Next-Lab

Next Generation Stakeholders and Next Level Ecosystem for Collaborative Science Education with Online Labs

Collaborative Project in European Union's 2020 research and innovation programme

Grant Agreement no. 731685

Deliverable D5.2

Data Management Plan

Editor(s)	Denis Gillet, Adrian Holzer (EPFL)
Date	30 June 2017
Dissemination Level	Public
Status	Final

The Next-Lab Consortium

Beneficiary Number	Beneficiary name	Beneficiary short name	Country
1	University Twente	UT	The Netherlands
2	École Polytechnique Fédérale de Lausanne	EPFL	Switzerland
3	IMC Information Multimedia Communication AG	IMC	Germany
4	EUN Partnership AISBL	EUN	Belgium
5	Ellinogermaniki Agogi Scholi Panagea Savva AE	EA	Greece
6	University of Cyprus	UCY	Cyprus
7	Universidad de la Iglesia de Deusto	UD	Spain
8	Tartu Ulikool	UTE	Estonia
9	Núcleo Interactivo de Astronomia Associacao	NUCLIO	Portugal
10	Ecole Normale Superieure de Lyon	ENS de Lyon	France
11	Turun Yliopisto	UTU	Finland
12	University of Leicester	ULEIC	United Kingdom

Contributors

Name	Institution
Diana Dikke (internal reviewer), Evgenij Myasnikov	IMC
Adrian Holzer, Denis Gillet, María Jesús Rodríguez-Triana, Andrii Vozniuk, Alex Wild, André Nogueira, Juan-Carlos Farah	EPFL
Effie Law, Matthias Heinz	ULEIC
Ton de Jong (internal reviewer)	UT
Evita Tasiopoulou	EUN
Jens Koslowski (internal reviewer)	EA

Legal Notices

The information in this document is subject to change without notice.

The Members of the Next-Lab Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the Next-Lab Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Executive summary

This deliverable presents the Data Management Plan of the Next-Lab project. Data in Next-Lab can be broadly divided in four categories: (1) platform content data, (2) platform usage data, (3) activity data and student output data, and (4) feedback data.

Platform content data mainly consists of data created by users on the Next-Lab sharing platform (Golabz) and the Next-Lab authoring platform (Graasp), such as Spaces, documents, links, discussions, etc. This data is essential for the Graasp and Golabz services to work. This data is mostly linked to login credentials (name, email, password) and possibly user profiles. User names are accessible to anyone on the platform, but emails and passwords are kept private.

Platform usage data consists of general Google Analytics traces on Graasp and Golabz. This data is used to provide the European Commission with evidence of impact. This data is not linked to identifiers such as email or names.

Activity data and student output data consists of activity traces of teachers and students, as well as of student productions (e.g., a concept map, a pdf report) inside an Inquiry Learning Space (ILS). Activity traces are used to provide feedback through teacher dashboards (e.g., Kibana, teacher Analytics apps) and student Learning Analytics apps. Student output data can be linked to a Nickname or be anonymous, depending on the settings of the ILS. Activity traces in an ILS are only recorded if the AngeLA learning analytics angel is present as a member of the ILS. If AngeLA is removed, no activity traces are recorded in the ILS.

Feedback data consists of data generated by the Go-Lab Community and Event interactions on Graasp, participatory design (PD) activities, help desk support activities and impact evaluation activities. The interaction of teachers and project partners through the Go-Lab Community and Event spaces is similar to other content data on Graasp with the exception of registration data that community and event members fill in. This form can contain their emails which - if so specified by users - are visible to space owners. Helpdesk support data and PD data deal in general with issues raised by users. This data is mainly used to better understand the needs of users and to fine tune the Next-Lab services to fit these needs. Furthermore, the data is also used to provide evidence to the European Commission on the performance of the project.

In order to ensure data preservation, the Next-Lab ecosystem runs on cutting edge infrastructure with full backup strategies. Data access is closely monitored (details are provided in this deliverable) to mitigate data security risks. Selected anonymized data and analytics will be extracted from our database for reporting to the European Commission or for publication in scientific venues. In the spirit of the open science movement, such data will be shared under Creative Commons CC-BY-NC.

Table of Contents

1. Introduction

This deliverable describes the data management plan for Next-Lab and the issues related to the collection, the exploitation, and the storage of data.

The objectives of data management are threefold:

1. Ensuring access to research data (open science) that can be used in studies conducted by Next-Lab partners in the framework of investigations related to the Next-Lab Innovation Action.

2. Assessing the qualitative and quantitative impact of Next-Lab for the European Commission.

3. Enforcing the European data protection rules[1] (to be implemented by May 2018), which are bringing additional restrictions on collecting, storing, exploiting, and disclosing data to ensure data protection and security. In a nutshell, these rules are mainly about increasing user awareness about what data is tracked, who has access to it, and what is done with it; requesting informed consent to data collection and exploitation; as well as accessing and controlling their data, being able to correct inaccuracies and delete them.

These objectives are complementary but also contradictory at times. For instance, on the one hand, Objectives 1 and 2 argue for the tracking and storage of as much data as possible including content data, activity data, user opinions and feedback. These objectives, but especially Objective 1, also argue to make the data freely and publicly available. Objective 3, on the other hand, constrains collection and usage to predefined purposes and limits the dissemination to predefined stakeholders.

To better understand these dimensions, a typical scenario illustrating the usage of the Go-Lab[2] ecosystem as promoted in the Next-Lab Innovation Action is detailed below.

1.1   Usage scenario

The typical usage scenario illustrating the data management elements of Next-Lab is the following. A teacher (from any country) discovers an interesting educational resource (e.g., an online lab) on the golabz.eu sharing platform (henceforth Golabz) which can be pursued as follows:

1. The teacher can freely use this online resource alone with his or her students without providing any identification on the Golabz sharing platform.

2. If the teacher wants to personalize (configure, embed) this resource (s)he needs to create an account on the Graasp authoring platform (graasp.eu). A full name, an email address, and a (encrypted) password are requested. The email and password are kept as credentials for further access.

3. With the Graasp account, the teacher can personalize the resource and share it as a single standalone Web page with selected students (typically, the students of one of her or his classes) using a secret URL. External Web applications, resources or

---

[1] European General Data Protection Regulation (GDPR) (EU) 2016/679: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

[2] Note that in terms of branding, Next-Lab is only used to define the co-funding source of the project. The community and technology are still referred to as the Go-Lab community and the Go-Lab ecosystem.

services can be freely integrated by the teacher in the open educational resource (OER), which is referred to as an online inquiry learning space (ILS).

