 5 interests identified in the profile. We explained to them the idea of the recommender and asking them to fill in an online survey form. In total, we have collected 40 responses (22 - female, 16 - male, 2 - preferred not to answer; median age 33 years).

Approach. Figure 4.8 summarizes the results of the first part of the survey. The users valued positively the idea of using their interests to guide the recommendations (mean $\mu = 4.00$). They were positive about the necessity of seeing their interests identified by the system ($\mu = 3.65$). At the same time, the users wanted to be able to edit their interests ($\mu = 4.12$) making it necessary to display the identified interests explicitly. When asked about the applications of the recommender, the users were positive about finding users by their interests ($\mu = 3.33$), and differently from the first survey were positive about other users seeing their identified interests ($\mu = 3.60$).

Usability. Based on the responses in the second evaluation, the SUS score was equal to 70, which is 7 points higher than in the first one. This means that the system still has an OK usability rating, and just a few points below the GOOD usability rating [6]. The Figure 4.9 presents an overview of answers to individual questions of the SUS. The users reported that they were more eager to use the recommender system frequently compared to the first survey

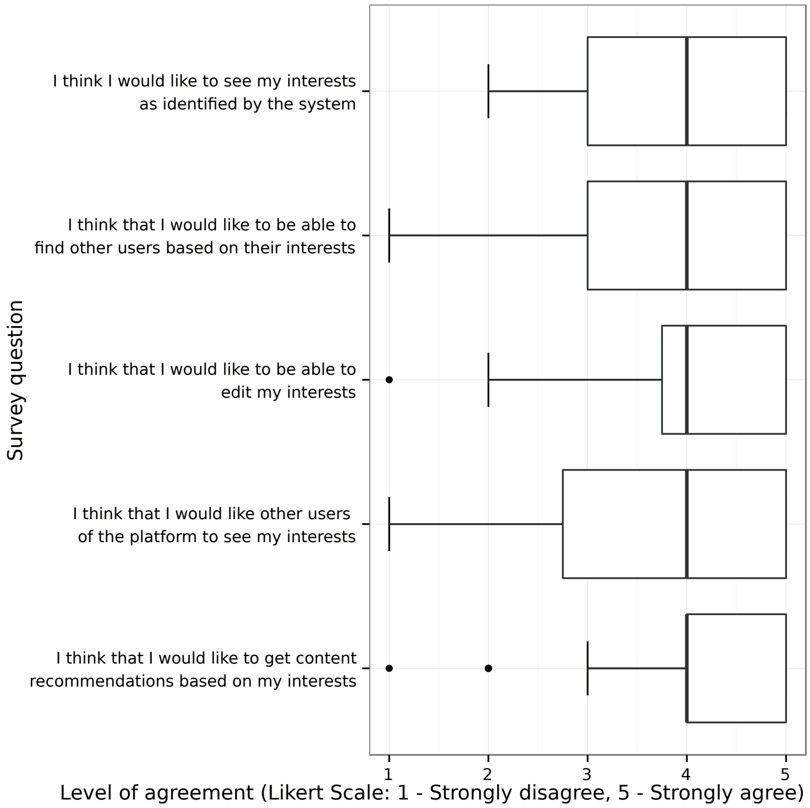


Figure 4.8 – A boxplot representing interests-based recommender perception in the second evaluation.

($\mu = 3.50$ vs. $\mu = 3.33$ in the first survey), they did not find it unnecessary complex ($\mu = 2.00$ vs. $\mu = 1.83$), they thought that it was easy to use ($\mu = 4.03$ vs. $\mu = 3.67$), they disagreed that they would need the support of a technical person ($\mu = 2.42$ vs. $\mu = 3.00$). Furthermore, the participants found various functions to be rather well integrated ($\mu = 3.67$ vs. $\mu = 3.17$), they did not find too much inconsistency ($\mu = 2.45$ vs. $\mu = 2.00$) they would imagine most people learn to use it quickly ($\mu = 3.17$ vs. $\mu = 3.17$). Our solution was not cumbersome to use ($\mu = 2.00$ vs. $\mu = 2.00$), they felt rather confident using it ($\mu = 3.17$ vs. $\mu = 3.80$) without the need to learn a lot of things before getting going ($\mu = 2.15$ vs. $\mu = 2.50$). In summary, the second evaluation demonstrated better results in all SUS question except the complexity and inconsistency, where the results were worth. This indicates directions for further investigation and improvement.

Accuracy. The results shown in Figure 4.10 indicate that both the extracted interests and recommended items were relevant ($\mu = 3.70$ and $\mu = 3.52$). The interests identification has improved from $\mu = 3.17$ in the first evaluation to $\mu = 3.70$ in the second one, possibly due to our improvements addressing the interests identification issue explained earlier. The users still agreed that the recommended items were relevant, but the level of agreement has slightly

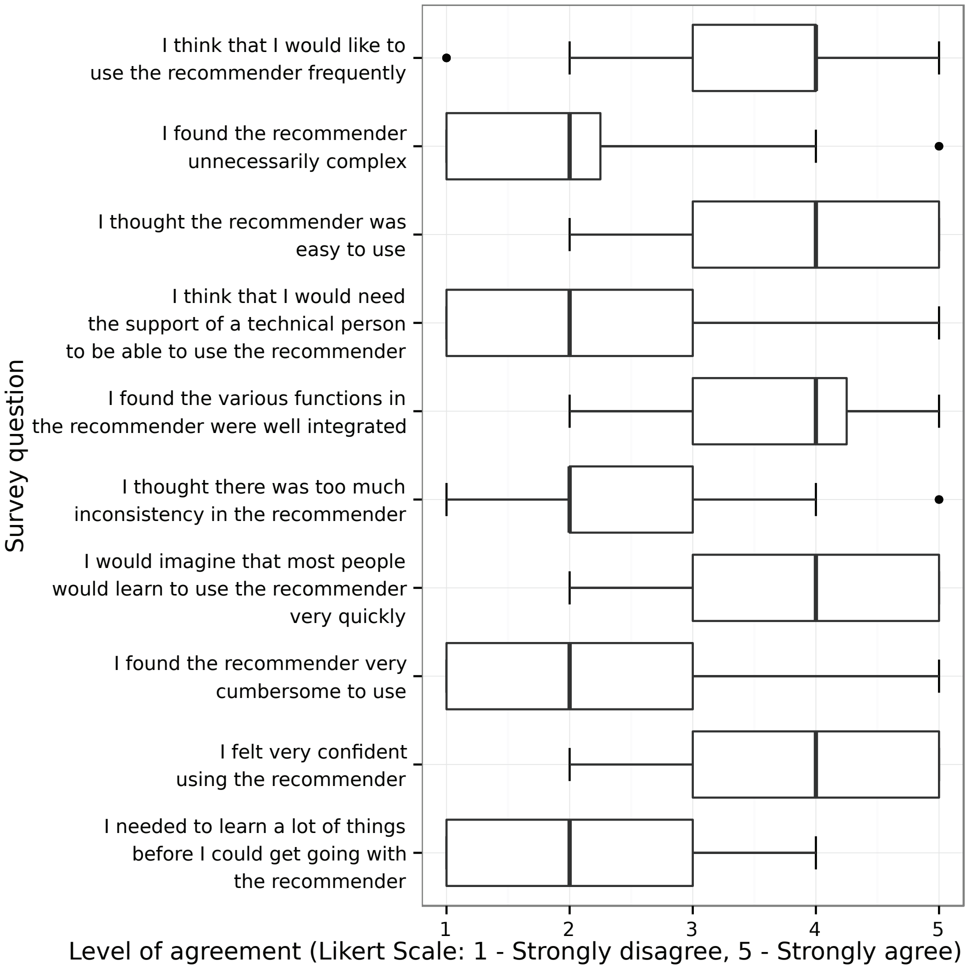


Figure 4.9 – System usability scale questionnaire results in the second evaluation.

decreased from $\mu = 3.67$ in the first evaluation to $\mu = 3.52$ in the second one. Similarly, while the interest-based recommendations were still perceived useful, the level of agreement slightly reduced from $\mu = 3.67$ in the first evaluation to $\mu = 3.60$.

In summary, the second evaluation has demonstrated better results in most of the aspects compared to the first one. The comparison of the results of the evaluations together with their relative change is presented in Table 4.1. The values in the table in bold indicate a positive change.

4.5.4 Actual Usage

In addition to collecting user feedback with the surveys, to understand if the recommender is being used, we have recorded user interactions with the recommender (**validation 3.c**). The recording period lasted from the 6th of May 2016 till the 10th of August 2016 (a bit more than

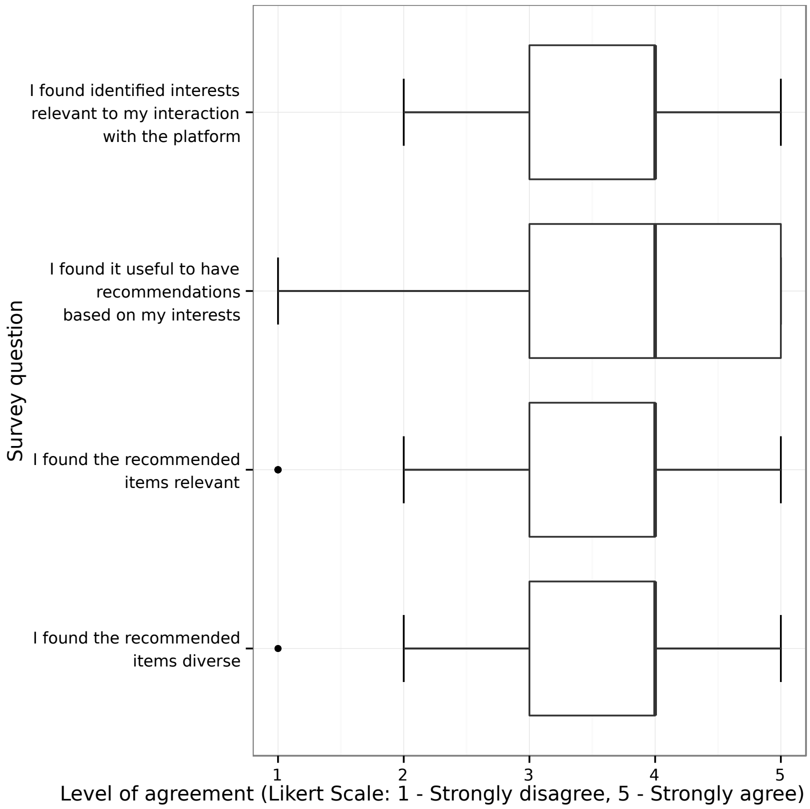


Figure 4.10 – A boxplot representing interests-based recommender accuracy in the second evaluation.

three months). During this period, 135 users have accessed the Interests tab for 363 times in total, the Suggested content tab was accessed by 115 users for 295 times and the Similar users tab was opened by 103 users in total 211 times. The users have as well adjusted their interests profile, in total 185 interests were removed by 22 users. These numbers indicate that the recommender is being adopted by the regular Graasp users.

Our goal was as well to understand for what percentage of users it was possible to generate an interests profile in practical settings. Out of 10210 users registered on Graasp as of sixth of May 2016, the recommender was able to identify interests for 707 (7%) users. We have looked closer into the data, and the primary reasons for this rather low number are inactivity of most of the users or their interaction with types of content (e.g. pictures and video) that the system was not able to identify concepts for at this stage. Out of 707 users with the identified interests, 135 users (19%) have accessed the interests tab and 22 (15%) out of them have removed at least one interest not corresponding to their real interests.

Question	Preliminary Evaluation Values	Long-term Users Evaluation Values
I think that I would like to get content recommendations based on my interests	3.17	4.00
I think I would like to see my interests as identified by the system	3.17	3.65
I think that I would like to be able to edit my interests	3.33	4.12
I think that I would like to be able to find other users based on their interests	3.33	3.75
I think that I would like other users of the platform to see my interests	2.67	3.60
I think that I would like to use the Interests-based recommender frequently	3.33	3.50
I found the Interests-based recommender unnecessarily complex	1.83	2.00
I thought the Interests-based recommender was easy to use	3.67	4.03
I think that I would need the support of a technical person to be able to use the Interests-based recommender	3.00	2.42
I found the various functions in the Interests-based recommender were well integrated	3.17	3.67
I thought there was too much inconsistency in the Interests-based recommender	2.00	2.45
I would imagine that most people would learn to use the Interests-based recommender very quickly	3.17	3.95
I found the Interests-based recommender very cumbersome to use	2.00	2.00
I felt very confident using the Interests-based recommender	3.17	3.80
I needed to learn a lot of things before I could get going with the Interests-based recommender	2.50	2.15
I found identified interests relevant to my interaction with the platform	3.17	3.70
I found the recommended items relevant	3.67	3.52
I found the recommended items diverse	3.50	3.62
I found it useful to have recommendations based on my interests	3.67	3.60

Table 4.1 – Comparison of the results from the two evaluations. The values presented in bold indicate a positive change.

4.5.5 Discussion

In this section, we conducted two evaluations of the recommender system proposed in Section 4.3. The evaluations unveiled advantages of the system as well as some of its limitations. First of all, it is worth mentioning that based on usage data the system was adopted by Graasp users, thus answering RQ3 for them. In private discussions, some of the users highlighted that they liked to see their interests explicitly listed and being able to adjust them. This requirement was as well confirmed in the evaluation surveys.

Two survey-based evaluations demonstrated that the system answered RQ3 since it was able to suggest relevant content to new Graasp users with a limited set of recorded activities as well as to long-term users. However, it is fair to highlight that in some cases our system constructed inaccurate user profiles leading to not so relevant recommendations due to misidentifying concepts present in the content. We addressed this issue by introducing two heuristics that aimed to increase the reliability of the identified concepts. A data-driven evaluation is required to understand their impact fully.

4.6 Implications for Humanitarian Settings

In this chapter, we focused on knowledge discovery in an educational setting. Similarly to the educational setting, proper knowledge sharing is essential for organizational learning in humanitarian contexts. This statement is particularly valid for humanitarian organizations, which are known to be heavily dependent on efficient knowledge management which is recognized to be key to their operations [69, 79]. In such organizations, whenever knowledge workers need to get access to knowledge relevant to their current tasks, they may interact with content produced by other organization members and located in knowledge sharing systems [59].

Due to the large size of current knowledge repositories in humanitarian organizations, the problem of getting relevant knowledge, including its discovery, is particularly pertinent. Our experience with MSF shows that it is not even sufficient for knowledge workers to have access to an advanced search engine or having the content structured following a well-defined taxonomy. Hence, the challenge of knowledge discovery needs to be addressed in humanitarian organizations as well. The mentioned issues make it necessary for the knowledge sharing system to assist the user in knowledge discovery by providing the user with content relevant to his or her interests.

By applying our proposed knowledge discovery approach presented in this chapter to a humanitarian setting, it is possible to build an interests profile of each of the organization members interacting with the knowledge repository and to recommend relevant content and people possessing relevant knowledge, as was explained earlier on in this chapter. Benefiting from the approach we proposed in this chapter, Amy, the doctor from MSF treating sleeping sickness, can discover content related to her mission that was created by her colleagues from other missions but was not explicitly included by the HQ into her workspace. Additionally, Amy may check the recommended content from time to time to see if any new information relevant to her interests appears at MSF.

4.7 Conclusions and Future Work

In this chapter, we aimed to answer RQ3 exploring how to enable discovery of knowledge relevant to user interests. We proposed an interactive interests-based recommender that constructs the user interests profile based on 1) content analytics providing the system with the concepts present in the content and 2) activity tracking allowing the system to know how the user interacted with the content. Afterward, the recommender uses the profile to recommend content and other users relevant to the user interests enabling in this way knowledge discovery and answering RQ3. There are two important benefits of our approach, namely interpretability and interactivity. The user can see the interests identified by the system, understand why specific content is being recommended according to these interests and adjust the interests list to match actual user preferences better.

We have implemented the proposed approach in Graasp that was used in a workshop to support teachers when building inquiry learning spaces for their students. We have conducted a preliminary evaluation during the workshop and also evaluated the approach with long-term Graasp users. The evaluations have demonstrated that our approach can identify relevant user interests and recommend relevant content based on the identified interests. At the same time, the evaluations have unveiled sensitivity of the approach to inaccurately identified concepts that could be further addressed in the future.

While we draw our experience and motivation from the educational context, our contributions have a broad impact and can be applied to content repositories, where it is possible to obtain content analytics and tracks of activities performed by the users (e.g., Google Drive and Dropbox).

Looking Outside. In this study, we analyzed the content and recorded the activities limited to the scope of the content repository. However, in the current technological landscape, the interactions are getting more distributed often spanning across multiple platforms. Studies suggest that combining data obtained from several platforms could allow obtaining a more accurate user interests profile [46]. In the future, we plan to extend the architecture of Graasp to index and analyze the content located outside of the system and track interactions with it.

Incorporating Relevance Scores. At the moment, when computing the similarity score for the recommended items we consider the concept presence but do not take into account the concept relevance scores. Though we take into account the scores at an earlier stage, when considering only the concepts with scores of 0.8 and more. Incorporating the concept relevance scores available for user interest concepts and content concepts when computing the user-content relevance score may lead to better quality of recommendations since it will promote the results with similar highly relevant concepts.

Recommender Adaptability. One potential downside of our approach could be related to the limited ability of the approach to react timely to change in the user interests reflected in her interactions. This is because the concepts that the user accumulated at some point through her interaction history maintain the same score indefinitely. One of the possible solutions to this problem is to introduce the forgetting function as suggested in [106], so that as time goes the concepts that are not encountered anymore get their relevance score reduced. With the current implementation of our approach, the user is still in control and is able to manually remove interests from her profile when they become irrelevant.

Content Types. One promising direction of improving the recommender is extending the types of content that the system can understand, track interactions with, and consequently recommend. Such types of content may include videos, where it is possible to do automatic speech recognition, or automatically identify objects present in the scene. As well certain kinds of content may allow fine-grained interaction tracking. For instance, instead of just tracking that the user has watched a video or has opened a presentation, it is possible to record specific pages visited, or specific parts of the video watched as well as time spent on individual pages. Having an understanding of concepts present in the specific parts of the content that the user has interacted with, may further refine the interests identification. At the same time, such fine-grained tracking further highlights the importance of taking into account the privacy dimensions addressed in Chapter 3.

