BREAKING EDUCATIONAL BARRIERS WITH CONTEXTUALISED PERVASIVE AND GAMEFUL LEARNING

Grant Agreement No. 687676

Innovation Action

ICT-20-2015

D1.8 Data Management Plan

Due date	Month 12
Actual date	Month 12
Deliverable author(s)	Jannicke Baalsrud Hauge; Jakob Baalsrud Hauge Nadera Sultana Tany; Anna Barenbrock
Contributing Partner(s)	BIBA, ATS, HWU, SEBIT, ORT, INESC
Version	1.0
Status	Final
Dissemination level	Public

Project Coordinator

Coventry University

Sylvester Arnab

Priory Street, Coventry CV1 5FB, UK

E-mail: s.arnab@coventry.ac.uk

Project website: http://www.beaconing.eu

This project has received funding under the Horizon 2020 Framework Programme of the European Union – Grant Agreement No. 687676

Version control				
Version	**Date**	**Author**	**Institution**	**Change and where applicable reason for change**
0.0	01.11.2016	J. Baalsrud Hauge	BIBA	ToC
0.1	02.12.2016	Jakob Baalsrud Hauge; Nadera Sultana Tany	BIBA	Initial Draft
0.2	13.12.2016	Nadera Sultana Tany	BIBA	Open Data Architecture Plan
0.3	16.12.2016	Jakob Baalsrud Hauge	BIBA	FAIR DATA, Integration of input from INESC
0.4	20.12.2016	Nadera Sultana Tany	BIBA	Elaboration on ODAP. Integration of input from HWU, ORT, SIVECO, SEBIT, ATS
0.5	20.12.2016	Jakob Baalsrud Hauge	BIBA	Draft Executive Summary
0.6	21.12.2016	Anna Barenbrock	BIBA	Captions, Spell Check, Layout
0.7	21.12.2016	Jannicke Baalsrud Hauge	BIBA	Final check, general improvements of text, sent to partners for cross-checking
0.7_ORT	22.12.2016	Francois Mohier	ORT	Remarks and Clarifying comments
0.7_ORT_SICVECO	22.12.2016	Marius	SIVECO	Update ethics
0.7_ORT_ATS	22.12.2016	Antoniu Stefan Ioana Stefan	ATS	Updates
0.8	23.12.2016	Anna Barenbrock	BIBA	Integration of remarks from ORT
0.9	29.12.2016	Anna Barenbrock	BIBA	Integration of review from HWU and ATS
0.9_baa	29.12.2016	Jannicke Baalsrud Hauge	BIBA	Clarification of final remarks and comments HWU, ATS, INESC, final check

Quality control				
QA Version	**Date**	**QA Responsible**	**Institution**	**Change and where applicable reason for change**
0.5	23/12/2016	Theodore Lim	HWU	Internal review
0.7	23/12/2016	Ioana Stefan	ATS	QM
0.9	30/12/2016	Jayne Beaufoy	COVUNI	Language check

Release approval				
Version	**Date**	**Name**	**Institution**	**Role**
1.0	30/12/2015	Ioana Stefan	ATS	QM

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

LIST OF ABBRIVATIONS

Abrv.	Description
API	Application Programming Interface
D	Deliverable
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable, Reusable
GLP	Gamified Lessons Path
LMS	Learning Management System

EXECUTIVE SUMMARY

The BEACONING consortium is specifically aware of the privacy issues the project is potentially facing. During the project, we have identified various needs of stakeholders, as well as of external entities concerning access to different data. However, since the BEACONING platform will offer personalised learning based on individual feedback, there are several issues related to privacy and data management that need to be regulated. In order to ensure the privacy of every participant and the integrity of the collected data, D1.8 Data Management Plan documents how the BEACONING consortium strives to achieve these goals. This deliverable describes in more detail the possibilities we have regarding data storage and access, as well as the restrictions for each pilot phase (according to national legislation in pilot countries and specific institution requirements). Since it has not been established in detail which data will be collected for each pilot, the description is still generic and will be better described as soon as the pilots are planned in more detail.

The Data Management Plan is developed according to the guidelines in the Horizon2020 Online manual. This document is the second of four deliverables regarding the Data Management Plan and ethics. The first one (D1.7 Data Management and Ethics Process Plan) was released at month 6, the next versions will be released in month 24 and 36.