4. Two dimensions can be configured by the teacher before sharing the ILS with the selected students:

    a. activity tracking can be enabled or disabled by inviting or not a virtual learning analytics agent (AngeLA) explicitly represented as a member of the space;

    b. access for students can be set as anonymous, nickname only, or nickname and password.

5. Activity traces and student outputs are kept in the space where they are created under the full control of the teacher.

6. Learning analytics visualization can be freely enabled by the teacher for self-awareness and reflection.
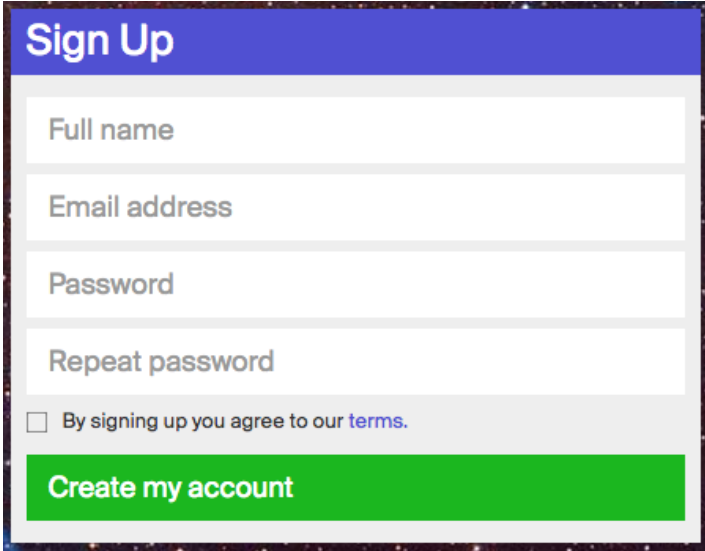
1.2 Roadmap

Broadly the data collected in Next-Lab can be divided into four categories, which guide the structure of this deliverable. Section 2 presents *platform content data*, which consists mainly of data created by teachers on Graasp and Golabz. Section 3 discusses *platform usage data*, which consists of Google Analytics traces on Graasp and Golabz used to provide evidence of impact to the European Commission. Section 4 presents *activity data and student output data*, which consists of teachers' activity traces and students' activity traces, as well as content produced by students in ILSs. Section 5 discusses *feedback data*, which consists mainly of surveys, participatory design data and data from the interactions on the Next-Lab helpdesk. Finally, Section 6 then discusses the data preservation issues before Section 7 wraps up with a conclusion.

2. Platform Content Data

Platform usage data includes all user data stored by software components used in Next-Lab. These include the Sharing Platform (golabz.eu) and the Authoring Platform (graasp.eu).

Graasp content data contains user data and data generated and uploaded by users. When signing up, users (typically teachers) provide email, full name, and password (as shown in Figure 1), the latter are saved in an encrypted format in the database. Data generated by users can contain anything from text to binary files. The content data in Graasp is organised in spaces, which can be described as online folders with permissions. A space has a list of members (owners, editors, viewers) and can contain a subspace hierarchy, links, documents, discussions and apps. Each one of these items also contains a description and all contain associated metadata (e.g., timestamps, creator id, file size). Users can also populate their profile in Graasp, which contains their usernames, possibly a picture, and a description. The database also stores the nicknames and sometimes passwords of users (typically students) who logged in through the standalone (student) view. Teachers are informed that to preserve anonymity, students should not use their real names as nicknames and they should change nicknames frequently (but not within one ILS or they will lose their data). Space owners can delete any content from the space. Once deleted, no copy of the data is kept on the server.

Figure 1. Graasp Sign Up dialogue.

Golabz receives user data from Graasp when a user (typically a teacher) logs in to Graasp from Golabz or when a user publishes an ILS from Graasp to Golabz. These data include: username, email, and Graasp user-ID. Golabz does not receive any data of the students. For the consortium members and external online lab and app providers, the accounts are created by the system administrator from the project consortium (Golabz administrator). These accounts contain email, username, and password (which is changed by the user when logged in for the first time).

2.1 Platform Content Data Consent

At the platform usage data level, for all those users obliged to enter personal data (i.e., typically teachers), an on-line consent form is used to provide information before users sign up to the platform. The users have at their disposal the description of terms and conditions in Graasp, which is the central place where users sign up (see Figure 2).

Figure 2. Consent form informing users when they sign up on Graasp.

2.2 Platform Content Data Storage

Graasp related data is stored in a secured data center on the EPFL campus in Lausanne, Switzerland. This data is backed up every day on a NAS provided by the data center.

Golabz data (incl. all content saved in Golabz and its metadata provided by online lab and app owners, like name and description of the software, screenshots, etc.) is stored at a HostEurope server (www.hosteurope.de/en/Server/Virtual-Server/; Enterprise tariff). The virtual server is hosted in datadock in Strasbourg, which fully complies with all quality and safety standards of Germany. HostEurope makes an automatic daily backup of the data; it is also possible to create snapshots to determine dates of backups and restorations. The data is also regularly saved locally at IMC AG, Saarbrücken, Germany. HostEurope assures an average availability of its servers of 99.9%. Using monitoring features, it is possible to supervise the running of the services and ports.

2.3 Platform Content Data Access

Graasp. The data in Graasp, like many cloud services, can either be accessed through regular usage by platform users or through database query by platform administrators.

- Regular usage access
  - *Private space data:* data uploaded to a space can be accessed by any members of that space with the adequate access rights (owner, contributor, viewer).
  - *Public space data:* data located in spaces set to *public* are accessible to anyone online.
  - *User profile data:* User profiles are public and accessible to anyone, but user emails are not accessible[3].
  - *Student data:* Data uploaded by students through the Standalone View of an ILS are accessible by space members (typically teachers).

- Database query access
  - The Graasp database can only be accessed by the Graasp platform managers (as of June 2017, Alex Wild, André Nogueira, Andrii Vozniuk, and Juan Carlos Farah), WP2 leader (Maria Jesus Rodriguez Triana) and the deputy coordinator (Denis Gillet). All these people are under EPFL contract and have to comply with the EPFL data management policy guaranteeing confidentiality. They lose their access if they leave EPFL.

Golabz data can be accessed by the Golabz platform managers (Evgenij Myasnikov, Diana Dikke). All these people are under IMC contract and have to comply with the IMC data management policy guaranteeing confidentiality. They lose their access if they leave IMC.

2.4   Platform Content Data Usage

The platform content data is stored primarily to allow the platform to function (i.e., user profile, content and activity traces are stored in order to allow users to exploit their personal and shared spaces, and to provide them with analytics and recommendations). It is also used by WP1 to provide analytics to partners, ambassadors and the European Commission about the project impact. The current script extracting analytics on the Graasp.eu database is listing the following information organized by tabs:

- *Users per day*: Date, number of standalone users (students), number of users until this date, min, max, average, mean.