5 Mobile Knowledge Delivery into Underconnected Environments

Knowledge is regarded as one of the most valuable resources available in humanitarian organizations [79]. Hence, organizations need to be able to access critical knowledge timely and reliably and build and share knowledge efficiently between the headquarters (HQ) and multiple, often geographically dispersed, field teams. At the same time, the field conditions are challenging and often lacking a reliable Internet connection, especially for organizations that operate in crisis situations. This chapter aims to overcome the knowledge delivery issues by addressing fourth research question:

RQ4 - Delivery: How to improve access to the Internet-located knowledge from within underconnected environments?

To answer this research question, this chapter proposes a novel content delivery model for underconnected environments. We propose to use (1) a peer-to-peer synchronization middleware for content delivery and (2) low-cost computers as intermediate nodes deployed in the field. Such nodes can serve as peers in the peer-to-peer delivery network and in addition provide access to the content by running a local webserver accessible via a direct WiFi connection. Since the data is replicated locally on the local server, it is fast to access. Thanks to peer-to-peer protocol the data is synchronized once to the device and then distributed to all other personal user devices (smartphones, laptops, tablets) locally hence saving the costly external bandwidth. Due to low power consumption of the device, it can run on a battery, and due to its small pocketable size it can be taken where the content is required (e.g. a remote hospital in a rural area) not depending on the network and power grid.

The content of this chapter was partially published in [111]:

A. Vozniuk, A. Holzer, S. Govaerts, J. Mazuze and D. Gillet. Graspeo: a Social Media Platform for Knowledge Management in NGOs. Seventh International Conference on Information and Communication Technologies and Development (ICTD 2015). Singapore, Singapore. May 15 - 18, 2015.

This chapter is structured as follows. In Section 5.1 we discuss our learnings from the field experience and state requirements for a content delivery system. Section 5.2 discusses related work addressing the challenges of underconnected environments. Section 5.3 presents our proposal and Section 5.4 describes the implementation of our general approach in Graasp. Section 5.6 presents evaluation results based on tests done in the lab and our experience of deploying the solution in the field in eight MSF missions during five field trips. Finally, Section 5.8 wraps up with conclusions and future work.

5.1 Background

To understand the knowledge delivery requirements and come up with a suitable approach, we have worked closely with MSF. MSF staff typically need to access essential documents during field trips. These essential documents cover thematics (i.e., malaria, sanitation) as in Figure 1.3 and geopolitical subjects (i.e., the situation in Yemen) and are frequently updated by the HQ staff as new information appears.

At the same time, the Internet connection type, speed, and reliability impacting the content access experience can differ substantially from mission-to-mission (**challenge 4**). As an example, Figure 5.1 shows the setup used to provide Internet access in the Kampala mission in Uganda. In Kampala, the mission is connected to the Internet with a 3G modem and afterward the connection is provided in the mission via a WiFi access point. MSF mission members access Internet by connecting to the access point as presented in Figure 5.2.

Figure 5.1 – A 3G mobile Internet stick mounted on the wall provides Internet access in the MSF mission in Kampala, Uganda.

Figure 5.2 – MSF mission members accessing the Internet through a local WiFi access point in the Kampala mission, Uganda.

5.1.1 Delivery Requirements

In order to better understand the requirements for a knowledge sharing system including field knowledge delivery, our colleagues from MSF conducted in-depth field studies that consisted of a total of 145 hours of interviews in Geneva, Niger and Swaziland between June and August 2013 (see more details in Section 1.2.2). Relying on this process, we have identified together with MSF key field delivery requirements listed below for a suitable field content delivery system:

Req 6.1 – Enable Content Access Offline: Situations without Internet and Power Supply are still common in rural areas [55]. Hence, the system should be capable of operating in an entirely autonomous mode.

Req 6.2 – Provide Fast Data Access: High latency when working with information technology is known to be a source of user frustration [22] and can lead to user inefficiency hindering the technology adoption [57]. Slow knowledge access is particularly severe in emergency cases requiring fast decision-making. This requirement makes it necessary for the system to enable content access with minimal latency.

Req 6.3 – Be Bandwidth Efficient: As mentioned earlier, the field staff often relies on a satellite or mobile Internet connection. These types of connections are usually costly, have high latency and limited bandwidth. Addressing these limitations, the system should be designed to use as little bandwidth as possible.

Req 6.4 – Be Portable: The systems should be deployable wherever it is needed, outside of the main base and on the road.

Req 6.5 – Enable Content Availability on the User Device: According to our field experience, most of the field staff have a personal mobile device prefer to have content on the device, so it is available to them any time wherever they go.

Req 6.6 – Provide Up-to-date Information: The system should update the content whenever it is possible, in order to provide access to the latest version available.

Req 6.7 – Require No Infrastructure Change: Changing IT infrastructure on the scale of MSF is labor-demanding and costly. The required solution should be just drop-in without the need to redesign the local IT infrastructure already in place.

A system meeting these requirements addresses the key challenges when accessing content from the field, namely latency, bandwidth and the data cost and necessity of being physically present in a location where the network and power grid connections are available.

5.2 Related Work

When sharing knowledge in areas with broadband Internet access, organizations can rely on mainstream cloud services such as *Evernote*[1], *Google Drive*[2] and *Dropbox*[3] that combine social media and knowledge sharing features. Unfortunately, these services are suboptimal in underconnected environments where humanitarian organizations operate since these services have a centralized cloud-based infrastructure which requires the clients to download the data from the cloud. There is currently a lack of compelling mainstream solutions enabling knowledge delivery into underconnected areas.

When content is delivered from an Internet server to an end-user device in the field, be it a tablet, a smartphone or a computer, it typically passes through a number of steps schematically represented in Figure 5.3. Existing research literature offers three main groups of solutions for knowledge delivery. Solutions of **the first group**, rely on mechanical content delivery (e.g., by a car or a helicopter) on a portable storage in case there is no network connection available at some part of the delivery chain in Figure 5.3. For instance, some humanitarian agencies rely on portable storage technologies, such as USB flash drives. Field staff would download the relevant content to the portable storage before going to the mission, then take it to the mission

[1]Evernote https://evernote.com/ (last accessed 10 November 2016)
[2]Google Drive https://www.google.com/drive/ (last accessed 10 November 2016)
[3]Dropbox https://www.dropbox.com (last accessed 10 November 2016)

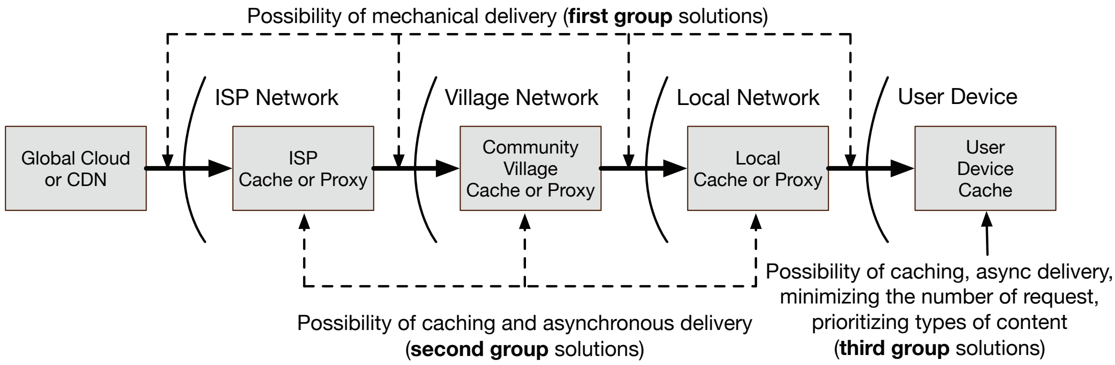


Figure 5.3 – A schematic representation of a typical content delivery chain from a server on the Internet to an end-user personal device in the field. Three groups of solutions are highlighted.

and use it offline. This type of technology has also been proposed to update software in the field [23]. Such solutions require manual upfront content download and afterward manual update and usually can be accessed by a single user at a time.

Solutions of **the second group**, target the delivery problem when the network is available but is of poor quality, for instance, unreliable or slow. These solutions are based on intermediate servers (e.g., [70], [55], [74]) and caching. In such settings, a local server plays the role of a proxy or relay to the Internet. When the connection is available, it caches the downloaded information so that it can be easily accessible locally. One of the common drawbacks of these solutions is that the hardware and software involved often make them not portable and inoperable in the field. Solutions of **the third group** aim to improve the end-use experience by applying techniques on the personal user device, for instance by using caching inside of mobile apps or by relying on browser local storage in case of web apps (e.g., [102]). Below, we review existing solutions in more details and highlight their advantages and limitations.

5.2.1 Mechanical Delivery with Portable Storage

Historically, mechanical content delivery was done with various media depending on its cost and availability. Just during the last two decades first, the CDs were employed for delivering content. Afterward DVDs and more recently USB memory sticks were used. Hereafter, we outline some of existing cases of mechanical delivery relying on various types of storage medium.

CDs and DVDs. DVDs were used in [4] for video content delivery for instruction in a low resource classroom. In 2010, DVD players and TVs were already widespread in developing areas, so it was reasonable to use interactivity features provided by DVD players to deliver interactive encyclopedias, language tutoring materials, and medical decision systems on a DVD without the need of having a computer which was considerably more expensive at that time. The core idea of that project was for local teachers to use for instructions in rural schools video content produced by more experienced teachers from better schools. This approach should expose the students to better quality content and in theory, improve their learning outcomes. The authors deployed Digital StudyHall (DSH) software making educational videos available on interactive DVDs having two applications: (1) displaying a PowerPoint presentation and (2) showing books for children. These two applications were evaluated in cases studies done in Indian schools. Based on the experience, the authors conclude that it may be reasonable to deliver high-quality cloud-based MOOCs to low-resource schools to study them together and improve the learning outcomes. In the case of MSF, the content delivery problem was first approached by issuing a CD to an MSF employee before going to the field with prerecorded files relevant to the particular mission. Later, a USB Stick with the files was given.

USB sticks. Differently from practically only readable on not writable CDs and DVDs, USB sticks are writable. This feature is employed by some existing proposals to allow updating the content delivered with a USB stick. For instance, MSF uses a system called LogKey to store relevant content on a USB stick and access the content by running a web application on a user computer. The web application has an option to update the content when an Internet connection is available if the user manually clicks the Update button. Another USB stick delivery example is presented in [23], where the authors propose to use USB flash drives to distributed software updates (in their case anti-virus bases). The insight is that since the sticks move around and are plugged into multiple computers, they can serve as a way to deliver software updates. When a drive is plugged into a computer with the FlashPatch software installed, it offers to write software updates to the drive and automatically monitors and applies new software updates delivered on the stick. With this approach, the updates will move around naturally as drives get plugged into computers.

HDDs. When the content size is becoming large, it is reasonable to employ hard disk drives (HDDs) having bigger capacity. For instance, HDDs were used in Cuba to deliver content locally without the Internet[4]. Mechanical delivery of HDDs can also be a reasonable option if the Internet connection speed is too low compared to the size of the content being transferred (e.g., AWS ImportExport Snowball[5]).

The described mechanical delivery solutions can be used in situations when no network is available at all, and the only option is to deliver content mechanically. The typical small form factor of the storage makes it easy to carry around. Unfortunately, most of the discussed storage solutions are challenging or not possible to connect to mobile phones or tablets since they usually lack a USB port or require special adapters for the connection. Another limitation is related to the fact that such storage solutions can usually be accessed only by a single user at a time since the access requires to connect the storage to the device. The solutions are also limited to providing access only to the content, while it can be beneficiary for the users

[4]Cuba Offline https://www.theguardian.com/world/2014/dec/23/cuba-offline-internet-weekly-packet-external-hard-drives (last accessed 10 November 2016)

[5]AWS ImportExport Snowball https://aws.amazon.com/importexport/ (last accessed 10 November 2016)

to have access to services built on top of the delivered content. Finally, the content on the storage needs to be updated manually to the latest version every time the Internet connection is available. We aim to keep the highlighted benefits and address the mentioned limitations in our proposal.

5.2.2 Intermediate Servers

When the network is unreliable or slow at some part of the delivery chain in Figure 5.3, a common approach used by a number of solutions is to deploy an intermediate server that will cache the data close to the end-user allowing several users to benefit from already downloaded data. The server can as well allow handling requests asynchronously downloading content on behalf of the user allowing her to come back when the downloading is done and saving the waiting time. Below, we considered several of existing solutions in more details.

TroTro [70] is a web application running in the browser that integrates with a shared proxy server. The web app can operate in three modes: (1) online (2) poor connectivity and (3) offline. In the poor connectivity mode, the web pages being downloaded are added to a queue, allowing asynchronous download and making the user aware of the downloads in progress and expected waiting time. When accessing the Internet content through the app, the data is being cached on a shared proxy server possibly benefiting the whole local community. In the offline and poor connectivity modes the user can navigate and search through the content available locally (downloaded before to the proxy) making it possible to find and browse some pages in zero-connectivity settings. The users are informed of the current mode through clear indicators in the web-interface. TroTro is particularly suitable for communities that share the same interests (for instance, cafes and school classes) due to increased chance of cache hits.

Kwaabana [55] addresses the problem of unreliable, slow and costly Internet in developing areas when sharing content via a social networks, in their case Facebook. Kwaabana consists of two parts: (1) a global Kwaabana server where the files shared outside of the village are put and (2) a village server storing the files shared locally between the members of the same village (located on the same network). Based on the social graph, the system can identify to which of those servers a particular shared file should be uploaded in the end. The files are always first uploaded to the village server, but if the sharing is with an external member, then the files are transferred to the global server. Kwaabana employs several techniques when transferring files. It uses a queue to order the uploads and make sure that eventually files get transferred even in the case of connectivity or power grid disruptions. Also, Kwaabana shapes the traffic minimizing the negative impact of long-running uploads on interactive (browsing) traffic. Together with the data in files, Kwaabana synchronizes the metadata stored in the database (users, their location, files metadata) by shipping the database SQL statements in a text file. As a result, the system speeds up the sharing and saves the bandwidth when sharing happens locally (often the case according to the social graph structure) and provides a reliable way to transfer files for global sharing even in the case of unreliable network and the power grid. The

Kwaabana authors pointed out that a P2P architecture may be a promising future direction, but no details of the idea were presented.

Offline Downloading [74] is a popular approach employed in China to overcome low download bandwidth. According to [74], there are two broad types of offline downloading approaches in China: (1) when an Internet Service Provider (ISP) provides a service for offline downloading and (2) using software installed on smart WiFi access points providing a download manager. In both cases, the idea is that the user would submit a download request specifying the file to be downloaded and the download would be handled by the ISP server or by the access point. After the download is finished, the user can have fast access to the file without no further waiting. The recommended approach to be used in a particular situation depends on several factors [74], including the quality of the connection between the end-user device and the server handling the offline download.

Content Delivery Networks (CDN) [98] replicate the data to the edge servers all over the world (usually, several servers per continent) aiming to improve the download speed by redirecting the client to the server with the faster connection to that client, usually the one located physically the closest. Technically, when a DNS request is made to download a web resource (for instance, an image), the hostname of the resource is resolved to IP address of the server that will be the best in serving the resource, normally the one located in the proximity to the requesting device.

5.2.3 Approaches on Personal Device

The third group of solutions applies various techniques on the user device, including in-app caching, asynchronous content download and upload and minimizing the number of requests to improve the response time. We review below some of the proposals.

COCO [102] is a framework allowing to improve the user experience when working with web applications in underconnected settings. COCO uses a local cache based on Google Gears[6] to allow the web app to operate completely offline in case of no Internet connection or reduce the latency by reducing the number of requests when operating online. The solution is particularly suitable for data submission scenarios when the recorded data is accumulated locally and pushed to the server when the connection becomes available. When version conflicts occur, COCO relies on a built-in mechanism for their resolution.

ODK Submit [18] targets the data collection and submission problem. ODK Submit uses contextually available information about the data (like the importance of the data, availability and price of different types of network connection) to identify which channel should be used to transfer which type (or part) of the data. For instance, information about the number of doctors in the hospital is not heavy and is important while their photos are heavy and not that important. So, in this case, it makes sense to send the number of doctors even via SMS (or

[6] Google Gears https://en.wikipedia.org/wiki/Gears_(software) (last accessed 10 November 2016)

pricey 3G when available), while the photos can be delivered later for free via WiFi when it is available. They propose a framework called Submit that can run on Android as a service that abstracts the networking and can be used by other Android applications when they need to send the data. Theoretically, the service allows to use 3G, SMS and supports peer-to-peer data transfer with QR codes, two versions of NFC, Bluetooth and WiFi Direct to transmit the data. In practice according to the user studies, WiFi Direct seemed to be a preferable solution. WiFi Direct can take a bit longer to setup compared to Bluetooth, but offers better speeds and is particularly suitable for files more than 1MB in size. The authors have demonstrated that Submit can act according to configured policies, for instance, it was able to conserve 3G traffic by sending only important data, and the rest of the data was sent using WiFi. In [18] they say that as the file size increases past 1MB, WiFi Direct is the best option regarding the speed.

5.2.4 Discussion

Based on the presented review of the delivery solutions, in our proposal, we aim to offer benefits of the first (Mechanical Delivery) and the second (Intermediate Servers) proposal groups. As in the first group, we aim to make our proposal portable allowing mechanical content delivery and as in the second group we want to reduce latency and save bandwidth by integrating intermediate servers into our proposal. In the next section, we describe the novel delivery model we propose.