1 INTRODUCTION

This section highlights the unresolved Data Management aspects that are yet undecided in D1.7. The deliverable provides an overview of the most relevant regulations for the BEACONING data management plan. The consortium has developed an Open Data Architecture Plan for the partners to abide. However, all pilot partners also have to consider their pilot specific needs and regulations. Consequently, the document also provides an overview of these.

1.1 SCOPE OF THIS DELIVERABLE

This document presents the Open Data Architecture Plan developed by the consortium and the pilot specific needs related to national legal and ethical regulations. It reflects upon the privacy concerns that may arise through the usage of generated and provided data during the implementation of the BEACONING project. It also comprises of a short description on how the research data will be openly accessible.

1.2 ROLE OF THIS DELIVERABLE IN THE PROJECT

This document is a more detailed elaboration of D1.7. This deliverable is to set the guidelines and governance for data management for the ownership of BEACONING data, the collection, storage, access to and use of the BEACONING ecosystem data, the treatment of IP, ethics and the measures of privacy, during and after the project. The immediate role of this deliverable is the governance of data regarding the BEACONING pilots and the participants. There will be additional ones regarding the data management plan in month 24 and 36.

1.3 APPROACH

The BEACONING consortium follows the Horizon2020 Programme guideline to ensure that our research data is approaching FAIR Data Management, as described in the DoA. The deliverable is developed in collaboration with pilot partners, as well as the main developers, in order to cover all aspects, in addition to achieving a mutual understanding.

1.4 STRUCTURE OF THE DOCUMENT

The deliverable is structured as follows: Chapter 1 gives an overview of the whole document. Chapter 2 explains how the privacy of the participants will be ensured. Chapter 3 outlines the progress, which has been done within the last 6 months regarding how to make the data FAIR, and Chapter 4 focuses on the ethical aspect of the pilot specific needs. Chapter 5 comprises of a short conclusion. The document should be understood as an extension to D1.7, i.e. there is no repetition of what was already described there.

2 DATA SECURITY

BEACONING pilots are producing and handling data that are sensitive and confidential from the end user point of view. Therefore, it is critical all partners ensure privacy of the user (de-personalize used data), as well as to save any collected data securely (with proper encryption) and controlling the access to the data through well-defined roles.

2.1 PRIVACY

End-user privacy is a key focus point for all partners, and the consortium is well aware of the legal restriction, as well as of the risks associated with the loss or misuse of sensitive personal data. All personal data for each user will be de-personalized at the very beginning when the student/teacher/parent logs into the system. Only authorised personnel have the right to trace, and that authorised person is either the one that generated the data (i.e. the user) or the parent or teacher. Any member of public can use the freedom of information act to ask for details, this will however be handled by the schools. The information that is made public is at the discretion of school authority. The identity of the users is located only by using specialized API functions that have been extensively validated and audited. All student records and information will be localised and is the responsibility of the school. The BEACONING platform will use surrogate IDs that cannot be used to identify students without access to the student personal data store component, which will be managed and under the control of the school.

The student information can be stored in the schools, under full control of local administrative resources.

2.2 SECURITY

In order to ensure the security and privacy of all parties involved in the BEACONING project, all data will be encrypted, while it is stored on servers or when it is being transported via the internet. We will use the encryption methods described in D1.7 and emphasize that also the data, which is not stored on BEACONING servers (e.g. the schools' servers), will be secured by these means.

3 FAIR DATA[1]

This chapter gives an overview about the progress regarding how we want to make our data FAIR.

3.1 ACCESSIBILITY OF DATA

The consortium agreed on making only totally anonymized data openly available as there are several partners, which legally are not allowed to provide other kind of data to third-parties, without the explicit permission of the person concerned. Even if the law states that data is open access, the quality of data should be controlled, i.e. no RAW data should be accessible to any party other than the owner. For third party, only filtered and condensed data, deemed appropriate by the owner, should be released. The provenance of data is still there but only enables very weak analysis for 3rd parties. Regions that do not allow to legally provide any kind of data to third-parties will be excluded from this measurement and will not provide any openly accessible data.

3.2 INCREASE THE RE-USING RATE OF DATA

Only data made openly available by the BEACONING project may be used by third-parties from the moment it is published.

The next subchapter describes the open data architecture plan as well as two problem scenarios and their related data flow plan, followed by an explanation of two architectures, which ensure data privacy at local level and anonymization for access from external parties (learning designers, game developers, and external researchers).