- *Users per country*: Country, number of registered users (teachers), number of creators (having created inquiry learning spaces), and number of potential[4] implementers (having created inquiry learning spaces used by a certain number of students).

- *Long tail*: Number of inquiry learning spaces versus their number of standalone users.

- *Evolution per month*: Number of registered users, number of standalone users, number of inquiry learning spaces (existing, created, co-authored, implemented with more than 5 or 10 students).

- *Co-authoring*: number of created, implemented and published ILS that were co-authored by teachers, or by teachers and Next-Lab members.

---

[3]   There is an exception to this claim in the case of event registration where Event organizers can export the list of emails of the participants given that they have provided their consent.

[4]   From the usage data it is not possible to know for sure that an ILS has been used in class by students. We infer the usage through the number of standalone users of an ILS. This might results in false negatives (real implementations with only a few students) or false positives (non implementations, such as ILS testing, with a lot of standalone users).

- *Implemented inquiry learning spaces*: Space ID, creation date, author category (project or external), space type, space language, published or not on the public repository (golabz.eu), number of copies, number of owners, number of editors, number of viewers.

- *User list*: Anonymized user ID (different from the internal user ID stored in the Graasp.eu database), country, registration date, account used (Facebook, Google+ of Graasp), language, number of ILS created, number of standalone users.

- *Apps and labs*: Number of times each app/lab was embedded in an ILS, created and implemented ILSs where the app/lab was embedded, users who embedded the app/lab, in general, in their ILSs and, in particular, in the potentially implemented ones.

These anonymized analytics are only accessible by the project partners (as an excel file) for the duration of the project. The raw data (see Section 2.3) exploited to produce these analytics are not shared with project partners or anyone else.

3. Platform Usage Data

Platform usage data is *anonymous* interaction data collected through mainstream tracking services installed on Graasp and Golabz i.e., Google Analytics. The data is anonymous in the sense that it is not linked to specific user identifiers.

Google Analytics: usage data on Google Analytics contains anonymized website traffic and navigation on Golabz and Graasp. Figure 3 shows the type of live information shown with Google analytics, whereas Figure 4 shows longitudinal data (here from January 1st to May 28th 2017). The Google Analytics data is stored by Google and comply to its own terms & conditions.[5] No explicit consent is given by people who do not sign up. However the terms and conditions inform users that platform usage data is collected. Furthermore users can block Google Analytics through browser plug ins, such as Ghostery.[6]
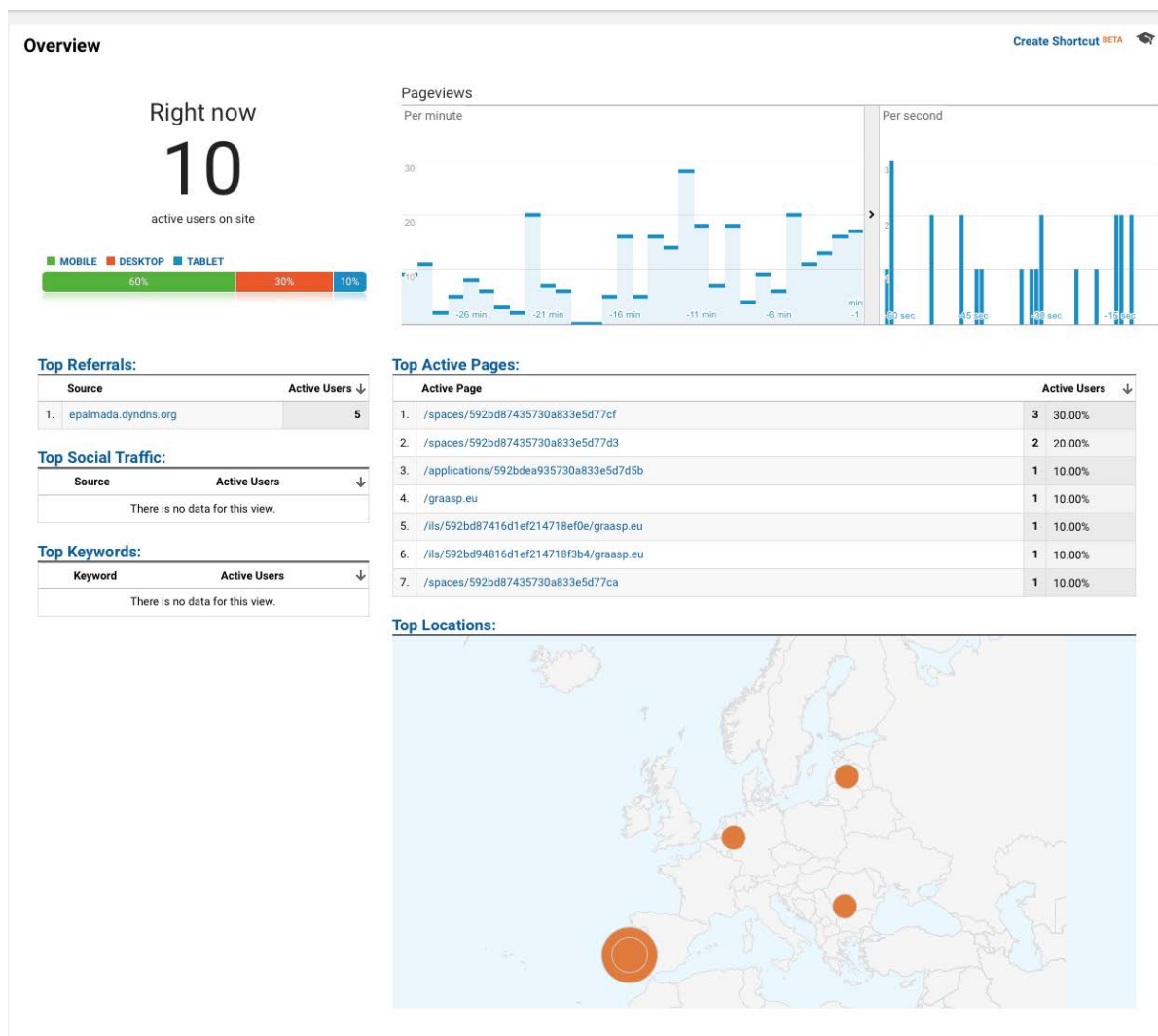
Figure 3. Live Google Analytics data on platform usage