5.3 Mobile Knowledge Delivery: P2P and Low-Cost Computers

To offer a system suitable for field conditions, we have revisited the assumptions underlying the design of existing knowledge delivery systems. Two prominent trends are underway in the recent years. The first one is the ongoing proliferation of personal computing devices including tablets, laptops and, particularly, smartphones. The personal devices are already ubiquitous in developed countries reflected by the Bring your own device (BYOD) policy and are becoming pervasive even in developing countries. Mobile phone subscriptions penetration in developing regions is approaching 100% [116]. While it is not yet the case with the personal computing devices availability, the ongoing spreading of smartphones allows to expect that the smartphones will become ubiquitous in developing regions as well in the near future like mobile phones recently did. In practice, the majority of MSF mission employees are equipped with at least one personal computing device, a laptop provided by MSF, and also possess a smartphone or a tablet. The second change is the availability of low-cost small yet powerful general-purpose computers, particularly single board computers (SBCs). It is now possible to buy a general-purpose SBC under 30 USD, an equivalent of a USB flash drive cost a few years ago. Such low-cost computers enable to build a set of services running locally and accessible in the underconnected areas. Our proposal builds on these two highlighted trends.

To facilitate knowledge delivery into underconnected settings and answer RQ4, we propose a novel model (**contribution 4**) to deliver knowledge from an Online System on the Internet to the end-user Personal System using: (1) a peer-to-peer (P2P) synchronization middleware, (2) a Local System running on a low-cost computer deployed in the field and (3) a Personal System providing content access on personal mobile devices. The model is schematically presented in Figure 5.4. Hereafter we discuss each of the components and their interplay more closely.

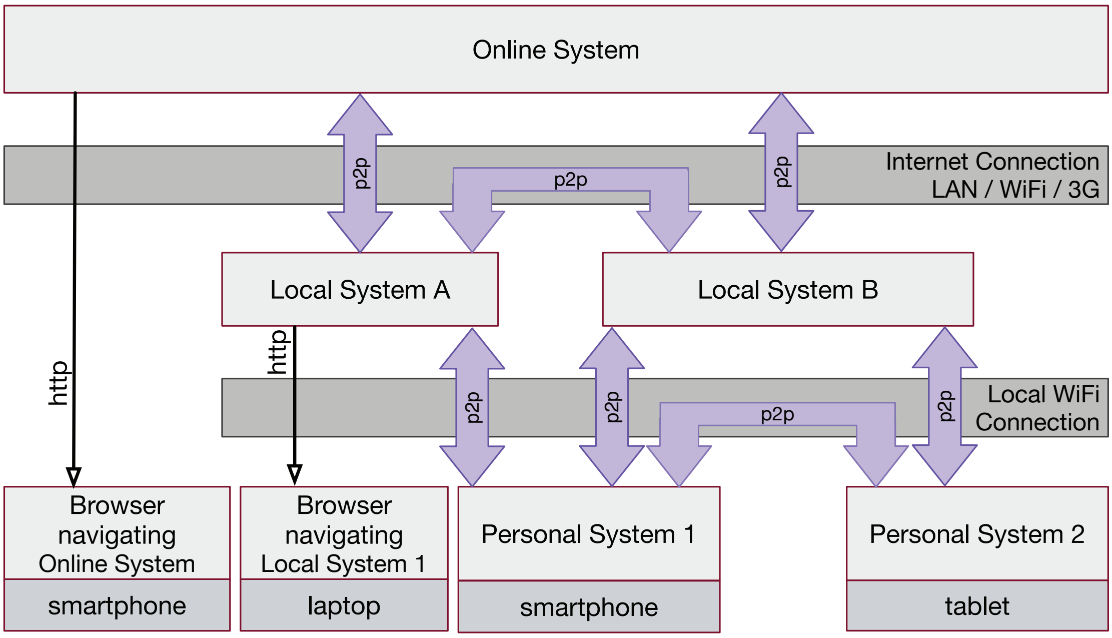


Figure 5.4 – The mobile delivery model with P2P and low-cost computers. The Local System is located inside of an underconnected environment (e.g., an MSF mission) and acts as a local replica containing relevant content.

Online System. When a high-quality Internet connection is available, the Online System is the preferred way of accessing organizational knowledge. The Online System can benefit from the computation power offered by the cloud and provide useful but at the same time computationally demanding services including search, analytics, and recommendations that are challenging to implement on low-cost computers. The HQ staff would typically use the Online System since good Internet connection is usually available in the HQ.

Local System. The Local System serves two primary purposes: (1) it acts as a local peer with the relevant content that can be synchronized from it and (2) it provides a web-access to access the content in the field by connecting to the Local System and using a web-browser even when the Personal System is not installed. To provide knowledge access in various field situations, possibly off-the-grid, the Local System should have a small form factor, low power consumption, and be able to run on a battery. Since often the field bandwidth is limited and expensive, the Local System should download the content from the Online System only once and then locally distribute it to Personal Systems. This is different from the mainstream cloud solutions where each client downloads from the cloud its own copy of the content. The Local System would be typically used by the field staff when accessing content in a permanent or temporary base, usually as part of a team.

Personal System. The Personal System is an application (mobile or desktop) running on a personal user device and has two functions: (1) handle P2P synchronization of the relevant content and (2) provide a UI for the personal device content access. Once the initial synchronization is done, the Personal System is the most portable option allowing the user to access relevant content at any time. The content can be updated by synchronizing with the Local System or the Online System depending on the available connection. The Personal System would usually be used by the field staff on the go, when away from a base. In case the user does not have the Personal System installed, she can still use the device to connect to the available Local System and navigate the content with the browser. This enables the content downloading but not synchronization.

P2P Middleware. Using a P2P middleware has proven to enable resilient and reliable content delivery on the global scale, for instance, CERN relies on P2P to deliver Large Hadron Collider experiment data to its partners worldwide. Because of its resilient properties, a P2P middleware is a suitable foundation for content delivery into underconnected environments. Thanks to the P2P middleware, the Online System, Local Systems and Personal Systems can all act as nodes in a P2P content delivery network. Because of the local peer discovery, the Local System and the Personal Systems can find each other and synchronize directly in the field (shown with horizontal arrows in Figure 5.4), without the need for transferring data via the central server, benefiting from a faster local network. The P2P middleware can adapt the delivery network structure to the connection speeds allowing to synchronize the data from the Online System to the Local System in the field only once and then distribute it to the Personal Systems hence minimizing the external bandwidth usage and saving often costly data.

5.4 Implementation: The GraaspBox

Following the model proposed in the previous section, we have developed a proof-of-concept solution for MSF using Graasp as the Online System, a low-cost computer running the Local System (called GraaspBox) and Bittorrent Sync[7] (BtSync) as P2P middleware.

Online System: Graasp. MSF relies on Graasp for their knowledge sharing as explained in Section 1.1. Each space in Graasp is mapped to a folder in the file system that can be synchronized to the Local System and the Personal System. To enable delivery of some non-traditional content types such as Space descriptions or online threaded discussions, we generate HTML files so the content can be viewed in browsers available virtually on any personal device. A space owner can enable synchronization of a space by pressing the “Enable peer-to-peer sync” button as shown in Figure 5.5 and disable using the “Disable peer-to-peer sync” button (see Figure 5.6). When synchronization is enabled, and the user has a connection to the Online System she can click the “Sync space content to my device” button (see Figure 5.6)

[7] http://www.bittorrent.com/sync (last accessed 10 November 2016)

to start syncing content to her personal device. In the case of absent Internet connection, the synchronization can be set up in a similar way by using the Local System.

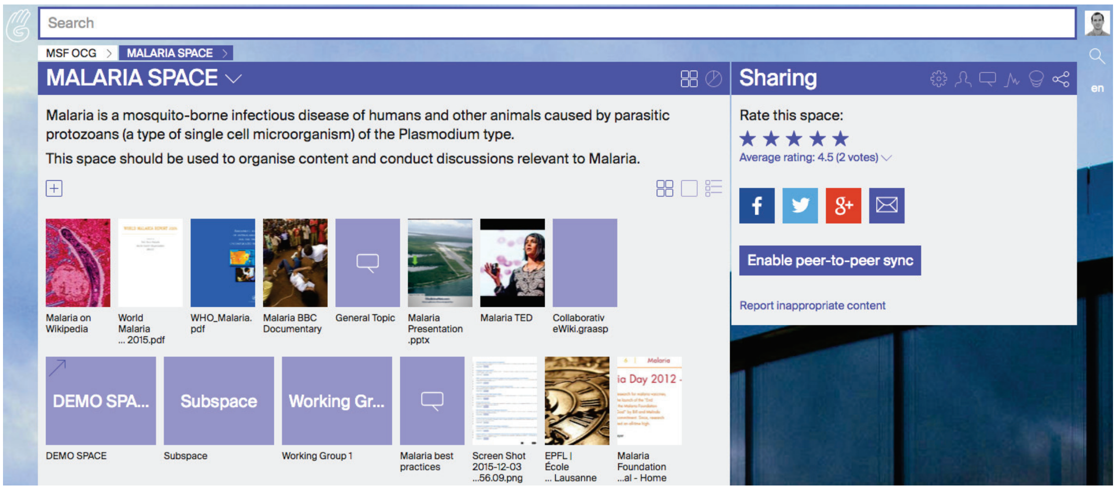


Figure 5.5 – A space owner can enable synchronization of the space by pressing the "Enable peer-to-peer sync" button on the sidebar.

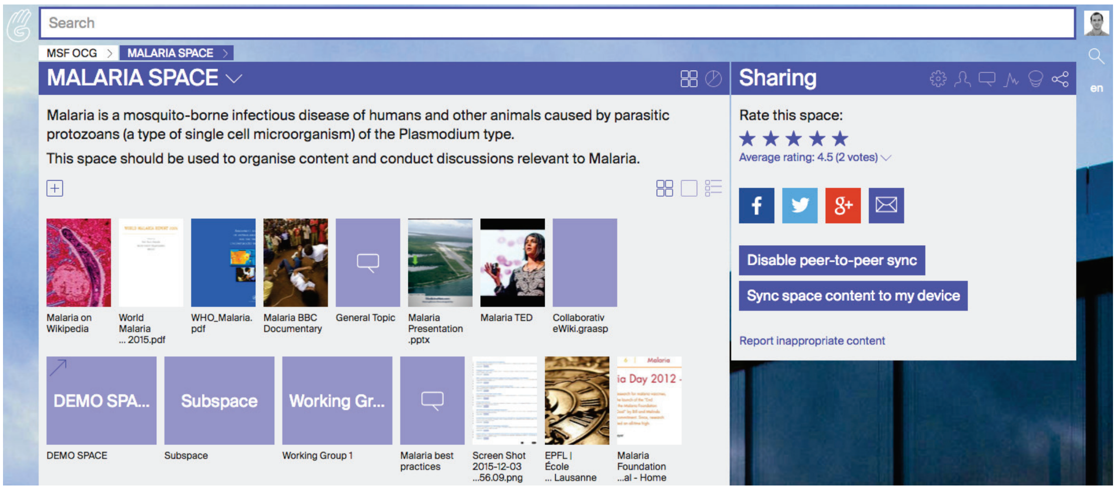


Figure 5.6 – A space owner can disable synchronization of the space by pressing the "Disable peer-to-peer sync" button on the sidebar. To start synchronizing the space to her personal device, she can press the "Sync space content to my device" button.

P2P Middleware: BtSync. We choose to use BtSync as a P2P middleware over open-source alternative including Hive2Hive[8] or Syncthing[9] since, according to our benchmarks, BtSync seemed to be more production-ready in terms of the APIs maturity and general stability compared to the open-source solutions. Also, the Bittorrent protocol that BtSync is built on has a proven resilience and performance. Still, our general delivery model is not coupled with a particular P2P middleware implementation. When syncing with BtSync, each space

[8]http://hive2hive.com/ (last accessed 10 November 2016)
[9]http://syncthing.net/ (last accessed 10 November 2016)

in Graasp has an associated secret Sync key accessible by the Local System. To manage the synchronization, the Online System, and the Local System interact with BtSync server via its APIs. The P2P BtSync middleware is aware of the network performance and makes sure that Personal Systems are downloading content from the GraaspBox, minimizing external Internet bandwidth usage.

Local System: The GraaspBox. GraaspBox is the Raspberry Pi single board computer running the Local System as shown in Figure 5.7. SBCs are low-cost, portable and usually energy efficient general purpose computers. When choosing a device for running the Local System, we have evaluated five different SBCs including Banana Pro, Orange Pi, Orange Pi Plus, Orange Pi Plus 2, and Raspberry Pi 3 and settled on Raspberry Pi 3 due to its adoption and community support. Reliability of the device hardware and software is crucial when deploying technology into the field since in case an issue occurs it is not always possible to fix it remotely. Higher adoption increases chances that errors are discovered and fixed early. There are already cases of employing Raspberry Pi in developing regions, for instance, when teaching practical Geography with sensors in a secondary school in Kenya [24]. Thanks to the low power consumption of Raspberry Pi, GraaspBox can run on a battery chargable in a car or using a solar panel. Due to its small pocketable size, the GraaspBox can be taken wherever the content is required (e.g., a hospital in a rural area or a car on its way to a remote mission) not relying on the network or power grid. To enable the local WiFi access of the content, we run a local DNS server with *dnsmasq* that resolves http://graasp.box to the local web server. This web server, built with Node.js, interacts with the P2P middleware managing the Sync keys and making sure that required content is synchronized. It as well allows the users to browse already synced content.

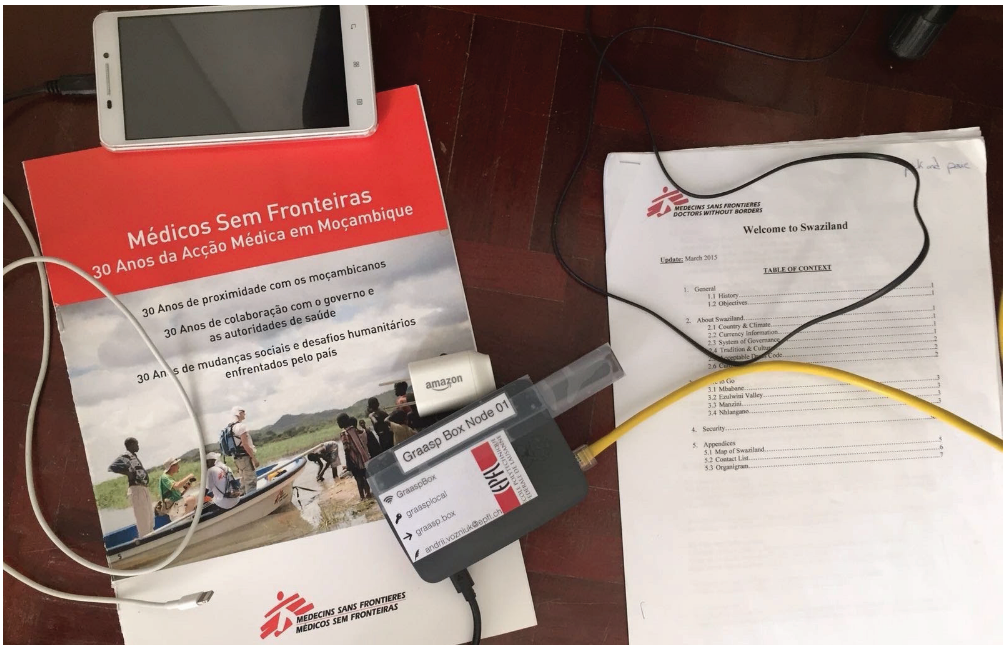


Figure 5.7 – A GraaspBox node in the Maputo mission, Mozambique. At this point, the node is connected to the Internet using the yellow LAN cable and is synchronizing the content.

Personal System. We use the free BtSync Android, iOS and Windows applications[10] to enable P2P synchronisation and provide content access on users' personal devices (see Figure 5.8 (2)). When the BtSync app is not available, a web-browser can be used to browse the content stored on the GraaspBox (see Figure 5.8 (1)).

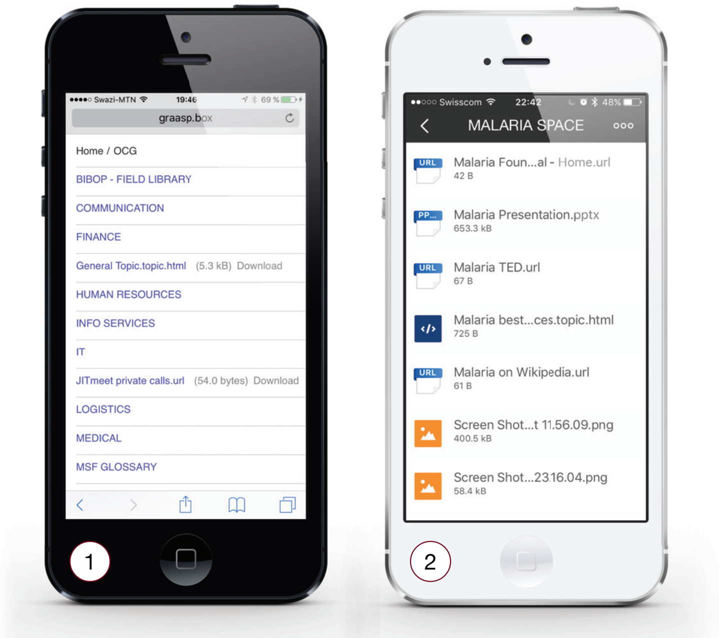


Figure 5.8 – (1) The MALARIA space on the GraaspBox accessed with a browser. (2) The MALARIA space on the phone managed with BtSync client.

In the next section, we explain how GraaspBox can be used in the field to improve content access.

5.5 GraaspBox Usage Scenario

When there is a fast and reliable Internet connection available, the user can access, view and download content directly on the Online System (Graasp). In addition to just content access, the user can benefit from other services provided by the Online System including advanced search, analytics, recommendations and a rich set of interaction options.

When the Internet connection is slow but still present, the user can use a hybrid scenario. In this scenario, she continues to navigate, search and interact using the online platform, but in case she needs to download a large file, she may download it quickly from the local GraaspBox by clicking on the "GraaspBox Download" button as shown in Figure 5.9. Pressing this button will redirect the user browser from the online content address starting with http://graasp.net/ to the GraaspBox content address starting with http://graasp.box/. For this redirect to work

[10] BtSync Apps https://getsync.com/download (last accessed 10 November 2016)

the user must be connected to the GraaspBox or the router in the user local network should have a corresponding rule in place.