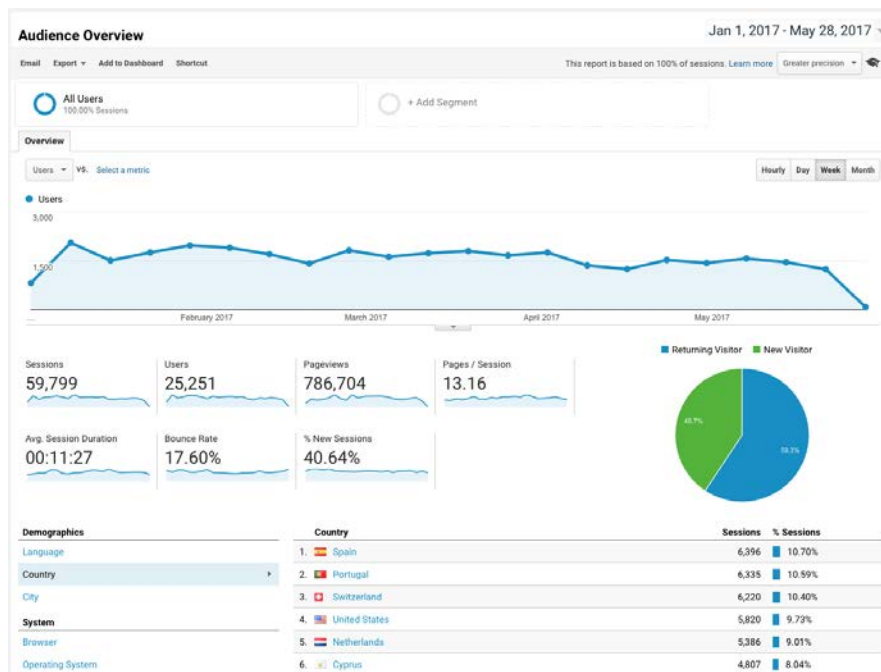[5]   https://www.google.com/analytics/terms/us.html

[6]

Figure 4. Google Analytics data on platform usage over time

The platform usage data is stored in order to provide usage statistics to the European Commission, partners and ambassadors by WP1. More concretely, Google Analytics help us to monitor project metrics such as number of visits per platforms, the number and length of the session and the bounce rate, as well as the number of users per country and city in a given period of time. Live data is also used in order to avoid making changes on the server when users are online. The Google analytics are only accessible by the platform managers (EPFL for graasp.eu and IMC for golabz.eu) and by the Next-Lab Coordinator for the duration of the project. No one else can request and get access to the corresponding google analytics accounts. However, synthetic graphs are shared with the project partners and with the European Commission to show them the overall impact of the project.

4. Activity Data and Student Output Data

Activity data is interaction data linked to specific user identifiers in the platform and used for a twofold purpose: first, to provide awareness and reflection services back to users through learning analytics apps and activity dashboards; and second, to keep track to the current status of the students work so that, when they open a new session, they can continue working on their ILSs (providing they use the same nickname). This activity data is also linked to platform content and more specifically, learning analytics apps can be linked to student outputs.

Graasp user activity (mainly teachers) contains actions performed by users inside a space, such as accessing an item, creating an item, deleting or modifying an item.

ILS user activity (mainly students) contains activity traces of standalone users. This activity relates to actions in the different inquiry learning apps and labs that support user tracking. The apps and labs can be both producers of activity data and consumers of activity data (e.g., to show which students are online, for example). Note, that a central feature of the Go-Lab ecosystem is that it allows users (teachers) to aggregate third party apps and labs into their learning spaces. How these apps and labs handle their data is not the responsibility of the Next-Lab consortium. Nevertheless, apps added to a space can only access data from other items in that space if the AngeLA activity tracker is enabled (teachers can disable it).

AngeLA. AngeLA, the learning analytics angel (agent), is a visual representation of the learning analytics tracking mechanism as a member of an ILS. If AngeLA is present in a space, then Student activity will be tracked and made available to LA apps. If AngeLA is not present, student activity is not tracked.[7] This implies that some apps will not work up to their full potential. Note that currently AngeLA sends activity traces to both the Vault[8] and a Learning Analytics backend located at the University of Duisburg Essen in Germany. This architecture is a leftover from the Go-Lab project in which Duisburg Essen was a partner. We are currently in the process of moving this backend on the Graasp infrastructure. Finally, adding AngeLA to the space is Opt Out for now, but we will change to Opt In in 2018 to comply with new EU privacy regulations.

ILS user output (mainly students) contains student productions, such as reports that they might have uploaded, or concept maps or other artefacts that they might have created within apps and labs. Again, apps and labs can both be producers and consumers of ILS user outputs.

4.1 Activity Data and Student Output Data Consent

Activity data is encompassed by the terms and conditions Graasp users (teachers) agree to when signing up. ILS users (students) do not sign up and thus do not formally provide their consent. However, like with other learning artefacts, the teachers are in charge of making choices for their students.

---

[7]   Note that even if one removes AngeLA, one can still access old data generated when it was activated.

[8]   The Vault is a subspace of an ILS where all the student outputs and the traces from the inquiry learning apps are stored.

4.2   Activity Data and Student Output Data Storage

Activity data is stored in Graasp. Student output data is stored in the Vault in the ILS (also in Graasp). Traces are digital log data stored in the Graasp.eu database in the form of a timestamped and contextualized (i.e., associated with a dedicated inquiry learning space) triplet of actor, verb, object which does not need calibration. The vocabulary is embedded in the platform, so there is no risk of vocabulary misuse (ActivityStreams and xAPI standard vocabulary).

An example of the raw data is: On "Date" (timestamp), "Anonymized_Actor_ID" (actor), "downloaded" (verb), "Object_ID" (object) in "Space_ID" (context).

An example of associated analytics will be: "Space_ID" has been accessed by "Access_Count" users of type "User_Type" from "Country_ID" in the period "Period_Descriptor", which are open Web standards typically exploited to provide learning analytics.

Data related to students and their activities are stored by design in an anonymous form in the Graasp.eu database, i.e., the actual identity of the students is never requested and they are only identified with nicknames they can change at their convenience and which can be different in each inquiry learning space created to support a different supervised classroom activity or learning session.

In future work, we plan to allow users to select their own learning record repository to store activity traces and learning outcomes (outputs). Additionally, we aim to provide the functionality to validate these records using blockchain or other cryptographic technologies. This will ensure that users cannot tamper with the contents of their learning repositories or falsify their educational records in addition to guaranteeing privacy.