If the Internet quality further degrades and the connection becomes unreliable or not available at all, the user can connect to the GraaspBox, navigate and access the content there offline. In case if the user would like to have the content available on her personal device, she can set up the synchronization with GraaspBox and afterward interact with the content locally on the device even when disconnected from GraaspBox.

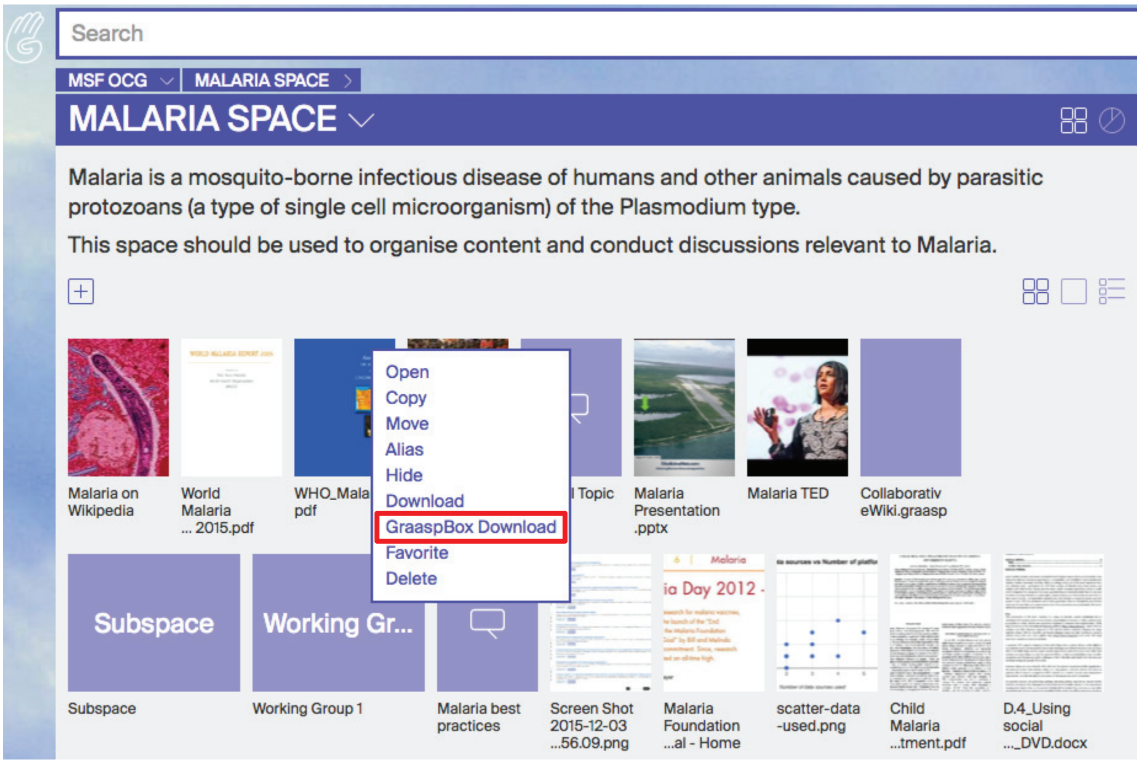


Figure 5.9 – A user can navigate the content using the Online System and at the same time download it from the Local System by clicking the “GraaspBox Download” button highlighted in the drop-down menu.

The next section focuses on the GraaspBox deployment in the field and user feedback.

5.6 Evaluation

To evaluate the GraaspBox, we conducted a series of laboratory functional and performance tests and deployed the devices in eight MSF missions during field trips discussed in this section.

Tests in the Laboratory. Before sending the device into the field, we have conducted a series of tests in the laboratory setting simulating the usage scenarios expected in the field (**validation 4.a**). In summary, the GraaspBox was able to synchronize the content and provide local access to it via WiFi without Internet access. It was able to act as a local peer for the peers connected to it via WiFi. When we enabled the Internet connection again, the GraaspBox downloaded updates from the online server. The maximum online synchronization speed (via LAN) observed in our tests, was around 90 Mbps. The download and upload speeds observed when using the WiFi stayed close to 20 Mbps. In our tests, Raspberry Pi was able to sustain maximum 32 simultaneous WiFi connections. The device was able to work continuously for nearly 45 hours (workload-dependent) on a single charge of a 20000 mAh battery. These numbers should be sufficient to support an MSF team in typical field scenarios.

Field Network Performance. To get an idea regarding the connection speed in the missions, network performance tests were conducted by MSF field staff representing two typical Graasp access scenarios. The first one measured the Graasp landing page loading time for not logged in users and the second measured loading time when the user was logged into her account on Graasp seeing her home page. To obtain the results reflecting the actual field condition the most accurately, the tests from the field were necessary. Realistic data is of particular importance since the Internet connection in the field is usually much less stable than in the HQ, even in the case where there are a couple of redundant connections available in the missions. The tests were done in August 2015 in Swaziland, Myanmar, Niger, Chad, Uganda, Kyrgyzstan and Mozambique. The test results are summarized in Table 5.1. Not surprisingly, in most of the cases, the measured loading times strongly correlated with the Internet connection bandwidth. It is worth noting that the obtained values varied during the day due to the Internet connection usage by the other users in the missions. Although the results of these tests are not completely reliable, they still provide useful insights. For instance, when a VSAT connection is used (as in Yangon, Myanmar, and N'Djamena, Chad), the response is slow due to the network latency, even when the bandwidth is sufficient. As for the other locations, when using an optical fiber or a DSL connection the Graasp website loads quickly. During the tests, we did not find a relation between the computer model and the loading times verifying that Graasp loading is not CPU bound. Interestingly, during the test the smallest loading time values were obtained on the oldest computer due to its good Internet connection. Regarding the tests generalization, it is worth keeping in mind that all the tests were conducted from the locations where the Internet connections are usually the best available to MSF staff in the country. Apart from the most developed countries, such as Kyrgyzstan, it is reasonable to expect that the most often Internet connection type available in the field is a VSAT connection.

Field Surveys. GraaspBox was deployed and evaluated with the help of an MSF knowledge manager during a total of eight field trips to Mozambique (Maputo and Gaza), Swaziland (Nhlangano, Mbabane, Hlathikhulu, Matsanjeni), Uganda (Kampala), and Chad (N'Djamena) (**validation 4.b**). Figure 5.10 shows schematically the way GraaspBox enables knowledge delivery for MSF. Two examples of the field study with the GraaspBox are presented in Figure 5.11. At the beginning of the trips, the manager demonstrated to the field staff several scenarios accessing the MSF content locally with the GraaspBox as explained in Section 5.5. He showed how to connect to the device, open content and establish synchronization with a personal

Location	Connection Type	Download speed	Upload speed	Landing page loading time (seconds)	Home page loading time (seconds)
Kyrgyzstan Bishkek	Optical Fiber	10M	Inside of country: 10 Mbps Outside of country: 1 Mbps	2.56	2.56
Myanmar Yangon	VSAT DSL	512 Kbps 1 Mbps	51 Kbps 512 Kbps	20.09	
Niger Niamey	Optical Fiber Wi-Max	1 Mbps 64 Kbps	1 Mbps 256 Kbps	17.13	12.45
Swaziland Nhlangano	DSL DSL	5 Mbps 1 Mbps	508 Kbps 252 Kbps	4.91	
Chad N'djamena	VSAT	2 Mbps	256 Kbps	34.50	
Uganda Kampala	Optical Fiber	1 Mbps	1 Mbps	3.24	8.89

Table 5.1 – Internet connection type and performance in MSF missions. The gray cells denote the data not obtained from the field.

device. Then he encouraged the mission staff to use the GraaspBox for their daily needs.

In total, the GraaspBox was used by 45 people, out of them individual interviews were conducted with 27 field employees, and of those $N = 14$ (11 - Male and 3 - Female; age ranged from 26 to 45 with median 31) have completed our questionnaire. Unfortunately, due to malfunction of the network performance monitoring software on the GraaspBox, we are not able to report the exact access patterns this time. The questionnaire (available in Appendix C) consisted of three parts: (1) evaluating the current situation with the Internet access; (2) collecting the needs regarding the knowledge delivery and (3) evaluating the field staff overall GraaspBox experience with the system usability scale (SUS) questionnaire [17].

Internet Access. The survey results have confirmed that slow and unreliable Internet connections were regular in the field. Six respondents (43%) said that they had a slow connection all the time, six (43%) witnessed it daily, and two (14%) - a few times each month. Unreliable connections were reported to be encountered all the time by six people (43%), daily - by five (36%), less often - by two (14%), and never - by one (7%). Moreover, five people (36%) reported to work all the time in situations without any Internet connection, four (29%) - work daily, three (21%) - a few times each month, and two (14%) - always have some way to access the Internet. Thirteen people (93%) expressed a strong agreement that they wanted to access the MSF content faster than they could. These numbers indicate the challenging access conditions in the field. After using the GraaspBox, 11 people (79%) expressed a strong agreement that MSF content access became faster than it was with the online platform.

Delivery Needs. According to the survey results, different ways of accessing the content are complementary, since users expressed their strong agreement with having relevant content available on their personal devices, on the GraaspBox and online (correspondingly, $\mu = 4.93$, $\mu = 4.57$ and $\mu = 4.50$ on the Likert scale). When asked regarding the single preferred access type, the majority of respondents picked their personal device (10 responses out of 14).

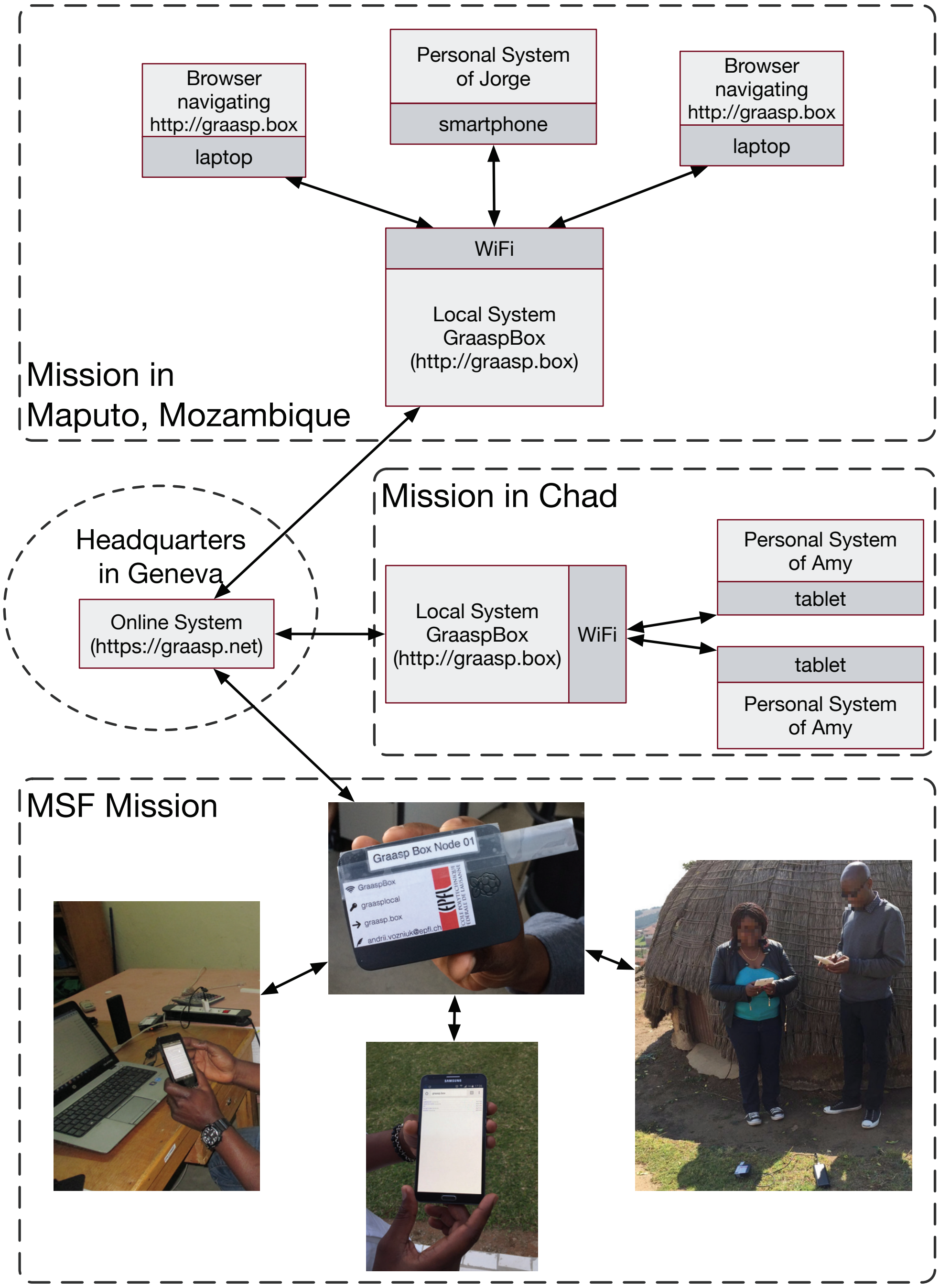


Figure 5.10 – A schematic representation of content delivery using GraaspBox from the HQ server located in Geneva to staff in multiple MSF missions. For the mission in the bottom, we present examples of devices used during field trips.

System Usability Scale. Based on the SUS questionnaire responses, the final GraaspBox SUS score was equal to 78, meaning that the system has a GOOD, close to EXCELLENT, usability score [6]. The Figure 5.12 presents an overview of answers to individual questions of the SUS.

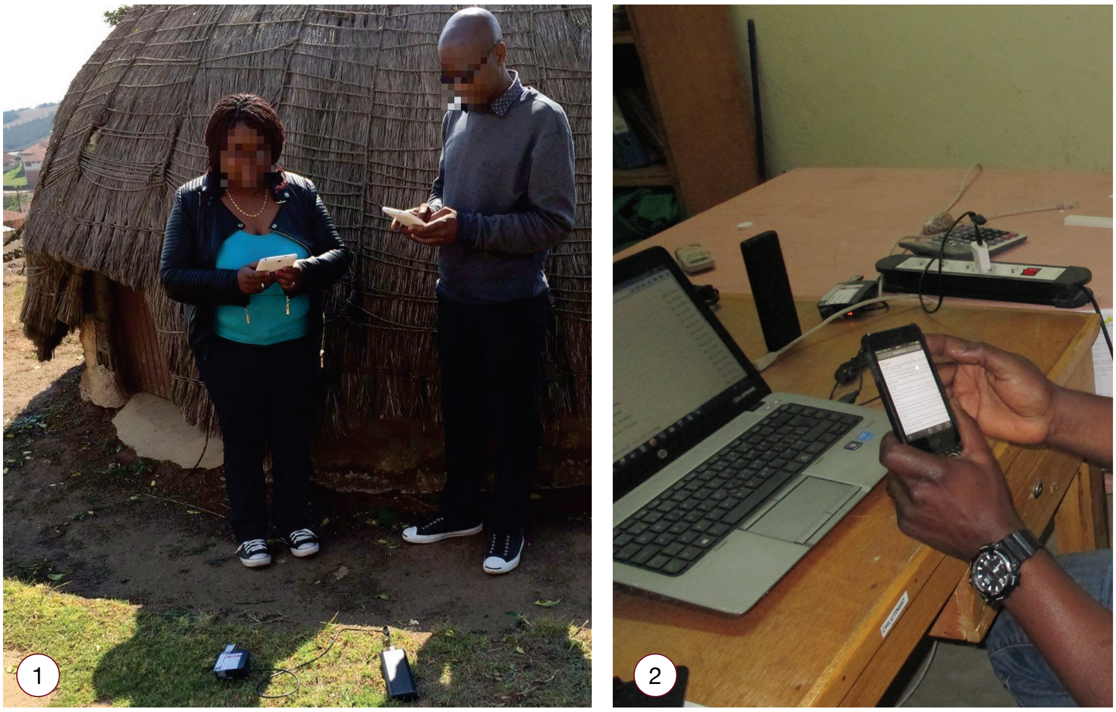


Figure 5.11 – (1) The MSF mission members are accessing content on a battery-powered GraaspBox in Maputo, Mozambique. (2) A mission member accessing the GraaspBox content using his phone and laptop in Chad.

The users reported that they were eager to use GraaspBox frequently ($\mu = 4.86, \sigma = 0.53$), they did not find it unnecessary complex ($\mu = 1.79, \sigma = 1.05$), they thought that it was easy to use ($\mu = 4.50, \sigma = 0.85$), they thought that they would not need the support of a technical person ($\mu = 2.00, \sigma = 1.04$) they found various functions to be well integrated ($\mu = 2.86, \sigma = 0.66$), they neither agree nor disagree regarding the system inconsistency ($\mu = 3.00, \sigma = 1.18$) they imagined most people would learn to use it quickly ($\mu = 4.79, \sigma = 0.58$). It was not cumbersome to use ($\mu = 1.57, \sigma = 1.02$), they felt rather confident using it ($\mu = 4.79, \sigma = 0.58$) without the need to learn a lot of things before getting going ($\mu = 2.29, \sigma = 1.59$). In addition to these positive results based on the survey responses, we have obtained positive feedback during face-to-face interviews with the field staff. The aspects of the system that received the lowest scores were the integration of different components (SUS Q5) and overall consistency (SUS Q6). This outcome is not surprising since the system is still in the early deployment stage providing space for iterative improvements. In the future, we are planning to conduct more detailed user interviews and think-aloud sessions to identify parts of the system that feel not consistent or not well integrated.