4.3   Activity Data and Student Output Data Access

Graasp user activity for a specific space can be accessed and visualized through the Kibana dashboard by space owners (teachers) as shown in Figure 5.
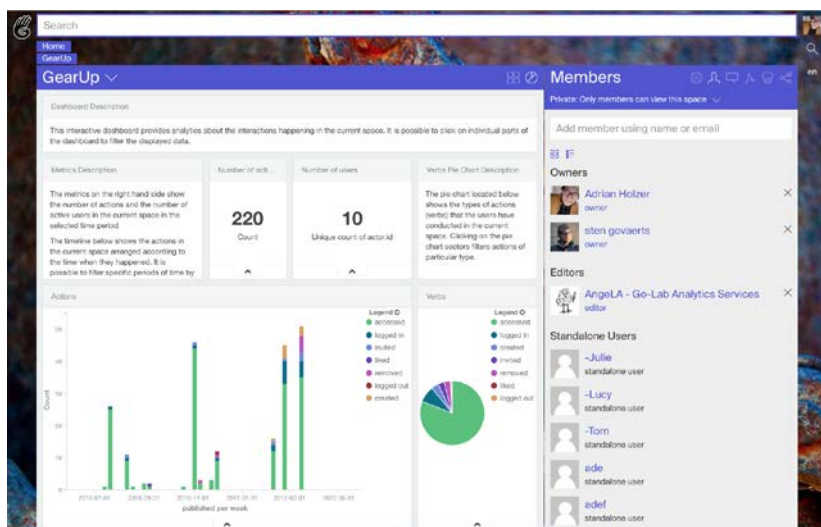

Figure 5. Kibana dashboard showing activity in a space.

Teachers, as owners of an ILS, have access to **student data** located in the Vault.

The **Graasp database,** where all activities are currently stored, can be accessed by the Graasp platform managers (see Section 2.3).

4.4   Activity Data and Student Output Usage

The trace data which will be extracted on a weekly or monthly basis from the Graasp database for assessing the impact of the project (as requested by the European Commission) or for scientific investigations will be anonymized during the extraction process and delivered as a file in the Excel or csv format (around 3MB per month). Naming will include the platform name and the date, i.e., Graasp.eu_Day_Month_Year.xlsx (or csv). The current script extracting data on the Graasp.eu database is listing the following information organized by tabs:

Graasp space activity: Space ID, creation date, number of actions for each Activity Streams or xAPI verb.

Labs and apps usage: Number of times an app or a lab on Golabz has been added in inquiry learning spaces.

Publish ILS: Inquiry learning spaces which have been created in Graasp and published by their owner(s) on the golabz.eu repository.

5. Feedback Data (pen paper, event community)
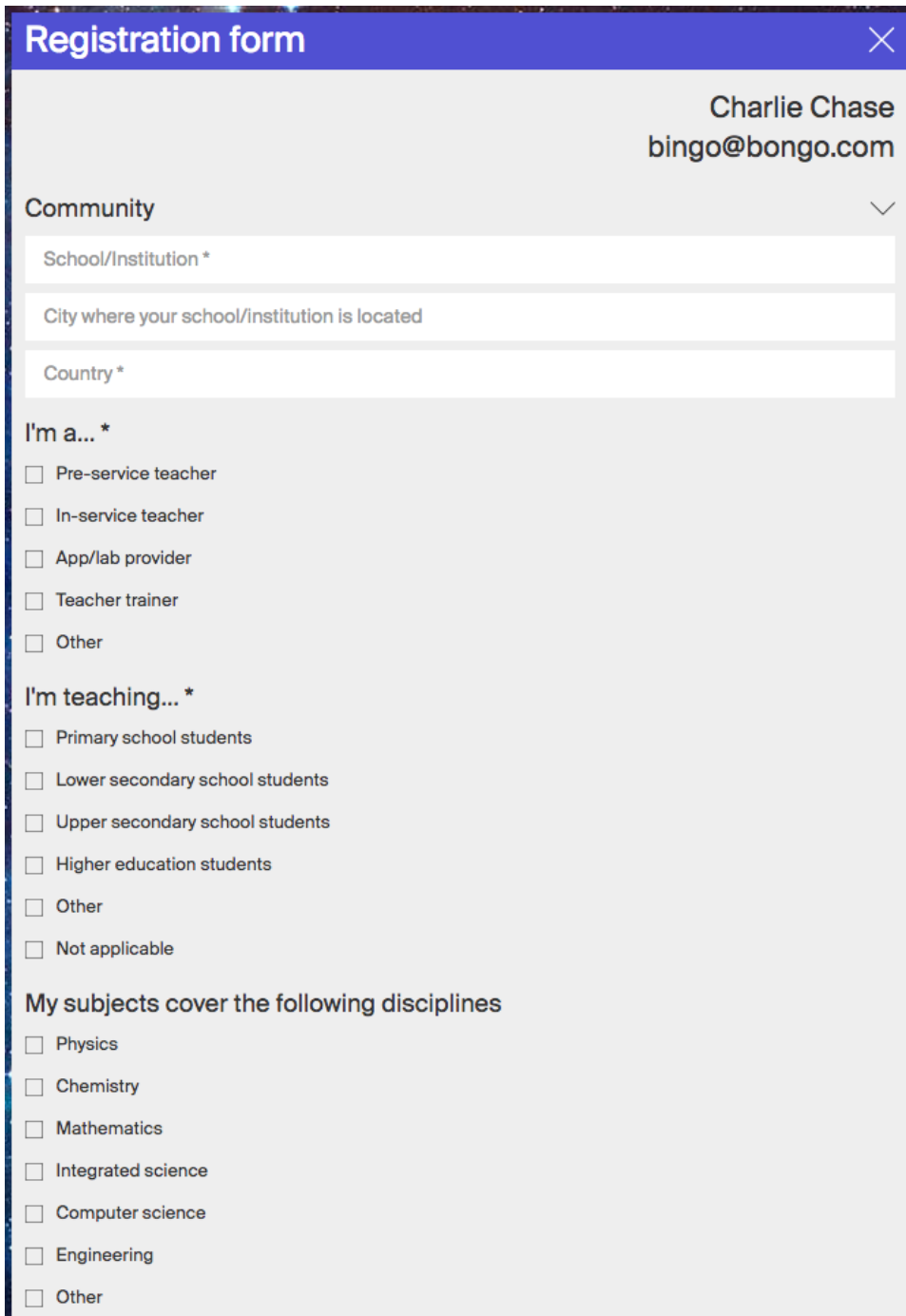
Under the umbrella of *feedback data* we consider not only feedback provided by the users e.g., data collected by WP1 and WP2 partners, as well as ambassadors in the event they organized, by the PD team, or by other partners for targeted research investigations, …) but also problems and questions asked by them (e.g., via the helpdesk). This data is used mainly to reinforce the co-design and to measure the impact of those functionalities and services offered in the project. Thus, ambassadors, partners and the European Commission will have access to the outcomes obtained from the data analyses.

Feedback and Participatory design (usability data) are gathered on:

- Graasp through the **Go-Lab Community space**[9]. As it is described in D2.1, the community space is used to support Events and peer interaction among community members (i.e., teachers and project partners). Teachers are typically invited to join the community in general or for a particular event (which automatically adds them to the community). When invited to join the community they fill in a registration form[10] shown in Figure 6. This form has a threefold purpose: collect the user profile (essential to measure the impact in T1.4); get the informed consent to use anonymous data regarding the activities carried out in the project and the platforms for research and improvement purpose; register under which conditions the users are joining the community. Apart from helping us to keep the community updated (e.g., sending information regarding training events and platform updates to those who subscribe), the conditions for joining the community allow us to detect users willing to provide user feedback (up to 2 questionnaires per year for assessing features and services).

---

[9]   Go-Lab community: http://graasp.eu/communities/5841a65260bb68357bb1c384

[10]   Registration form: http://graasp.eu/form/58ebae9d16d1ef214765bac1

Figure 6. Go-Lab Community registration form.

Figure 7 shows that the event registration is just an extension of Community registration form, where we ask for consent to take pictures or record videos during the sessions.

Figure 7. Event registration form.

- **Intercom Helpdesk.** A direct line of support with the teachers in Next-Lab is provided through the Intercom Helpdesk. Intercom streamlines the creation and management of support tickets, allowing project partners to collaboratively answer questions and resolve issues put forth by current and potential users. Interactions that occur on Intercom are stored on Intercom's infrastructure. Intercom stores browser information such as language and location (as illustrated by Figure 8 which shows the Helpdesk users over the last 3 months). Furthermore, the names of Graasp and Go-Lab users are shared with Intercom, however the associated Graasp userID is not directly shared with Intercom, but it is hashed. This mechanism allows Graasp managers, but no one else, to make the link between Intercom users and Graasp users. A user can always sign up separately to Intercom to share his/her information, though this is not a requirement to access the helpdesk feature.
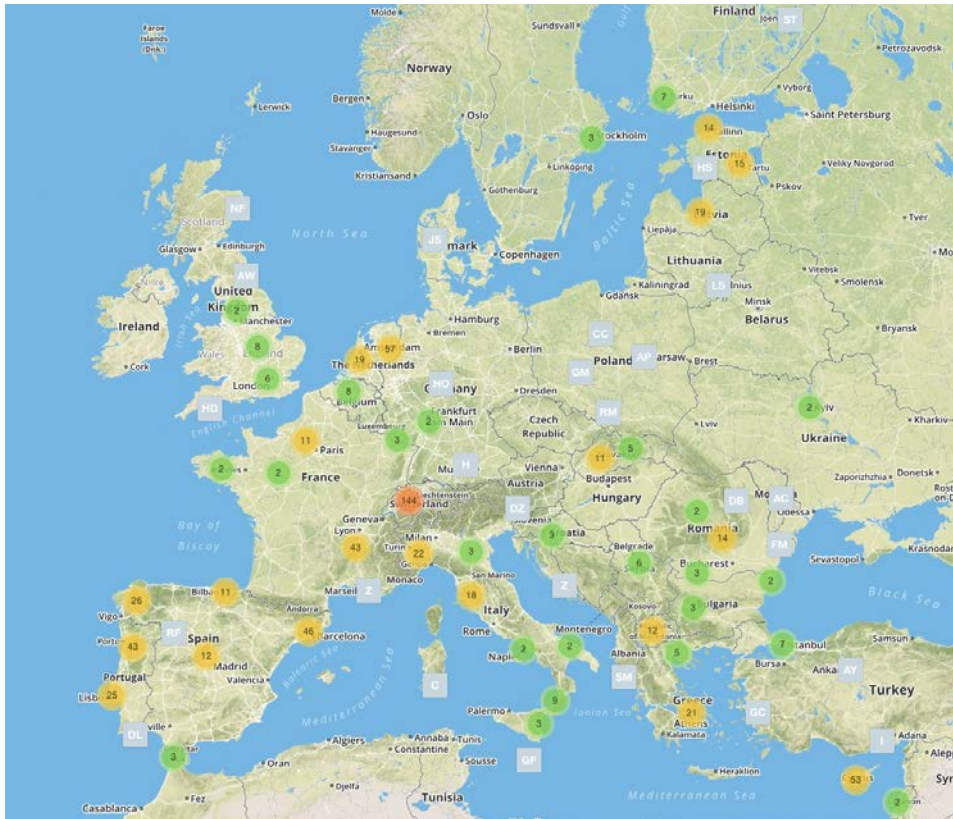
Figure 8. Intercom Helpdesk user location from March to June 2017.

- **Participatory design data.** Data will be gathered by means of interviews, observations, questionnaires, etc. either in face-to-face PD (participatory design) events or through online mechanisms (e.g., online questionnaires, PDotCapturer[11], etc.). PD data will be gathered anonymously, meaning it will not be linked with personal information of the participants providing it. For some inferential statistics and to get background information on the participants, general demographic data on the person (e.g., age) and their teaching/learning background (e.g., primary or secondary school) might be gathered and taken into consideration for the data analysis.

- **Ambassador Outreach Data.** We collect feedback data from outreach activities from Go-Lab ambassadors through online surveys. This data contains information about presentations (see Figure 9)[12] and social media dissemination (see Figure 10)[13] performed by the Ambassadors. These surveys also include personal data about the Ambassadors such as: name, surname, email address, city and country where they teach, school name, school postal address, subjects they teach. When it comes to the events/presentations/trainings they carry out as part of their outreach the following information is collected: type of activity, dates, country, city, language,

---

[11] Heintz, M., Law, E. L.-C., Govaerts, S., Holzer, A., and Gillet, D. (2014). Pdot: participatory design online tool. In CHI'14 Extended Abstracts on Human Factors in Computing Systems, pages 2581-2586. ACM.

[12] SurveyMonkey form used by Go-Lab Ambassadors to report on presentations, trainings: https://www.surveymonkey.com/r/next-lab_dissemination

[13] SurveyMonkey form used by Go-Lab Ambassadors to report on social media dissemination: https://www.surveymonkey.com/r/next-lab_socialmedia

name of the event, link to website (if available), type of participants, number of participants.

* Required

1. Please select your name from the dropdown menu.

[dropdown]

* 2. Type of dissemination activity:

◯ Presentation on Next-Lab (face to face)

◯ Next-Lab Teachers' training (face to face)

◯ Direct email to mailing list

◯ Article on website/newsletter/blog/print

◯ Other (please specify)

[text field]

* 3. Date of the dissemination activity:
For one day events, fill in the same date in both fields.