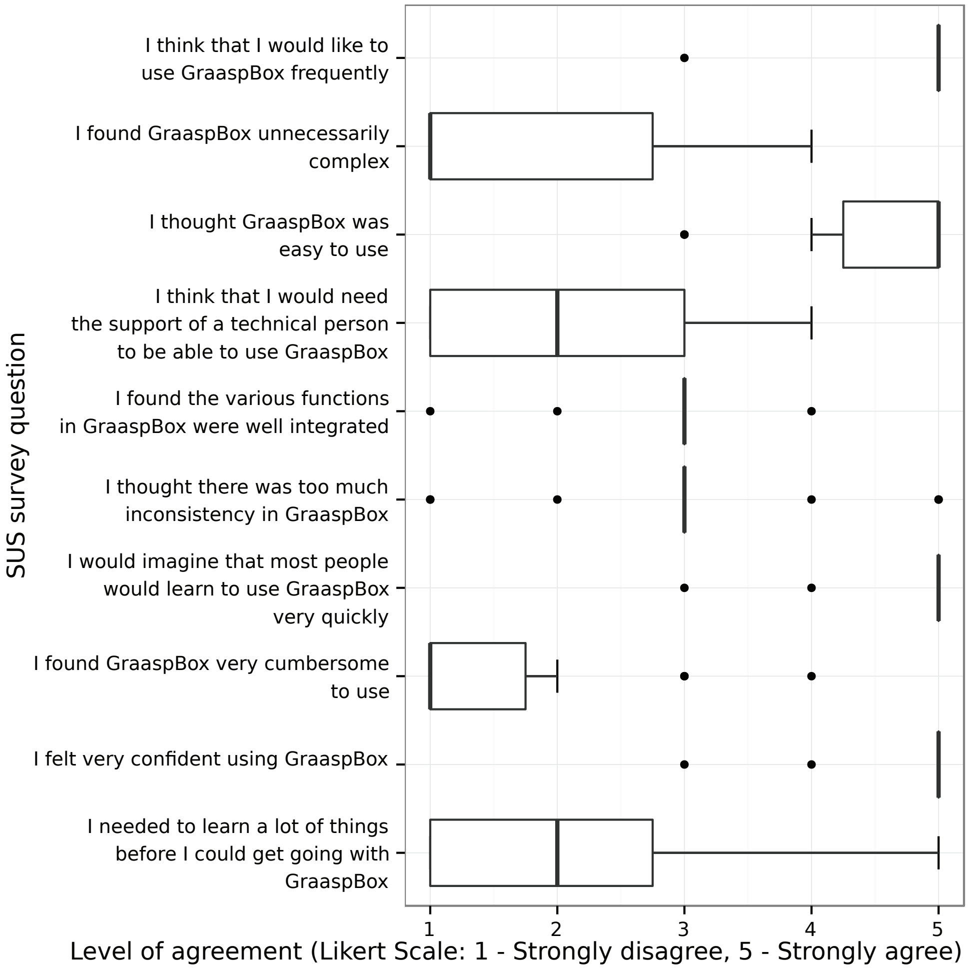


Figure 5.12 – GraaspBox system usability score survey results.

5.7 Implications for Educational Settings

According to the United Nations "The State of Broadband 2015" Report[11], in 2015 at least four billion people still did not have reliable Internet access, so they could not benefit from the knowledge shared online. Following our proposed approach, it is possible to make learning resources available on a GraaspBox in a school located in a remote area with limited Internet access. Such high-quality educational resources can improve learning outcomes as was demonstrated in [87], where the authors have run an experiment of supporting teaching activities with digital content in a low-resource classroom in peri-urban India. The paper demonstrated that having electronic teaching aid delivered to low-resource classrooms can be beneficial regarding the learning outcomes. This result motivates why providing digital content into such low-resource classrooms is important. Another confirmation of our delivery

[11] http://www.broadbandcommission.org/Documents/reports/bb-annualreport2015.pdf (last accessed 10 November 2016)

model relevance for educational applications is the fact that MSF considers using GraaspBox to deliver their organizational learning resources into the field.

In our scenario in Section 1.2 John, the physics teacher, teaches Physics not only in his school but also volunteers to visit developing countries. He often finds himself in a situation where there is only a limited Internet connectivity in the school. Hence often he carries the materials on a USB stick and distributes them manually in the classroom. John needs a better technology to support his teaching in such environments where the cloud access is not guaranteed. John can take the GraaspBox with him into a rural area together with smartphones for the students (if needed) and conduct teaching sessions without an Internet connection by providing offline access to ILSs or other open educational resources. In another case, when John teaches at his school with limited bandwidth, he can use the GraaspBox to reduce the content (e.g., video) loading time for his students by asking the students to load it locally from the GraaspBox, thus avoiding waiting during his class. Whenever he has a good Internet connection and updates content on the Graasp.net (Online System), the content gets automatically synchronized to GraaspBox (Local System) as soon as the box gets again connected to the Internet.

5.8 Conclusions and Future Work

In this chapter, we targeted the RQ4 studying how to facilitate knowledge delivery into underconnected environments. We proposed a novel knowledge delivery model that answers this question by enabling knowledge delivery to mobile devices using peer-to-peer middleware and low-cost computers. We also presented a novel proof-of-concept implementation of the model called GraaspBox for the knowledge sharing platform Graasp. Finally, we evaluated the implemented model in the laboratory through a series of tests and in the field by deploying the GraaspBox in eight MSF missions in Mozambique, Swaziland, Uganda, and Chad. The evaluations showed that the proposed model and the corresponding implementation allow to retain knowledge access in cases when there is no Internet connection and speed up the access when the connection is slow.

Our proposal allows a humanitarian organization to deliver and keep up to date knowledge in their underconnected missions. Using our model, other social media platforms can as well benefit providing offline access to content hosted there. It is worth highlighting an additional benefit of our model. Since the data is replicated from the Online System to multiple Local Systems in different locations, this provides additional data backup and improves the overall data storage reliability.

The focus of our proposal was on delivering organizational knowledge for humanitarian organizations, where field workers are exposed to underconnected environments on a daily basis. At the same time, the proposed model and its implementation with GraaspBox can be used to deliver training and educational resources (especially, cloud-based MOOCs) to educational institutions in developing countries located in underconnected areas [33]. Also, our model can be employed to deliver software updates and improve overall IT infrastructure security in the field [23].

Although GraaspBox fulfills the field delivery requirements formulated in Section 5.1.1, our field experience has uncovered some shortcomings of our approach discussed below.

Complexity of P2P. We observed that for regular users it was hard to understand principles of P2P networks. Delivering data locally and not from the cloud is something they are not used to. In most of the cases, GraaspBox just worked as expected. But when some issues occurred (e.g. the updated file was not synced yet), it was challenging to explain to the user's specific technical problems underlying the issue. In the future versions of the GraaspBox, we are planning to make it more transparent to the end-user how the content gets delivered by explaining it better in the documentation.

P2P App Installation. P2P synchronization to the user device requires an app (Personal System) to be installed. Since the installation often needs to take place in the field without an Internet connection, we store the apps on the GraaspBox so they can be downloaded locally. It is possible to install the apps in this way on Android and Windows or Linux, but not on iOS (without the jailbreak). iOS allows installing apps only from the App Store, for which the Internet connection is required. When the app can not be installed, GraaspBox provides a way to access the content using the web-browser available on virtually any platform. But in such case, the content is not synchronized to the personal device. Another promising direction is to enable P2P synchronization directly in the user browser without the need to install a standalone applications. Several browser-based P2P architectures exist built on top of the WebRTC standard[12], but still, not all of the features are supported even by the latest versions of the mainstream browsers. In practice, browsers used in the field can be outdated lacking the standard support at all.

Below, we discuss some of the promising directions for the future research.

Delivering Knowledge From Underconnected Environments. In this study, we targeted knowledge delivery to underconnected environments. We would also like to be able to deliver knowledge from the field to the HQ. Uploading data from poor network environments is a known problem, where it was measured that more than 75% of uploads fails in such conditions [55]. The GraaspBox users in the field have suggested using GraaspBox to manage the unreliable uploads so the users can just put the files on the personal device and have them eventually uploaded. Technically, our architecture also allows uploads, but uploads bring challenges related to handling conflicted versions of the same content that was modified by multiple users. As part of the future work, we plan to enable one-way content delivery from the field to the HQ. In such a case the Local System (GraaspBox) can serve as a staging node where data is stored until the Internet connection becomes available. Afterward, the accumulated content is synchronized to the server in the HQ.

Delivering Services. We are currently extending GraaspBox to provide not only local access

[12] WebRTC Project https://webrtc.org/ (last accessed 10 November 2016)

to the content in the field but also to run a set of services locally. Such services can be beneficial for local collaboration and communication. For instance, one of such services we are testing on GraaspBox is SpeakUp [51], a system allowing anonymous brainstorming that can elicit the best ideas in a group or best questions in q&a sessions. It is possible to run a standalone SpeakUp server on GraaspBox and provide local brainstorming services in the field by forwarding requests coming from the SpeakUp mobile app to the local server instead of the server located online.

6 Conclusions and Perspectives

In the recent years, social media platforms have become a global scale ubiquitous communication medium. At the same time, they are increasingly often deployed on a smaller scale to facilitate knowledge sharing inside of organizations in various domains. Based on our investigation context, in this thesis, we focused on two domains, namely the educational and the humanitarian, where social media platforms are used respectively to facilitate blended learning and enable distributed knowledge sharing. When mainstream social media platforms are deployed in these domains they face challenges undermining their adoption.

The primary goal of this thesis was to improve social media platforms to make them more suitable for educational and humanitarian knowledge sharing. To achieve this aim we have targeted four aspects of social media platforms, namely *analytics*, *privacy*, *discovery*, and *delivery*. When conducting the study, we followed an iterative and incremental process involving target users at every step starting from understanding their needs and formulating their requirements. This participatory design approach is well accepted in the HCI community and known to be effective. We think that an important and valuable general aspect of our work is that all proposed approaches have been deployed in authentic settings and adopted by their target users.

This chapter outlines the main contributions of this study, highlights its limitations and offers perspectives for further research investigations.

6.1 Key Contributions

In this section, we summarize the main contributions of this thesis regarding the research questions formulated in Chapter 1.

6.1.1 Analytics-related Contributions

Increasingly more types of interaction between people become computer-mediated. This trend has a potential to continue in the future with the current developments of virtual reality environments. Analytical tools and visualizations are a suitable solution to enable awareness and reflection in such computer-mediated environments. In the context of this thesis we specifically considered how support for awareness and reflection can be provided on social media platforms for facilitating knowledge sharing. We aimed at answering the following research question:

RQ1 - Analytics: how to provide user-oriented analytics in knowledge sharing systems to support awareness and reflection?

To address this research question, we proposed in Chapter 2 an embedded contextual analytics model making the analytics user-oriented, i.e. targeting the end-users of the platform (**contribution 1.a**). This is different from other approaches often focusing on offering analytics to researchers, analysts or institutions. Since in the proposed model analytics is a part of the interaction context, it is particularly suitable for enabling user awareness and reflection, be it in a blended learning session or when sharing knowledge in a workgroup. To materialize the proposed model in social media platforms, we proposed two general architectures enabling integration of embedded contextual analytics into the platforms. The first architecture is based on analytical web apps that obtain the data from the platform via a set of standardized APIs, analyze the data and present results to the user (**contribution 1.b**). The second architecture relies on an backend analytics engine (**contribution 1.c**). Each of these architectures possesses unique properties discussed in the thesis and making them suitable for different scenarios. The apps-based approach allows providing real-time domain-specific analytics for short interaction sessions, where the amount of data is limited and personalization is critical. At the same time, the approach based on the analytics engine is scalable and is suitable for interactive analysis of long-term recording involving many platform users.

We have validated the proposed architectures on the technical level by building analytical tools used in educational and humanitarian settings and integrating them into the Graasp social media platform. Following the apps-based approach and based on identified teacher needs we have built four analytical apps to support teachers and students in blended learning scenarios implemented in Graasp. For the second approach using an analytics engine, we have designed together with MSF knowledge managers a dashboard and deployed it to support awareness in the knowledge sharing process. We have evaluated both approaches with their target users. The evaluation indicated that the teachers, students, and knowledge workers have adopted embedded contextual analytics tools to support their awareness and reflection.

6.1.2 Privacy-related Contributions

Providing adequate activity analytics in a knowledge sharing platform requires recording and processing of digital traces that users leave behind when interacting with the platform. The storage and analysis of user interaction traces are often associated with privacy concerns. These concerns are triggered by possible negative implications of data exposure when tracked in sensitive contexts, as well as the legal framework often in place in such contexts as in the case of dealing with educational or medical data. Users require control over the traces being recorded depending on the context they work and over the storage location of the traces. Hence, we formulated the following research question:

RQ2 - Privacy: what privacy management interfaces and mechanisms are suitable for knowledge analytics and learning analytics?

To answer this question, we proposed in Chapter 3 a novel agent-based privacy management interface and a corresponding mechanism. Differently from settings-based privacy management interfaces employed in mainstream social media platforms, our approach uses a metaphor of physical presence in an interaction context. In such a metaphor, an agent tracks the actions it can observe similarly to how a person physically present in a room could observe and take notes of everything happening there. The presence of the agent makes the users in an interaction context aware of the tracking happening and allows to manage the tracking policy in a way similar to the physical world by inviting or removing the agent. The agent brings flexibility regarding the format used for representing the actions and regarding the service it transmits the data to.

We have implemented the proposed privacy interface and mechanism in Graasp and applied it in the context of two European projects, i.e., Go-Lab and SiWay. These projects used ActivityStreams and xAPI respectfully to represent user activities and relied on two different types of analytics infrastructure. The general model of the agent-based privacy management was able to meet the requirements in both cases. Also, we conducted an evaluation of the implemented model in Graasp and obtained a positive feedback from the users.

6.1.3 Knowledge Discovery-related Contributions

Finding relevant content is one of the core activities of users interacting with a content repository, be it knowledge workers using an organizational knowledge management system at a workplace or self-regulated learners collaborating in an online learning environment. Due to the number of content items stored in such repositories potentially reaching millions, and quickly increasing, it can be challenging for the user to find relevant content by browsing or relying on the available search engine. Hence, we formulated the following research question:

RQ3 - Discovery: how to enable discovery of knowledge relevant to user interests?

We addressed this problem in Chapter 4 by providing content and people recommendations based on user interests enabling discovery of relevant knowledge. We propose an approach to building a user interests profile automatically by combining content analytics and activity tracking. Our approach allows the user to see her profile that the recommender uses when suggesting relevant items. Also, the user can interact with the recommender and manually adapt her interests when required. We have implemented the proposed recommender system in Graasp and conducted two evaluations with users in both the educational and humanitarian domains.

The evaluations demonstrated the ability of the approach to identify relevant user interests and to recommend relevant content. Also, it can suggest people with similar interests, a challenge which is mission-critical in the humanitarian context.

6.1.4 Knowledge Delivery-related Contributions

At the moment of writing in 2016, near half of the world's population still does not have reliable Internet access, so they cannot benefit from the knowledge shared online. As one example, the places where humanitarian action is needed are often those where Internet connectivity is limited or not available at all. Hence, we formulated the following research question:

RQ4 - Delivery: How to improve access to the Internet-located knowledge from within underconnected environments?

This thesis proposes in Chapter 5 a novel knowledge delivery model that relies on a peer-to-peer middleware and uses low-cost computers for local knowledge replication. Social media platforms built following this model enable knowledge delivery into underconnected environments.

Following the proposed model, we have developed a novel system called GraaspBox allowing MSF to deliver knowledge into the underconnected field from a social media platform employed by the headquarters. With the help of MSF, the system was deployed and evaluated in the course of eight field trips to Mozambique, Swaziland, Uganda, and Chad. The evaluation demonstrated that the proposed model and its implementation in GraaspBox are adequate means to facilitate knowledge access by field staff operating in underconnected missions.

6.1.5 Other Contributions

In addition to the described contributions, we think that another outcome of this thesis work is worth mentioning. In the course of my PhD, the Graasp platform was redesigned and rebuilt from scratch providing an improved user experience, a better internal system design, and a set of novel components compared to the previous versions. My primary roles in this redesign and rebuilt of Graasp were the design of the system architecture and initial leadership of the system development. Developing a system of the scale of Graasp that is actively used in production is a team's effort. Hence, I contributed to almost every aspect of the system as well as other members of the team did. Nevertheless, it is worth mentioning specific components where my contribution was significant. I have designed and developed the user activity tracking infrastructure that provides the data necessary for analytics, recommendations, and gamification. Also, I closely supervised the development of two analytics components and developed the recommendation and the search components. Additionally, I designed and developed the knowledge delivery infrastructure consisting of the Graasp component integrating Graasp with a P2P synchronization middleware and the GraaspBox system running on a low-cost computer. As results of these developments, Graasp became a better foundation for integrating proposals from this thesis as well as other novel features developed in projects our group is involved in. Now, Graasp is a core part of three European projects, including two large-scale projects Go-Lab and Next-Lab, is used by multiple humanitarian organizations and is a reliable and extensible foundation for the future research projects. Currently, more than ten thousands people rely on Graasp for knowledge sharing and benefit from the proposals developed in this thesis. For me, this is an inspiring outcome of this work.

6.2 Limitations

In this section, we discuss some of the general limitations of this study that have an impact on the reliability and generalizability of the results.

Knowledge Sharing Platforms. When evaluating our proposals, we have mainly targeted web platforms. But there are mobile-only systems, now conversational systems built around a discussion chat as a knowledge sharing channel as well as virtual reality. While the models that we proposed are general and can be applied to other interaction media, our developments were done in the context of a web-based social media platform. Other platforms may require using other ways to implement and integrate our proposals as for instance in the case of mobile applications.

Impact of the User Interface. We involved target users into the design and integration of our proposals. But the way the proposals are integrated into the user interface can influence the adoption, usefulness, and impact the user behavior. For instance, in the case of recommendations, they are presented in separate tabs in the user interface, which may be not visible immediately, hence potentially leading to fewer people using them.

Evaluation Methodology. This study has generalizability limitations inherent to studies employing surveys and case study-based evaluation methods. Since the case studies and surveys were done in a specific context with a group of users with a particular background, this may limit the generalizability of the findings. In our case, we approached this issue by conducting multiple studies, often across educational and humanitarian domains. Another evaluation limitation is related to the use of self-reported scales employed in some surveys from this thesis involving the possibility of bias common to this method. A more preferred way is whenever possible to get objective information regarding the question being asked, instead of relying on subjective reporting. As such, we were reporting the actual usage data for our proposals whenever such data was available. Unfortunately, it was not always the case.