	DD	MM	YYYY
Starting date:	[ ] /	[ ] /	[ ]
Ending date:	[ ] /	[ ] /	[ ]

* 4. Country where the activity took place:
For online activities, please indicate the target audience.

[dropdown]

Other (please specify)

[text field]

* 5. City where the activity took place:
For online activities, please fill in N/A

[text field]

* 6. Language of the intervention:

[dropdown]

Please fill in as many fields as possible / relevant.

* 7. Name of the event:

The name of the conference, event, or magazine where the project was presented. For articles and webinars, please specify the web portal where they are published; for newsletters, please add the edition, specifying the issue date.

[text field]

* 8. Purpose of the event:

Provide a short description of the event and its aims.

[text field]

* 9. Link to the event page/ the event programme:

[text field]

* 10. Title of your dissemination activity:

Please add the title of your workshop/presentation (or webinar/article/email subject where applicable) as it appears in the event programme.

[text field]

* 11. Short description of the activity:

Please summarise the content of your presentation / workshop / webinar / article etc.

[text field]

* 12. Type of participants:
You can select more than one. If you are reporting about an article/publication, please select the target audience.

☐ Primary school teachers

☐ Secondary school teachers

☐ Researchers / Academic staff

☐ Pre-service teachers

☐ Project managers / project representatives

☐ Policy makers

☐ Representatives of private companies

☐ Other (please specify)

[text field]

* 13. Number of participants: (if you know the exact number just write the same number in both boxes)

Between [ ]

And [ ]

14. Provide here the links to relevant pictures and videos taken at the event.
If pictures from the event are not available online, you can also upload them to a Google drive folder or Dropbox and put the link to access them here.
Otherwise, you can send them by email to: enrique.martin@eun.org, including the date of the event, the country and city in the subject of the email.

[text field]

Figure 9. Ambassador presentation dissemination report surveys.

Figure 10. Ambassador social media dissemination reporting surveys.

5.1 Feedback Data Consent

- **Graasp Community and event spaces**. When joining the Go-Lab Community in Graasp, teachers:
  - must agree to let Go-Lab & Next-Lab use anonymous data regarding their activities in the project and the platforms for research and improvement purposes
  - can agree to let Go-Lab & Next-Lab send them questionnaires (max. twice per year) for assessing current and new features and services offered by golabz.eu and graasp.eu
  - can agree to let Go-Lab & Next-Lab send them information regarding training events and platform updates

When joining an event, they can also decide whether or not to appear in pictures or video recordings taken during the event for dissemination and research purposes.

- **Participatory design**. Teachers will be approached by the Next--Lab consortium on the basis of their experience with Go-Lab ecosystem, local contacts with schools using Go-Lab, longer standing cooperations, and/or specific user characteristics. Students (minors) will never be approached directly but always through their teachers or schools. All participants, irrespective of ages, are required to sign a consent form to protect their rights of participation in empirical studies. Of particular important is that they agree on the data so produced being published anonymously for research purposes and that they have full rights to withdraw from any study without the need of giving any reason. In case of interviews, questionnaires, and the online feedback mechanisms, participants give consent by participating in the data collection. In case of observations, consent will be gathered in advance. All participants participate on a voluntary basis. Participants (or when appropriate, their legal representatives) will be informed about the data gathering and the way the data are used (which will always be done anonymously). For participatory design and feedback data where minors are involved (mainly concerning the inquiry learning spaces) we use (passive) informed consent forms as they are in use at UT and ULEIC. ULEIC and UT data gathering is subject to prior approval of the ethics committees of these two institutions. As part of the consent procedure, teachers and students (and their legal representatives) will be informed about the goal of the study and the way the data will be processed and published. The information given will ensure that participants or their legal representatives have sufficient information to enable them to decide on their consent and it will explain in a clear way participant's rights.

- **Intercom Helpdesk.** Use of the Intercom helpdesk implies acceptance to Intercom's privacy policy[14]. If and when users sign up individually to Intercom, they will be prompted to accept this privacy policy, along with their terms of service. However, users do not provide their explicit consent when they simply use the service without signing in.

- **Ambassador Outreach Data.** The organizing of events is part of the Ambassadors tasks, which they agreed to fulfil as part of the MOU they have signed (see Figure 11). In the Open Call for the Ambassadors, teachers had to reply yes/no to some statements, including this one regarding their contact details: "Whether I am selected or not, European Schoolnet may contact me for other projects / events".

---

[14]   https://www.intercom.com/privacy

Figure 11. Ambassador MOU.

5.2   Feedback Data storage

Participatory design and feedback data with minors on Next-Lab learning spaces will be gathered by ULEIC and UT (and potentially additional partners), participatory design data and feedback data with adults will be collected by most partners under the leadership of EUN.

- **Graasp Community and event spaces**. The same applies to other Graasp data (see Section 2.3).

- **Participatory Design data** collected by ULEIC will be stored on ULEIC servers (in case of ULEIC online tools used to collect data) or servers of questionnaire service providers (e.g., Google Forms) for digital data collection. For paper-based data collection the feedback data will be stored in a locked office in the Informatics Department of the University of Leicester (for data collected by ULEIC).

- **Intercom Helpdesk.** As stated in their privacy policy, Intercom "complies with the EU-U.S. Privacy Shield Framework and the Swiss-US Privacy Shield Framework as set forth by the U.S. Department of Commerce regarding the collection, use, and retention of personal information from European Union member countries."

- **Ambassador Outreach Data.** The data is stored on the SurveyMonkey server under EUN's professional account.

5.3 Feedback Data Access

- **Graasp Community and event spaces**. The same applies to other Graasp data (see Section 2.3) with the exception of user registration data, which does not exist in regular Graasp spaces. Such registration information can be accessed by community and event owners.

- **Participatory design** Only the partner that performed the PD activity (mostly ULEIC) will have access to the raw data collected. On rare occasions they might share the data either with a partner to analyse the data (e.g., if a partner other than ULEIC conducts the event but the data analysis task lies with ULEIC) or with the partner developing the artefact of interest in the PD activity (e.g., if there are benefits of accessing the anonymized raw data over receiving a report).

- **Intercom Helpdesk** Besides the platform administrators at Intercom, 52 people involved in providing help to users have access to Intercom data. These include Ambassadors and project partners. Among the 52 people, 15 have full access (i.e, these are partners from EPFL, IMC, Nuclio, or EA), the others having restricted access (Intercom app settings, Intercom members and Intercom billing can't be accessed).