Impact on Knowledge Sharing. During our evaluations, we collected user feedback regarding our proposals and reported the usage data. While these metrics can be a good indicator of the proposal's adoption and value, it can still be worth measuring the impact the proposal has on the knowledge sharing. This requires long-term studies of user behavior on the platform, for instance, by using A/B testing where a randomized half of the users are exposed to the system with the integrated proposal and another part works without the proposal deployed. With a long-term study, the behavior of these groups could be compared. Conducting such evaluations can be a part of the future work.

6.3 Future Research Directions

In this section, we highlight future research directions when applying social media platforms for knowledge sharing in the educational and humanitarian domains.

Default Privacy Policy. When addressing the challenge of enabling user data privacy, an open research question remains regarding the default data tracking policy used by the platforms. In the course of this study, we as well witnessed the ethical aspects related to the granularity of the tracked interactions. Some users reported feeling uncomfortable since their interactions could be seen and analyzed by other members of the space. We faced several times the conflict between having a complete picture of user activities requested by knowledge managers and not exposing too detailed information that was of concern for regular users. Another point raised when deploying our proposals was related to the interaction data ownership. Do interaction traces belong to the user, to the context where these interactions have occurred, to the platform owners or the organization where the user is employed? What about collaborative activities, who owns the traces and can remove them in this case? Can the user take the traces away when leaving or removing her account or we should consider history irreversible and should traces remain forever? Similarly to the question of ownership it was not clear who can decide where the traces should be physically stored and who can define the tracking policy: a space owner, an organization or the user herself.

Conversational Interface to Analytics. While we used embedded contextual dashboards to deliver analytics in our proposal, it requires a high level of digital literacy to be able to understand and interpret the presented metrics. Conversational interfaces currently experience a renaissance with the increased popularity of chat applications and chatbots. One promising direction to investigate in our opinion is the difference in user experience between the analytics delivered via dashboards and analytics obtained through a conversation with a chatbot providing conversational interface to analytics.

Adoption and Participation. First of all, as with any technology developed to be used by people, the question of adoption is fundamental. It is too naive to assume that if the technology is deployed and considered as useful it is also naturally adopted. In fact, a considerable part of our efforts during our study was pursuing people to try our developments and provide the feedback necessary for improvements and evaluation. It remains an open question how to extend the participation of users in knowledge sharing.

A productive way to identify promising future research directions is to take a look at issues that arise when combining several of our proposals together in a single system. For instance, when the analytics and delivery are combined, two open questions come to mind: (1) how to record interaction traces happening offline and aggregate them into the online platform to get a consistent overview of the user activities (this question also highlights the importance of managing online-offline user identity) (2) how to deliver analytics to the users interacting offline with GraaspBox to enable awareness and reflection. The apps-based approach can enable such functionality if an offline activity record storage is implemented on GraaspBox and accessible via a set of standardized APIs.

Privacy & Discovery. If we consider privacy and discovery together, we can see a conflict between these two aspects. Allowing recommendations of content relevant to user interests requires the system to have a consistent overview of the user activities. At the same time the privacy management mechanism can be used to disable activity tracking in the specific context, hence not recording some of the activity traces that the recommender relies on. This partial absence of the activity information can impact the user interests profile and the recommendations and make them not reflect real user interests, particularly when most of the user interactions happen in such contexts without tracking.

Privacy & Analytics. Issues similar to the ones from the previous paragraph arise when thinking of privacy and analytics. Changing privacy policy has an impact on the analytics. For instance, if a user is an active contributor in contexts where tracking is disabled, this may be not reflected in her overall performance indicators not representing the situation objectively. This issue highlights the importance of the permissions management, namely who should be able to enable and disable tracking? Is it a teacher, a project manager, a user, or an institution? This question remains open.

Privacy & Delivery. Currently, the user activities on the GraaspBox are not captured. But if we would like to provide analytics offline, such tracking would need to be put in place. User tracking brings forward the question of imposing on GraaspBox the same privacy policies that are set up in the online system. To be able to put such policies in place, the local system (GraaspBox) should, together with content, synchronize and store the tracking policies and update them when they are changed in the online system. A potential issue with this approach is that in the situations when the device is mostly offline, such update in the tracking policy will be transmitted to the GraaspBox only when it connects again to the Internet. Such delayed policy change may result in not recording interactions for a considerable period. Another approach could be to record everything but send only data that complies with the recording policy when a connection is available and delete the rest. However, this may result in sending data for the periods when the tracking was disabled, but the device was not aware of that. At the same time, storing user activities on the local system may produce additional privacy risks as the device can get stolen. Hence, the local user traces should be adequately protected, for instance, by employing an encryption.

Discovery & Delivery. When combining discovery and delivery aspects, an interesting question is how to enable access to the recommended relevant content in the field with periods of absent Internet connection. When using an online platform, the user can simply click on the recommended items and view them. The same is not possible offline in case the content was not synchronized beforehand when the connection was available. One option could be to deliver (synchronize) recommended content proactively based on the user interests so that it is available when the user needs it. Such approach may have a considerable data usage overhead since potentially many content items will be delivered even if they are not used afterward.

The outlined questions are examples of the complex challenges emerging from an interplay of components when a system with a rich set of functionalities is deployed into production. These questions do not seem to have a straightforward answer and require a careful examination of the tradeoffs to find a suitable solution. The upcoming European project Next-Lab aims to address some of the mentioned challenges and will use the Graasp platform with integrated proposals from this thesis as foundation.

A Appendix

A.1 Appendix A: AngeLA Evaluations Survey

The survey form below was used to evaluate AngeLA, an agent-based privacy management interface described in Chapter 3.

AngeLA System Usability Scale

The use of learning analytics tools implies recording and processing of student activities conducted in Graasp. AngeLA is an approach allowing to provide the teacher with full control over student activity tracking in Graasp. AngeLA mimics in a virtual space the privacy control mechanism inherently present in a physical room: if a person is present in a room, she is able to observe all activities happening in the room.

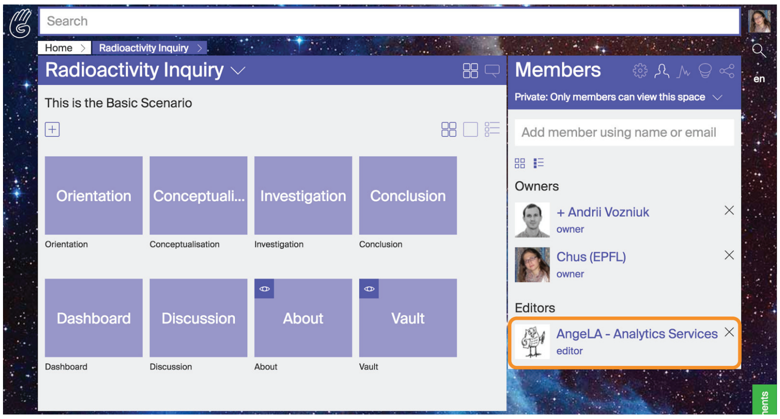
When AngeLA is present as a member of a space (like on the figure below), she records all the actions happening in the space so it is possible to use any of the Learning Analytics apps designed to support teachers and students. The apps allow to better understand the learning process and the usage of the inquiry learning space. Similarly, when AngeLA is removed (by clicking on the “x”), so she is no longer a member of space, she will stop recording the actions and the Learning Analytics apps will stop receiving new data. You may find further information about AngeLA in the following link: http://graasp.eu/users/5405e202da3a95cf9050e8f9

We have implemented AngeLA in Graasp, but this approach can be applied to various learning environments and social media platforms. Now we would like to get your feedback.

Could you please answer the questions below based on your experience with AngeLA in Graasp. It should take under 5 minutes.

* Required

AngeLA is presented as a user in the Radioactivity Inquiry space



1. **I think that I would like to use AngeLA privacy mechanism frequently:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

2. **I found AngeLA privacy mechanism unnecessarily complex: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

3. **I thought AngeLA privacy mechanism was easy to use: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

4. **I think that I would need the support of a technical person to be able to use AngeLA privacy mechanism: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

5. **I found the various functions in AngeLA privacy mechanism were well integrated: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

6. **I thought there was too much inconsistency in AngeLA privacy mechanism: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

7. **I would imagine that most people would learn to use AngeLA privacy mechanism very quickly: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

8. **I found AngeLA privacy mechanism very cumbersome to use: ***
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

9. **I felt very confident using AngeLA privacy mechanism:** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

10. **I needed to learn a lot of things before I could get going with AngeLA privacy mechanism:** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

11. **Please provide your suggestions for improving AngeLA privacy mechanism (optional)**

12. **Please specify your age** *

Mark only one oval.

- Under 20
- 20-24
- 25-29
- 30-34
- 35-39
- 40-44
- 45-49
- 50-54
- 55-59
- Over 60

13. **Please specify your gender** *

Mark only one oval.

- Female
- Male
- Other:

14. **Please provide your email in case we can contact you for more information (optional)**

A.2 Appendix B: Recommender Evaluation Survey

This appendix contains a survey form used to evaluate the interests-based recommender system presented in Chapter 4.

10/18/2016 Interests-based Recommender Evaluation (10)

Interests-based Recommender Evaluation (10)

The interests-based recommender implemented in Graasp aims to provide you with content corresponding to your interests (see the figures below). The recommender has two steps, it:

1) identifies your interests based on your interaction with the content on the platform (Figure 1 below);
2) suggests relevant content based on the identified interests (Figure 2 below).

We would like to learn your opinion regarding the recommender in Graasp and invite you to answer the questions in this survey.

If you are interested learning more about the recommender internals, please check our paper: https://goo.gl/fe3MBn

The survey should take under 10 minutes to complete. The answers are anonymous unless you explicitly provide the email in the end of the survey.

* Required

Figure 1. Interests identified by Graasp based on the user-platform interaction

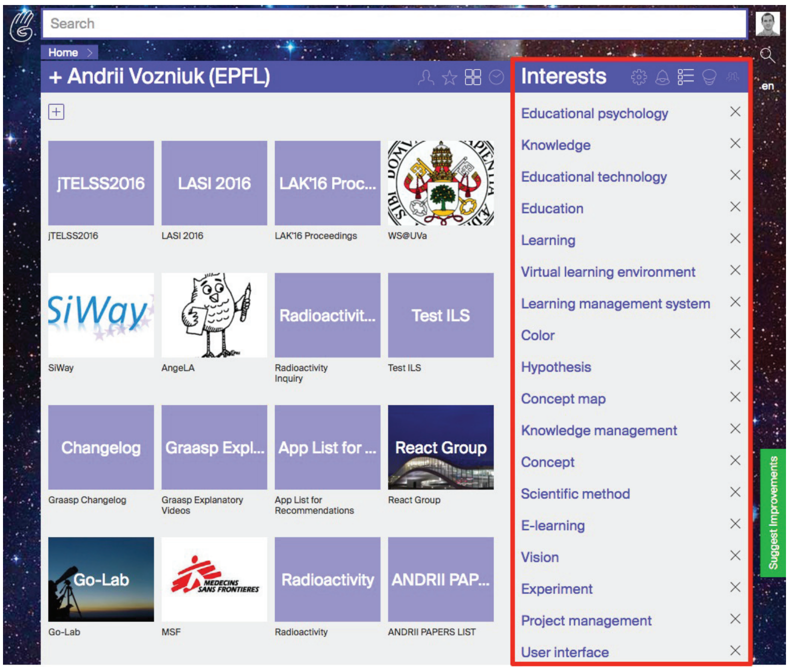
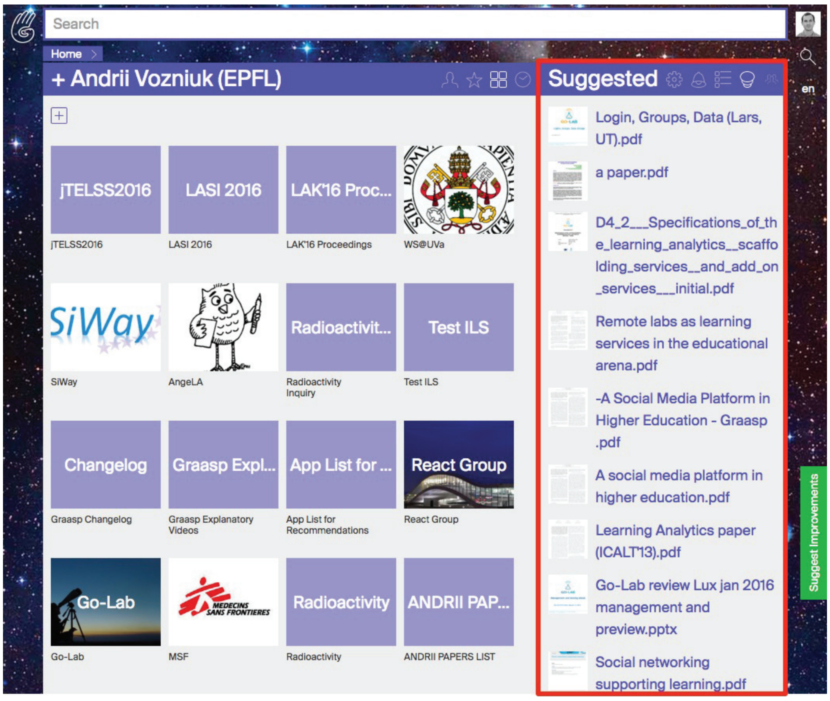


Figure 2. Content suggested by Graasp based on the identified interests

https://docs.google.com/a/vozniuk.com/forms/d/1kG7K7dIBB02hKl1Xt49sx09dlqNWEs_VJuKd9qH3DBs/edit 1/7



1. **Now please log into Graasp and take a look at the interests and suggestions made for you and then go to the next step.** *
   Check all that apply.

   ☐ I have logged into Graasp

Interests-based Recommender Potential

Please select if you agree with the statements below. Your level of agreement can range from "Strongly disagree" to "Strongly agree"

2. **I think I would like to see my interests as identified by the system:** *
   Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

3. **I think that I would like to be able to find other users based on their interests:** *
   Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

4. **I think that I would like to be able to edit my interests:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

5. **I think that I would like other users of the platform to see my interests:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

6. **I think that I would like to get content recommendations based on my interests:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

7. **I think that I would like to get similar users recommendations based on my interests:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

Interests-based Recommender Usability

Please select if you agree with the statements below. Your level of agreement can range from "Strongly disagree" to "Strongly agree"

8. **I think that I would like to use the Interests-based recommender frequently:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

9. **I found the Interests-based recommender unnecessarily complex:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

10. **I thought the Interests-based recommender was easy to use:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

11. **I think that I would need the support of a technical person to be able to use the Interests-based recommender: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

12. **I found the various functions in the Interests-based recommender were well integrated: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

13. **I thought there was too much inconsistency in the Interests-based recommender: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

14. **I would imagine that most people would learn to use the Interests-based recommender very quickly: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

15. **I found the Interests-based recommender very cumbersome to use: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

16. **I felt very confident using the Interests-based recommender: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

17. **I needed to learn a lot of things before I could get going with the Interests-based recommender: ***
    Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

Identified Interests Quality (Interests Tab)

Please check the interests as identified by Graasp based on your interaction with the platform. Please select if you agree with the statements below. Your level of agreement can range from "Strongly disagree" to "Strongly agree".

18. **I found identified interests relevant to my interaction with the platform** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strongly agree

19. **If you have at least 10 interests in Graasp, please specify how many of the top 10 interests are relevant to you interaction with the platform** *
Mark only one oval.

1	2	3	4	5	6	7	8	9	10
⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭

20. **Please specify what your liked about the identified interests** *

21. **Please specify what your did not like about the identified interests** *

22. **Please suggest improvements for the interests identification** *

23. **Please now go through the list of your interests on Graasp and remove the ones you consider irrelevant by pressing the X button on the right-hand side of the interest**

Recommended Content Quality (Suggested Tab)

Please check the content suggested to you in Graasp by the Interests-based recommender and select if you agree with the statements below. Your level of agreement can range from "Strongly disagree" to

"Strongly agree".

24. **I found it useful to have recommendations based on my interests:** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

25. **I found the recommended items relevant** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

26. **If you have at least 10 suggestions in Graasp, please specify how many of the top 10 suggested content are relevant to your interests** *
Mark only one oval.

1	2	3	4	5	6	7	8	9	10
◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

27. **I found the recommended items diverse** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

28. **Please specify what you liked about Interests-based recommender** *

29. **Please specify what you did not like about Interests-based recommender** *

30. **Please provide your suggestions for improving the Interests-based recommender** *

Additional information

Please provide some details about you so we know you better

31. **Please specify your age** *

32. **Please specify your gender** *
Mark only one oval.
    - Female
    - Male
    - Prefer not to answer

33. **Please provide your email in case we can contact you for more information (optional)**

Powered by Google Forms

A.3 Appendix C: GraaspBox Evaluation Survey

The survey form below was used to evaluate the GraaspBox knowledge delivery system presented in Chapter 5. Note, that when the survey was created, the system used to have a temporary name "Graspberry".

10/18/2016 Graspberry Box Evaluation

Graspberry Box Evaluation

We would like to hear from you about your needs and to see if the Graspberry Box can fulfil them. Please answer the questions below. Your answer is anonymous unless you specify your email at the end of the survey.

* Required

1. **How would you describe the usual speed of your Internet connection in the field?** *

Mark only one oval.

	1	2	3	4	5	
Slow	⬭	⬭	⬭	⬭	⬭	Fast

2. **How often do you have a slow Internet connection?** *

Mark only one oval.

- ⬭ All the time
- ⬭ Daily
- ⬭ Weekly
- ⬭ A few times each month
- ⬭ Never

3. **I would like to be able to access relevant MSF content faster that I currently can** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	⬭	⬭	⬭	⬭	⬭	Strong agree

4. **Is your usual Internet connection reliable (if you experience drops)?** *

Mark only one oval.

	1	2	3	4	5	
Unreliable	⬭	⬭	⬭	⬭	⬭	Reliable

5. **How often do you experience unreliable Internet connection?** *

Mark only one oval.

- ⬭ All the time
- ⬭ Daily
- ⬭ Weekly
- ⬭ A few times each month
- ⬭ Never

https://docs.google.com/a/vozniuk.com/forms/d/17VBQ8IT5g0J9tD2SugTE95a0wBOU_icopWnWkjsHBOU/edit 1/5

6. **How often do you work without any Internet connection available?** *
Mark only one oval.

- All the time
- Daily
- Weekly
- A few times each month
- Never

7. **How would you describe your current experience accessing MSF content?** *
Mark only one oval.

	1	2	3	4	5	
Slow						Fast

8. **I would like to have relevant MSF content available on my own device (mobile or laptop)** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree						Strongly agree

9. **I would like to have relevant MSF content available on the Graspberry Box** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree						Strongly agree

10. **I would like to have relevant MSF content available online** *
Mark only one oval.

	1	2	3	4	5	
Strongly disagree						Strongly agree

11. **If the same relevant MSF content is available online, on Graspberry Box, and on your device (phone or laptop), which would you use the most?** *
Mark only one oval.

- Online Platform
- Graspberry Box
- My device
- Other: ....

12. **Please explain your answer to the previous question** *

Graspberry Box Usability

You have tried accessing content locally from the Graspberry device. Please answer if you agree with the questions below based on your experience

13. **I think I would like to use Graspberry Box frequently** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

14. **I found Graspberry unnecessary complex** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

15. **I thought Graspberry was easy to use** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

16. **I think that I would need the support of a technical person to be able to use Graspberry** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

17. **I found various functions in Graspberrry were well integrated** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

18. **I though there was too much inconsistency in Graspberry** *

Mark only one oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

Bibliography

[1] The top 100 best NGOs 2013. *The Global Journal. Special Issue. 15* (2013).

[2] Acquisti, A., Alsenoy, B. V., Balsa, E., Berendt, B., Clarke, D., Diaz, C., Gao, B., Gürses, S., Kuczerawy, A., Pierson, J., Piessens, F., Sayaf, R., Schellens, T., Stutzman, F., Vanderhoven, E., and Wolf, R. D. SPION Deliverable D2.1 - State of the Art. Cosic internal report, 2011.

[3] Anand, A., and Walsh, I. Empirical Analysis of Factors Influencing Knowledge Sharing Facilities and Barriers in Indian SMEs: A Comparative Case Study. In *European Conference on Knowledge Management* (2015), Academic Conferences International Limited, pp. 887–895.

[4] Anderson, R., Robertson, C., Nabi, E., Sahni, U., and Setia, T. Facilitated video instruction in low resource schools. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development* (2012), ACM, pp. 2–12.

[5] Avison, D. E., and Shah, H. U. *The Information Systems Development Life Cycle: A First Course in Information Systems.* McGraw-Hill, 1997.

[6] Bangor, A., Kortum, P. T., and Miller, J. T. An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction 24*, 6 (2008), 574–594.

[7] Betty, P. Assessing homegrown library collections: Using Google Analytics to track use of screencasts and flash-based learning objects. *Journal of Electronic Resources Librarianship 21*, 1 (2009), 75–92.

[8] Blattner, M. M., Glinert, E. P., Jorge, J. A., and Ormsby, G. R. Metawidgets: towards a theory of multimodal interface design. In *Proceedings of Sixteenth Annual International Computer Software and Applications Conference (COMPSAC)* (1992), IEEE, pp. 115–120.

[9] Bogdanov, E., El Helou, S., Gillet, D., Salzmann, C., and Sire, S. Graaasp: a web 2.0 research platform for contextual recommendation with aggregated data. In *CHI'10*

Extended Abstracts on Human Factors in Computing Systems (2010), ACM, pp. 3523–3528.

[10] BOGDANOV, E., LIMPENS, F., LI, N., EL HELOU, S., SALZMANN, C., AND GILLET, D. A social media platform in higher education. In *EDUCON* (2012), pp. 1–8.

[11] BOGDANOV, E., SALZMANN, C., AND GILLET, D. Contextual Spaces with Functional Skins as OpenSocial Extension. In *4th International conference on advances in computer-human interactions* (2011), pp. 158–163.

[12] BOGDANOV, E., SALZMANN, C., AND GILLET, D. Widget-based approach for remote control labs. *9th IFAC Symposium on Advances in Control Education 45*, 11 (2012), 189–193.

[13] BOSNIĆ, I., VERBERT, K., AND DUVAL, E. Automatic keywords extraction - a basis for content recommendation. In *Proceedings of the Fourth International Workshop on Search and Exchange of e-le@rning Materials* (2010), pp. 51–60.

[14] BOYD, D. *Taken out of context: American teen sociality in networked publics.* PhD thesis, University of California-Berkeley, 2008.

[15] BOYD, D., AND MARWICK, A. E. Social Privacy in Networked Publics: Teens' Attitudes, Practices, and Strategies. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society* (2011).

[16] BRADFORD, P., PORCIELLO, M., BALKON, N., AND BACKUS, D. The Blackboard Learning System: The Be All and End All in Educational Instruction? *Journal of Educational Technology Systems 35*, 3 (2007), 301–314.

[17] BROOKE, J. SUS: A quick and dirty usability scale. *Usability evaluation in industry 189*, 194 (1996), 4–7.

[18] BRUNETTE, W., VIGIL, M., PERVAIZ, F., LEVARI, S., BORRIELLO, G., AND ANDERSON, R. Optimizing Mobile Application Communication for Challenged Network Environments. In *Proceedings of the 2015 Annual Symposium on Computing for Development* (2015), ACM, pp. 167–175.

[19] BUTOIANU, V., VIDAL, P., VERBERT, K., DUVAL, E., AND BROISIN, J. User Context and Personalized Learning: a Federation of Contextualized Attention Metadata. *Journal of Universal Computer Science 16*, 16 (2010), 2252–2271.

[20] CHARLEER, S., SANTOS, J. L., KLERKX, J., AND DUVAL, E. Improving teacher awareness through activity, badge and content visualizations. In *International Conference on Web-Based Learning* (2014), Springer, pp. 143–152.

[21] CHATTI, M. A., SCHROEDER, U., AND JARKE, M. LaaN: Convergence of knowledge management and Technology-Enhanced learning. *IEEE Transactions on Learning Technologies 5*, 2 (2012), 177–189.

[22] CHEN, J., AMERSHI, S., DHANANJAY, A., AND SUBRAMANIAN, L. Comparing web interaction models in developing regions. In *Proceedings of the First ACM Symposium on Computing for Development* (2010), ACM, p. 6.

[23] CORRIGAN-GIBBS, H., AND CHEN, J. FlashPatch: Spreading Software Updates over Flash Drives in Under-connected Regions. In *Proceedings of the Fifth ACM Symposium on Computing for Development* (2014), ACM, pp. 1–10.

[24] CUOMO, F., MIBUARI, E., WELDEMARIAM, K., AND STEWART, O. Leveraging Raspberry Pi for interactive education. In *Proceedings of the 4th Annual Symposium on Computing for Development* (2013), ACM, p. 16.

[25] DE JONG, T., LINN, M. C., AND ZACHARIA, Z. C. Physical and virtual laboratories in science and engineering education. *Science 340*, 6130 (2013), 305–308.

[26] DE JONG, T., RIESEN, S., KAMP, E., ANJEWIERDEN, A., BOLLEN, L., LAW, E., RUDINSKY, J., HEINTZ, M., MÄEOTS, M., PEDASTE, M., ET AL. D1.1: Go-Lab learning spaces specification. Tech. rep., Go-Lab European Project, 2013.

[27] DE JONG, T., SOTIROU, S., AND GILLET, D. Innovations in STEM education: The Go-Lab federation of online labs. *Smart Learning Environments 1*, 3 (2014).

[28] DOUGIAMAS, M., AND TAYLOR, P. Moodle: Using learning communities to create an open source course management system. In *World conference on educational multimedia, hypermedia and telecommunications* (2003), vol. 2003, pp. 171–178.

[29] DRACHSLER, H., VERBERT, K., SANTOS, O. C., AND MANOUSELIS, N. Panorama of recommender systems to support learning. In *Recommender Systems Handbook*. Springer, 2015, pp. 421–451.

[30] DUVAL, E. Attention please! learning analytics for visualization and recommendation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (2011), ACM, pp. 9–17.

[31] EL HELOU, S., SALZMANN, C., AND GILLET, D. The 3A Personalized, Contextual and Relation-based Recommender System. *Journal of Universal Computer Science 16*, 16 (2010), 2179–2195.

[32] ERDT, M., FERNANDEZ, A., AND RENSING, C. Evaluating Recommender Systems for Technology Enhanced Learning: A Quantitative Survey. *IEEE Transactions on Learning Technologies 8*, 4 (2015), 326–344.

[33] ESCHER, G., NOUKAKIS, D., AND AEBISCHER, P. Boosting Higher Education in Africa through Shared Massive Open Online Courses (MOOCs). In *Education, Learning, Training*. Brill, 2014, pp. 195–214.

[34] FERGUSON, R., BUCKINGHAM SHUM, S., AND DEAKIN CRICK, R. EnquiryBlogger: using widgets to support awareness and reflection in a PLE setting. In *Proceedings of the Workshop on Awareness and Reflection in PLEs, Personal Learning Environments Conference* (2011).

[35] FEW, S. Dashboard confusion revisited. *Visual Business Intelligence Newsletter* (2007).

[36] GILLET, D. Tackling engineering education research challenges: Web 2.0 social software for personal learning. *International Journal of Engineering Education 26*, 5 (2010), 1134–1143.

[37] GILLET, D., AND BOGDANOV, E. Cloud-savvy contextual spaces as agile personal learning environments or informal knowledge management solutions. In *Information Technology Based Higher Education and Training (ITHET), 2013 International Conference on* (2013), IEEE, pp. 1–6.

[38] GILLET, D., LAW, E. L., AND CHATTERJEE, A. Personal learning environments in a global higher engineering education Web 2.0 realm. In *Education Engineering (EDUCON)* (2010), IEEE, pp. 897–906.

[39] GILLET, D., AND NA, L. Designing PLE for Higher Education. In *Responsive Open Learning Environments*, S. Kroop, A. Mikroyannidis, and M. Wolpers, Eds. Springer, 2015, pp. 115–133.

[40] GILLET, D., NGOC, A. V. N., AND REKIK, Y. Collaborative web-based experimentation in flexible engineering education. *IEEE Transactions on Education 48*, 4 (2005), 696–704.

[41] GILLET, D., VOZNIUK, A., RODRIGUEZ TRIANA, M. J., AND HOLZER, A. C. Agile, Versatile, and Comprehensive Social Media Platform for Creating, Sharing, Exploiting, and Archiving Personal Learning Spaces, Artifacts, and Traces. In *The World Engineering Education Forum* (2016).

[42] GOVAERTS, S., CAO, Y., VOZNIUK, A., HOLZER, A. C., GARBI ZUTIN, D., SAN CRISTÓBAL RUIZ, E., BOLLEN, L., MANSKE, S., FALTIN, N., SALZMANN, C., TSOURLIDAKI, E., AND GILLET, D. Towards an Online Lab Portal for Inquiry-based STEM Learning at School. In *The 12th International Conference on Web-based Learning (ICWL)* (2013), Lecture Notes on Computer Science, Springer, pp. 244–253.

[43] GOVAERTS, S., VERBERT, K., DAHRENDORF, D., ULLRICH, C., SCHMIDT, M., WERKLE, M., CHATTERJEE, A., NUSSBAUMER, A., RENZEL, D., SCHEFFEL, M., FRIEDRICH, M., SANTOS, J. L., DUVAL, E., AND LAW, E. Towards responsive open learning environments: The role interoperability framework. In *Towards Ubiquitous Learning: 6th European Conference of Technology Enhanced Learning, EC-TEL* (2011), C. D. Kloos, D. Gillet, R. M. Crespo García, F. Wild, and M. Wolpers, Eds., Springer Berlin Heidelberg, pp. 125–138.

[44] Gross, R., and Acquisti, A. Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (2005), ACM, pp. 71–80.

[45] Gutwin, C., and Greenberg, S. The importance of awareness for team cognition in distributed collaboration. In *Team cognition: Understanding the factors that drive process and performance*. APA Press, 2004, pp. 177–203.

[46] Guy, I., Avraham, U., Carmel, D., Ur, S., Jacovi, M., and Ronen, I. Mining expertise and interests from social media. In *Proceedings of the 22nd International Conference on World Wide Web* (New York, 2013), ACM, pp. 515–526.

[47] Harkous, H., Rahman, R., Karlas, B., and Aberer, K. The Curious Case of the PDF Converter that Likes Mozart: Dissecting and Mitigating the Privacy Risk of Personal Cloud Apps. In *16th Privacy Enhancing Technologies Symposium (PETS)* (Darmstadt, 2016).

[48] Harrer, A., Martínez-Monés, A., and Dimitracopoulou, A. Users' data: collaborative and social analysis. In *Technology-Enhanced Learning. Principles and Products*, N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, S. Barnes, and L. Montandon, Eds. Springer, 2009, ch. 9, pp. 175–193.

[49] Heppen, J. B., and Therriault, S. B. Developing Early Warning Systems to Identify Potential High School Dropouts. Tech. rep., American Institutes for Research, National High School Center, 2008.

[50] Holt, J. *UML for Systems Engineering: watching the wheels*. The Institution of Engineering and Technology, 2004.

[51] Holzer, A. C., Govaerts, S., Ondrus, J., Vozniuk, A., Rigaud, D., and Garbinato, B. SpeakUp – A Mobile App Facilitating Audience Interaction. In *The 12th International Conference on Web-based Learning* (2013).

[52] Hull, G., Lipford, H., and Latulipe, C. Contextual gaps: Privacy issues on Facebook. *Ethics and information technology 13*, 4 (2011), 289–302.

[53] Iannella, R., and Finden, A. Privacy awareness: Icons and expression for social networks. In *8th International Workshop for Technical, Economic and Legal Aspects of Business Models for Virtual Goods Incorporating the 6th International ODRL Workshop* (2010), pp. 1–15.

[54] Jivet, I. Master thesis. The Learning Tracker. A Learner Dashboard that Encourages Self-regulation in MOOC Learners. Delft University of Technology. 2016.

[55] Johnson, D. L., Belding, E. M., and Mudenda, C. Kwaabana: File sharing for rural networks. In *Proceedings of the 4th Annual Symposium on Computing for Development* (2013), ACM, p. 4.

[56] KENSING, F., AND BLOMBERG, J. Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW) 7*, 3-4 (1998), 167–185.

[57] KIM, H.-W., AND KANKANHALLI, A. Investigating user resistance to information systems implementation: a status quo bias perspective. *Management Information Systems Quarterly 33*, 3 (2009), 567–582.

[58] KIM, J., LI, S.-W., CAI, C. J., GAJOS, K. Z., AND MILLER, R. C. Leveraging video interaction data and content analysis to improve video learning. In *Proceedings of the CHI 2014 Learning Innovation at Scale workshop* (2014).

[59] KING, W. R. *Knowledge Management and Organizational Learning*. Springer US, 2009.

[60] KITTO, K., CROSS, S., WATERS, Z., AND LUPTON, M. Learning analytics beyond the LMS: the connected learning analytics toolkit. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK)* (2015), ACM, pp. 11–15.

[61] KLAMMA, R., CHATTI, M. A., DUVAL, E., HUMMEL, H., HVANNBERG, E. T., KRAVCIK, M., LAW, E., NAEVE, A., AND SCOTT, P. Social Software for Life-long Learning. *Journal of Educational Technology & Society 10*, 3 (2007), 72–83.

[62] KOUTROPOULOS, A., GALLAGHER, M. S., ABAJIAN, S. C., DE WAARD, I., HOGUE, R. J., KESKIN, N. Ö., AND RODRIGUEZ, C. O. Emotive Vocabulary in MOOCs: Context & Participant Retention. *The European Journal of Open, Distance and E-Learning 15*, 1 (2012).

[63] KOVANOVIĆ, V., JOKSIMOVIĆ, S., GAŠEVIĆ, D., HATALA, M., AND SIEMENS, G. Content analytics: the definition, scope, and an overview of published research. *Handbook of Learning Analytics* (2015).

[64] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.

[65] KRÜGER, N., DANG-XUAN, L., SCHNEIDER, A.-M., AND STIEGLITZ, S. Usage of Social Media for External Stakeholder Relationship Management – A Study of German Companies and International Non-government Organizations. In *Advanced Information Networking and Applications Workshops (WAINA)* (2013), IEEE, pp. 1479–1482.

[66] LAI, Y.-L., AND HUI, K.-L. Internet Opt-in and Opt-out: Investigating the Roles of Frames, Defaults and Privacy Concerns. In *Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research* (New York, USA, 2006), ACM, pp. 253–263.

[67] LARMAN, C., AND BASILI, V. R. Iterative and incremental development: A brief history. *IEEE Computer*, 36 (2003), 47–56.

[68] LAUDON, D., AND LAUDON, J. *Management Information Systems: Organization and Technology in the Networked Enterprise*, 4th edition ed. Prentice Hall, 2001.

[69] LETTIERI, E., BORGA, F., AND SAVOLDELLI, A. Knowledge management in non-profit organizations. *Journal of Knowledge Management 8*, 6 (2004), 16–30.

[70] LI, L. D., AND CHEN, J. TroTro: web browsing and user interfaces in rural Ghana. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development* (2013), ACM, pp. 185–194.

[71] LI, N., HOLZER, A. C., GOVAERTS, S., AND GILLET, D. Enforcing Privacy for Teenagers in Online Inquiry Learning Spaces. In *Understanding Teen UX workshop at CHI* (2014).

[72] LI, N., INGRAM-EL HELOU, S., AND GILLET, D. Using Social Media for Collaborative Learning in Higher Education: A Case Study. In *Proceedings of the 5th International Conference on Advances in Computer-Human Interactions* (2012).

[73] LI, N., NAJAFIAN-RAZAVI, M., AND GILLET, D. Trust-aware Privacy Control for Social Media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (2011).