- **Ambassador Outreach Data.** Access to the SurveyMonkey data is accessible to the EUN team (Evita Tasiopoulou, Enrique Martin) and the person responsible for the impact (Task 1.4), i.e., María Jesús Rodríguez Triana.

5.4 Feedback Data Usage

- **Graasp Community and event spaces**. This data is used to report about the training events and provide partners, Ambassadors and the Commission with evidence on the project impact.

- **Participatory design**. For the most cases the results and outcome of PD activities will be analysed by the partner conduction the activity (or ULEIC) and shared with the respective partners in the form of anonymized and aggregated reports. These reports can be enhanced with quotes from the raw datasets where appropriate.

- **Intercom Helpdesk.** The Intercom helpdesk data is used to provide help to users who request it. In the future, we will use the data to better understand the recurring issues and provide FAQ type support to users. Finally, we use data to provide feedback to the European Commission on the workload and the performance of the Helpdesk.

- **Ambassador Outreach Data.** Ambassadors' personal info, contact, school details are collected in order to facilitate our communication with them and provide us with some demographics like the type of areas we cover, possible audiences we can attract and indicate needs that might arise in the future (i.e. travel limitations etc.). Events data is collected mainly for reporting purposes and for providing an as detailed as possible overview of the outreach activities that our Ambassadors are carrying out and their possible impact (when we combine this info with the metrics for examples, we might get some interesting insights). We also use this info to evaluate the Ambassadors performance and these reports with partly determine our future collaboration with them (we hold the right to replace them if they do not perform as agreed).

6. Data Preservation

Backups of the Graasp.eu server are kept at least for the full duration of the Next-Lab project (January 2017 to December 2019) plus one year. After the end of the project we will fall back in our usual backup scheme of keeping backup at least for one year (longer if human and IT resources available allow it). Backups of the extracted data files are also kept for the full duration of the project plus one year. The public data set will be kept according to the policy of the public scientific repository which will be selected in agreement with the project partners.

Thanks to these backups, the full database of the Graasp.eu platform can be regenerated at any time and analytics can be extracted.

The Graasp.eu server and one backup storage unit are currently stored in the data center of the SV building at EPFL. A second backup is made in the EE building also at EPFL. In July 2017, the graasp.eu server will be moved to another EPFL data center located in the MA building.

Backups of Golabz are kept for the full duration of the Next-Lab project plus one year after the project end. After that, the database will be archived locally at IMC and can be restored at any moment, if needed.

6.1 Formal information/data security standards

The Graasp.eu server is following the EPFL standard for open access platform and is audited yearly for possible intrusion risks. The computers of the Data Management Leader and of T1.4 Leader are not open on the Web and OS security patches are applied when available.

The HostEurope virtual server (where Golabz is hosted) is hosted in datadock in Strasbourg, which fully complies with all quality and safety standards of Germany. The datadock is one of the safest and greenest data centres in Europe and was awarded the highest possible rating of 5 stars in the recent eco Datacentre Star Audit. The computer of the Golabz system administrator is not open on the Web and security updates are applied regularly.

6.2 Main risks to data security

The only personal data which are stored in the Graasp database are emails and names or nicknames of the registered users. These emails are only accessible by the EPFL server managers and the Data Management Leader (see 2.3) belonging to the EPFL React Group developing the Graasp platform, all have a regular EPFL contract). Only these authorized managers with the password to access the server can see this information. The administrator password is changed regularly, including every time there is a change in the personal administrating the server. So, the main risks include an intrusion or a breach in the server (which is well protected against this) or a manager sharing intentionally or unintentionally the data (which could trigger a legal procedure).

As mentioned before, all data extracted from the server database for the Task T1.4 Leader (responsible for assessing the impact of the project) are anonymized during the extraction process (randomized identifiers). So, user identities are never made and are not available outside the server database (which needs it to provide its services).

Golabz stores the following personal data: users' emails, usernames, and passwords. These data is accessible for the Golabz system administrator and main developer (Evgenij Myasnikov, IMC) and Golabz product manager (Diana Dikke, IMC) both having permanent contract at IMC. The risk of an intrusion or a breach in the server is low, as the HostEurope servers are well protected against such attacks.

6.3   Open Science and Data sharing

All activity traces of the spaces in the Graasp.eu platform will be automatically recorded, except the data of the spaces in which tracking has been disabled by the users themselves (opting-in or opting-out for data tracking is available per ILS in Graasp through the AngeLA mechanism[15]). However, only selected analytics extracted from the digital logs will be shared for reporting to the European Commission or in case of associated scientific publications. A description of the data will be provided in the repository which will be selected in association with related publications. See an example [here](). The data will be curated and formatted according to relevant guidelines[16] and shared under Creative Commons CC-BY-NC.

No restrictions apply on the anonymized data, as there are no commercial dimensions in the Next-Lab project fully dedicated to promoting and exploiting open access platforms and open educational resources.

The Next-Lab Data Management Leader and the other authorized people who have access to data specified above will have to use the data accessible to them only for their contractual duties. They will have no rights to share their credentials to access such data with others and they will be responsible to keep them in a safe place.

---

[15] Note that normal spaces (as opposed to Inquiry learning spaces a.k.a ILS) do not provide an opt out mechanism, activities are always tracked.

[16] https://www.openaire.eu/h2020-oa-data-pilot

7. Conclusion

This deliverable presented the Data Management Plan for the Next-Lab Project. This deliverable highlighted how the Next-Lab consortium tackles the tension between storing and sharing the greatest amount of data in the spirit of open science and restricting data collection to the minimum to respect user privacy and ensure informed consent and data control.

This Data Management Plan has been evaluated and accepted by the EPFL Data Management Team and by the EPFL Ethics Committee working closely together. They focused on the authoring and exploitation infrastructure (graasp.eu and the associated storages of spaces, traces and learning outputs) which are under the responsibility of EPFL, while also considering the interplay with the other platforms, services, and data management activities. This process was helpful to understand the challenges and to develop best practices for an academic institution like EPFL in offering cloud services with worldwide open access.

As a matter of fact, privacy and ethics related to educational services and data are part of the research investigations carried out in the framework of the Next-Lab Innovation Action, and advances on these dimensions for open access digital education will be regularly reported in scientific publications acknowledging the co-funding of Next-Lab and in upcoming related Next-Lab deliverables.

Thanks to this document, the Next-Lab beneficiaries should have a clear understanding of their duties as services providers, data managers, or data consumers. We focused on having people individually listed with fully-defined responsibilities and duties. Only people requiring access to servers or data for the operation of the infrastructures or the exploitation of the data are granted with such an access. No complimentary access is granted to beneficiaries not requiring it for the completion of their tasks or to any third parties.