[74] LI, Z., WILSON, C., XU, T., LIU, Y., LU, Z., AND WANG, Y. Offline Downloading in China: A Comparative Study. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference* (2015), ACM, pp. 473–486.

[75] LIKERT, R. A technique for the measurement of attitudes. *Archives of psychology 22*, 140 (1932).

[76] MALDONADO, R. M., KAY, J., YACEF, K., AND SCHWENDIMANN, B. An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *International Conference on Intelligent Tutoring Systems* (2012), Springer, pp. 482–492.

[77] MAN, H., CHEN, H., AND JIN, Q. Open learning: A framework for sharable learning activities. *International Conference on Web-Based Learning* (2010), 387–392.

[78] MARSHALL, B., ZHANG, Y., CHEN, H., LALLY, A., SHEN, R., FOX, E., AND CASSEL, L. N. Convergence of knowledge management and e-learning: the GetSmart experience. In *Digital Libraries, 2003. Proceedings. 2003 Joint Conference on* (May 2003), pp. 135–146.

[79] MATSCHKE, C., MOSKALIUK, J., AND CRESS, U. Knowledge exchange using web 2.0 technologies in NGOs. *Journal of Knowledge Management 16*, 1 (2012), 159–176.

[80] MATUK, C., AND TISSENBAUM, M. Introduction. In *Dillenbourg, P. (discussant), M. Tissenbaum & C. Matuk (co-organizers). Real-Time Visualization of Student Activities to Support Classroom Orchestration. Symposium conducted at the International Conference for the Learning Sciences, Singapore* (2016), vol. 2, pp. 1120–1127.

[81] MIKOWSKI, M., AND POWELL, J. *Single Page Web Applications: JavaScript end-to-end.* Manning Publications, 2013.

[82] Minner, D. D., Levy, A. J., and Century, J. Inquiry-based science instruction - what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching 47*, 4 (2009), 474–496.

[83] Mooney, J. Bringing portability to the software process. Technical Report TR 97-1. Department of Statistics and Computer Science. West Virginia University, 1997.

[84] Moore, M. G. Editorial: Three types of interaction. *The American Journal of Distance Education 3*, 2 (1989), 1–6.

[85] Muller, M. J., and Kuhn, S. Participatory design. *Communications of ACM 36*, 6 (1993), 24–28.

[86] Najjar, J., Duval, E., and Wolpers, M. Attention metadata: Collection and management. In *WWW2006 workshop on logging traces of web activity: the mechanics of data collection* (2006), pp. 1–4.

[87] Nixon, L., Roy, J., Latha, H., Rao, S., Rangachar, B., and Thies, W. Measuring the learning benefits of electronic teaching aids in low-resource classrooms. In *Proceedings of the Fifth ACM Symposium on Computing for Development* (2014), ACM, pp. 127–128.

[88] Pohs, W., Pinder, G., Dougherty, C., and White, M. The Lotus Knowledge Discovery System: Tools and experiences. *IBM Systems Journal 40*, 4 (2001), 956–966.

[89] Ramos, J. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (2003).

[90] Raymond, N., Achkar, Z. A., Verhulst, S., and Berens, J. OCHA Think Brief: Building Data Responsibility into Humanitarian Action. OCHA, May 2016.

[91] Raynes-Goldie, K. Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday 15*, 1 (2010), 1–4.

[92] Rekik, Y., Gillet, D., El Helou, S., and Salzmann, C. The eLogBook Framework: Sustaining Interaction, Collaboration, and Learning in Laboratory-Oriented CoPs. *International Journal of Web-based Learning and Teaching Technologies 2*, 3 (2007), 61–76.

[93] Rodriguez Triana, M. J., Govaerts, S., Halimi, W., Holzer, A. C., Salzmann, C., Vozniuk, A., de Jong, T., Sotirou, S., and Gillet, D. Rich Open Educational Resources for Personal and Inquiry Learning. Agile Creation, Sharing and Reuse in Educational Social Media Platforms. In *International Conference on Web & Open Access to Learning* (2014).

[94] Rodriguez Triana, M. J., Prieto Santos, L. P., Vozniuk, A., Shirvani Boroujeni, M., Schwendimann, B. A., Holzer, A. C., and Gillet, D. Monitoring, Awareness and Reflection in Blended Technology Enhanced Learning: a Systematic Review. *In Press. International Journal of Technology Enhanced Learning* (2016).

[95] ROMERO, C., AND VENTURA, S. Educational Data Mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 40*, 6 (2010), 601–618.

[96] ROSENBERG, M., AND FOREMAN, S. White Paper. Capturing, Sharing, and Using Collective Knowledge to Drive Learning and Productivity. *The eLearning Guild* (2016).

[97] SANTOS, J. L., VERBERT, K., KLERKX, J., DUVAL, E., CHARLEER, S., AND TERNIER, S. Tracking Data in Open Learning Environments. *Journal of Universal Computer Science 21*, 7 (2015), 976–996.

[98] SAROIU, S., GUMMADI, K. P., DUNN, R. J., GRIBBLE, S. D., AND LEVY, H. M. An Analysis of Internet Content Delivery Systems. In *Proceedings of the Fifth Symposium on Operating Systems Design and Implementation (OSDI)* (Boston, USA, 2002), ACM, pp. 315–328.

[99] SCHWARTZ, B. *The Paradox of Choice: Why More is Lesss.* Harper Perennial, 2005.

[100] SCHWENDIMANN, B. A., RODRIGUEZ TRIANA, M. J., VOZNIUK, A., PRIETO, L. P., SHIRVANI BOROUJENI, M., HOLZER, A. C., GILLET, D., AND DILLENBOURG, P. Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies PP*, 99 (2016), 1–1.

[101] SCLATER, N. Code of practice for learning analytics. a literature review of the ethical and legal issues. Tech. rep., Jisc, 2014.

[102] SHAH, S., AND JOSHI, A. COCO: a web-based data tracking architecture for challenged network environments. In *Proceedings of the First ACM Symposium on Computing for Development* (2010), ACM.

[103] SHANI, G., AND GUNAWARDANA, A. Evaluating Recommendation Systems. In *Recommender Systems Handbook* (2011), Springer, pp. 257–297.

[104] SIEMENS, G. Learning analytics: envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (2012), ACM, pp. 4–8.

[105] SIEMENS, G., GASEVIC, D., HAYTHORNTHWAITE, C., DAWSON, S., SHUM, S., FERGUSON, R., DUVAL, E., VERBERT, K., AND BAKER, R. Open Learning Analytics: an integrated & modularized platform. Tech. rep., 2011.

[106] THÜS, H., CHATTI, M. A., BRANDT, R., AND SCHROEDER, U. Evolution of interests in the learning context data model. In *Design for Teaching and Learning in a Networked World: 10th European Conference on Technology Enhanced Learning* (2015), Springer, pp. 479–484.

[107] TSAI, J., EGELMAN, S., CRANOR, L., AND ACQUISTI, A. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research 22*, 2 (2011), 254–268.

[108] VERBERT, K., GOVAERTS, S., DUVAL, E., SANTOS, J. L., VAN ASSCHE, F., PARRA, G., AND KLERKX, J. Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing 18*, 6 (2014), 1499–1514.

[109] VOZNIUK, A., GOVAERTS, S., BOLLEN, L., MANSKE, S., HECKING, T., AND GILLET, D. AngeLA: Putting the Teacher in Control of Student Privacy in the Online Classroom. In *Information Technology Based Higher Education and Training (ITHET)* (2014).

[110] VOZNIUK, A., GOVAERTS, S., AND GILLET, D. Towards portable learning analytics dashboards. In *13th International Conference on Advanced Learning Technologies (ICALT)* (2013).

[111] VOZNIUK, A., HOLZER, A. C., GOVAERTS, S., MAZUZE, J., AND GILLET, D. Graspeo: a Social Media Platform for Knowledge Management in NGOs. In *Seventh International Conference on Information and Communication Technologies and Development (ICTD)* (2015), ACM.

[112] VOZNIUK, A., RODRIGUEZ TRIANA, M. J., HOLZER, A. C., AND GILLET, D. Combining Content Analytics and Activity Tracking to Identify User Interests and Enable Knowledge Discovery. In *Proceedings of the 6th International Workshop on Personalization Approaches in Learning Environments (PALE) at the Conference on User Modeling, Adaptation and Personalization (UMAP)* (2016), ACM, pp. 42–48.

[113] VOZNIUK, A., RODRIGUEZ TRIANA, M. J., HOLZER, A. C., GOVAERTS, S., SANDOZ, D., AND GILLET, D. Contextual Learning Analytics Apps to Create Awareness in Blended Inquiry Learning. In *14th International Conference on Information Technology Based Higher Education and Training (ITHET)* (2015).

[114] WILLIAMS, L., AND COCKBURN, A. Guest editors' introduction: Agile software development: It's about feedback and change. *IEEE Computer*, 6 (2003), 39–43.

[115] WOLPERS, M., NAJJAR, J., VERBERT, K., AND DUVAL, E. Tracking actual usage: the attention metadata approach. *Journal Of Educational Technology and Society 10*, 3 (2007), 106–121.

[116] WYCHE, S. Exploring mobile phone and social media use in a Nairobi slum: a case for alternative approaches to design in ICTD. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development (ICTD)* (2015), ACM, p. 12.

[117] ZALDIVAR, V. A. R., GARCÍA, R. M. C., BURGOS, D., KLOOS, C. D., AND PARDO, A. Automatic discovery of complementary learning resources. In *Towards Ubiquitous*

Learning - 6th European Conference of Technology Enhanced Learning (EC-TEL) (2011), pp. 327–340.

[118] ZHANG, H., ALMEROTH, K., KNIGHT, A., BULGER, M., AND MAYER, R. Moodog: Tracking students' online learning activities. In *Proceedings of World conference on educational multimedia, hypermedia and telecommunications* (2007), vol. 25, pp. 4415–4422.

[119] ZHANG, H., DE CHOUDHURY, M., AND GRUDIN, J. Creepy but Inevitable? The Evolution of Social Networking. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW)* (2014), ACM, pp. 368–378.

Andrii Vozniuk

version: 05.01.2017

Andrii@Vozniuk.com
Lausanne, Switzerland
linkedin.com/in/vozniuk
go.epfl.ch/vozniuk-publications
slideshare.net/vozniuk
github.com/voz

EDUCATION

École Polytechnique Fédérale de Lausanne (EPFL) — 2011- 2017

PhD in Computer Science

Thesis: Enhancing Social Media Platforms for Educational and Humanitarian Knowledge Sharing: Analytics, Privacy, Discovery, and Delivery Aspects.

Kiev Polytechnic Institute (NTUU KPI) — 2003 - 2009

MSc and BSc with distinction in Applied Mathematics

RESEARCH EXPERIENCE

Go-Lab Project — 2012 - 2016

Research Assistant

Go-Lab was an FP7 EU-funded project investigating inquiry-based learning with remote laboratories. I studied how social media platforms can be enhanced to better support educational scenarios by providing learning analytics, privacy management, and recommendations. I actively participated in the design and development of the Graasp platform and integrated my proposals into it.

Graspeo Project — 2014 - present

Research Assistant

Graspeo is a project co-funded by MSF investigating next-generation knowledge sharing platforms for humanitarian organizations. I study how to support humanitarian knowledge workers with knowledge analytics and discovery, and how to enable knowledge delivery into underconnected environments.

SpeakUp Project — 2013 - present

Research Assistant

SpeakUp is a project co-funded by EPFL and UNIL investigating ways to improve audience-speaker interaction. I study how the integration of real-time discussion technology impacts the students' participation in a classroom. I have built the back-end of the system and am involved in its design and development.

TEACHING & MENTORING

Supervised over 10 student projects
Reviewed over 50K lines of code in various languages
Teaching Assistant and Lecturer at the Social Media course

HONOURS

EPFL FIT Innogrant CHF100'000 holder as a Graspeo co-founder	2015
Venture Kick Stage I and II winner as a Graspeo co-founder	2013 - 2014
Start Lausanne Stage II finalist as a Graspeo co-founder	2013
EPFL fellowship PhD student	2011 - 2012
Winner of Regional and finalist of National Ukrainian Physics competitions	2000 - 2003

PUBLICATIONS

The up-to-date list is available at: go.epfl.ch/vozniuk-publications

1. **A. Vozniuk**, A. C. Holzer and D. Gillet. Peer Assessment Dataset, in Journal of Learning Analytics, vol. 3, num. 2, p. 322-324, 2016.
2. D. Gillet, **A. Vozniuk**, M. J. Rodriguez Triana and A. C. Holzer. Agile, Versatile, and Comprehensive Social Media Platform for Creating, Sharing, Exploiting, and Archiving Personal Learning Spaces, Artifacts, and Traces. The World Engineering Education Forum, Seoul, Korea, November 6-10, 2016.
3. B. A. Schwendimann, M. J. Rodriguez Triana, **A. Vozniuk**, L. P. Prieto and M. Shirvani Boroujeni et al. Perceiving learning at a glance: A systematic literature review of learning dashboard research, accepted in IEEE Transactions on Learning Technologies, 2016.
4. A. C. Holzer, **A. Vozniuk**, S. Bendahan and D. Gillet. Rule of Thumb – Effect of Social Button Icons on Interaction. 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, Tuscany, Italy, September 6-9, 2016.
5. **A. Vozniuk**, M. J. Rodriguez Triana, A. C. Holzer and D. Gillet. Combining Content Analytics and Activity Tracking to Identify User Interests and Enable Knowledge Discovery. 6th International Workshop on Personalization Approaches in Learning Environments (PALE), Halifax, Nova Scotia, Canada, July 13-16, 2016.
6. B. A. Schwendimann, M. J. Rodriguez Triana, **A. Vozniuk**, L. P. Prieto Santos and M. Shirvani Boroujeni et al. Understanding learning at a glance: An overview of learning dashboard studies. International Learning Analytics and Knowledge, Edinburgh, UK, April 25-29, 2016.
7. M. J. Rodriguez Triana, L. P. Prieto Santos, **A. Vozniuk**, M. Shirvani Boroujeni and B. A. Schwendimann et al. Monitoring, Awareness and Reflection in Blended Technology Enhanced Learning: a Systematic Review, accepted in International Journal of TEL.
8. A. C. Holzer, **A. Vozniuk**, S. Govaerts, H. Bloch and A. Benedetto et al. Uncomfortable yet Fun Messaging with Chachachat. The ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY), London, England, October 5-7, 2015.
9. M. J. Rodriguez Triana, A. C. Holzer, **A. Vozniuk** and D. Gillet. Orchestrating Inquiry-Based Learning Spaces: an Analysis of Teacher Needs. International Conference on Web-Based Learning, Guangzhou, China, November 5-8, 2015.
10. **A. Vozniuk**, M. J. Rodriguez Triana, A. C. Holzer, S. Govaerts and D. Sandoz et al. Contextual Learning Analytics Apps to Create Awareness in Blended Inquiry Learning. 14th International Conference on Information Technology Based Higher Education and Training, Lisbon, Portugal, June 11-13, 2015.
11. **A. Vozniuk**, A. C. Holzer, S. Govaerts, J. Mazuze and D. Gillet. Graspeo: a Social Media Platform for Knowledge Management in NGOs. Seventh International Conference on Information and Communication Technologies and Development (ICTD 2015), Singapore, Singapore, May 15 - 18, 2015.
12. M. J. Rodriguez Triana, S. Govaerts, W. Halimi, A. C. Holzer, C. Salzmann and **A. Vozniuk** et al. Rich Open Educational Resources for Personal and Inquiry Learning. Agile Creation, Sharing and Reuse in Educational Social Media Platforms. International Conference on Web & Open Access to Learning, Dubai, United Arab Emirates, November 25-27, 2014.
13. **A. Vozniuk**, S. Govaerts, L. Bollen, S. Manske and T. Hecking et al. AngeLA: Putting the Teacher in Control of Student Privacy in the Online Classroom. ITHET 2014, York, England, September 11-13, 2014.
14. A. C. Holzer, S. Govaerts, **A. Vozniuk**, B. Kocher and D. Gillet. SpeakUp in the Classroom: Anonymous Temporary Social Media for Better Interactions. ACM CHI Conference on Human Factors in Computing Systems, Toronto, Canada, April 26 - May 1, 2014.

PUBLICATIONS:CONTINUED

15. **A. Vozniuk**, A. C. Holzer and D. Gillet. Peer Assessment Based on Ratings in a Social Media Course. The 4th International Conference on Learning Analytics and Knowledge LAK14, Indianapolis, Indiana, USA, March 24-28 2014.
16. S. Govaerts, Y. Cao, A. **Vozniuk, A**. C. Holzer and D. Garbi Zutin et al. Towards an Online Lab Portal for Inquiry-based STEM Learning at School. The 12th International Conference on Web-based Learning (ICWL 2013), Kenting, Taiwan, October 6-9, 2013.
17. A. C. Holzer, S. Govaerts, J. Ondrus, **A. Vozniuk** and D. Rigaud et al. SpeakUp – A Mobile App Facilitating Audience Interaction. The 12th ICWL, Kenting, Taiwan, October 6-9, 2013.
18. **A. Vozniuk**, S. Govaerts and D. Gillet. Towards portable learning analytics dashboards. International Conference on Advanced Learning Technologies, Beijing, China, July 15-18, 201.

WORK EXPERIENCE

Seance Association
2013 - present

Co-founder, Software Engineer & Treasurer

Graspeo
2013 - 2015

Co-founder

Jaccomo
2009 - 2010

Software Engineer

Center for Medical & Biotechnical Research NAS of Ukraine
2008 - 2009 part-time

Software Engineer

Research Center Jülich
2006 - 2008 part-time

Research Intern & Software Engineer

Institute of Mathematics NAS of Ukraine
2006

Research Intern

SOFTWARE ENGINEERING SKILLS

Systems Architecture
Data Engineering
Analytics Infrastructures
Cloud & Scale
Data Visualization
Web Development
Development Tools and Processes

OTHER

Languages
English, French (B1/B2), Ukrainian, Russian.

Volunteer Activity
Software engineer & treasurer at the Seance Association building free educational technology.