3.3 OPEN DATA ARCHITECTURE PLAN

The core system architecture uses a centralized data store that will be hosted in the European Union (to be decided if this will be on the premises of one of the partners, or using a cloud solution provider such as Microsoft Azure or Amazon AWS).

However, student personal identification data will be decoupled from the main data structures and will be made available as a separate Personal Data Store (PDS) component with a specialized API layer that implements access control rules and performs extensive auditing of all access requests. Personal data in the central repository will be stored using de-personalized identity tokens that cannot be traced back to a real person without access to the PDS.

The architecture allows personal data to be split across multiple PDS components, each stored in a different location (country or on the premises of the educational institution) so that it is separately administered and controlled.

Finally, administrative roles in the system will be created using separation of duties principles, which allows implementing scenarios where a single administrator does not have access to both personal identification data from personal data stores and the actual data contents stored in the main data store. This will prevent malicious administrators from circumventing regular access interfering with access auditing policies.

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Using standardized API interfaces for personal data access also allows existing systems that already track student data (e.g. LMS) to be extended in the future and act as a PDS themselves.

The table below (Table 1) presents a detail of multiple deployment scenarios using this approach. These can be mixed depending on each implementing entity requirements.

Table 1: data storage

No	Description	Advantages	Risks & Challenges
1	Personal data is stored using the same central infrastructure as the rest of the system, in an EU datacentre.	Simplified and reduced cost rollout to countries and entities that adhere to common EU privacy laws. Reduced infrastructure and personnel required from the implementing entity side. Specialized team that handles system security and maintenance.	The administration team should be correctly vetted, with separation of duties and access to private data on a need to know basis only.
2	Personal data is stored separately, in the jurisdiction that matches the end-user privacy requirements.	Provides compliance with specialized privacy requirements of jurisdictions. Centralized management provides opportunities for cost reductions and streamlined maintenance.	Additional infrastructure required for each deployment.
3	Personal data is stored separately, under sole control of the implementing entity.	Personal data management is under the control of a separate entity with full control over access rules and auditing practices.	Increased maintenance costs on the implementing entity side for infrastructure, personnel training and ongoing maintenance.
4	The entire system is deployed using the infrastructure that is controlled by the implementing entity.	Dedicated control that is fully managed by the implementing entity, according to their internal privacy rules or jurisdiction legislation.	High cost of implementation and maintenance. Specialized personnel required to handle all maintenance activities, as well as procedures for app deployment onto end-user devices (MDM or ad-hoc).

3.3.1 Stakeholders

In this Open Data Architecture Plan, seven stakeholders have to be considered for the collected data (see also Figure 1). Namely,

1. Students
2. Teachers
3. School administration
4. Learning designer
5. Game developer, and
6. External researcher
7. Parents

The different stakeholders have various levels of authorization and permission of access to different data.

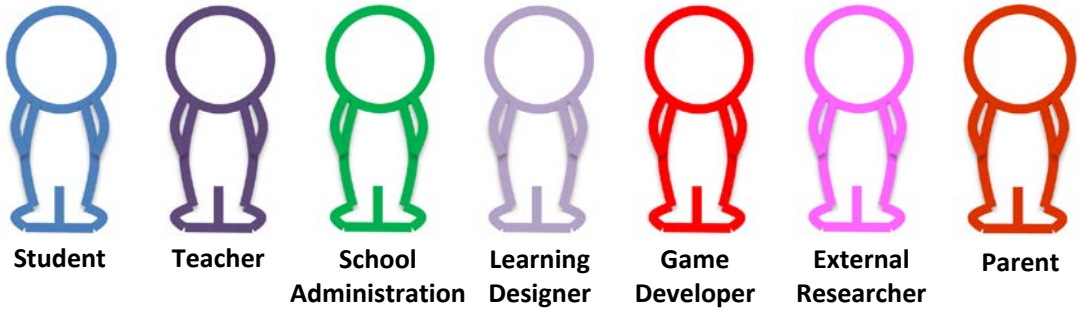

Figure 1. Stakeholders

3.3.2 Scenario-1: Teacher Sets Up Gamified Lesson Plan

This first scenario shows what happens, when the teacher creates a gamified lesson plan, Figure 2. To set this plan up for the pupils, the teacher would access the BEACONING ecosystem and find a suitable gamified lesson. Following this selection, the ecosystem will provide an authoring interface where the teacher would be able to set up the class size, individual student identity, learning contents, privacy and security aspects according to the pilot specific needs, etc. in order to set the options for the class.

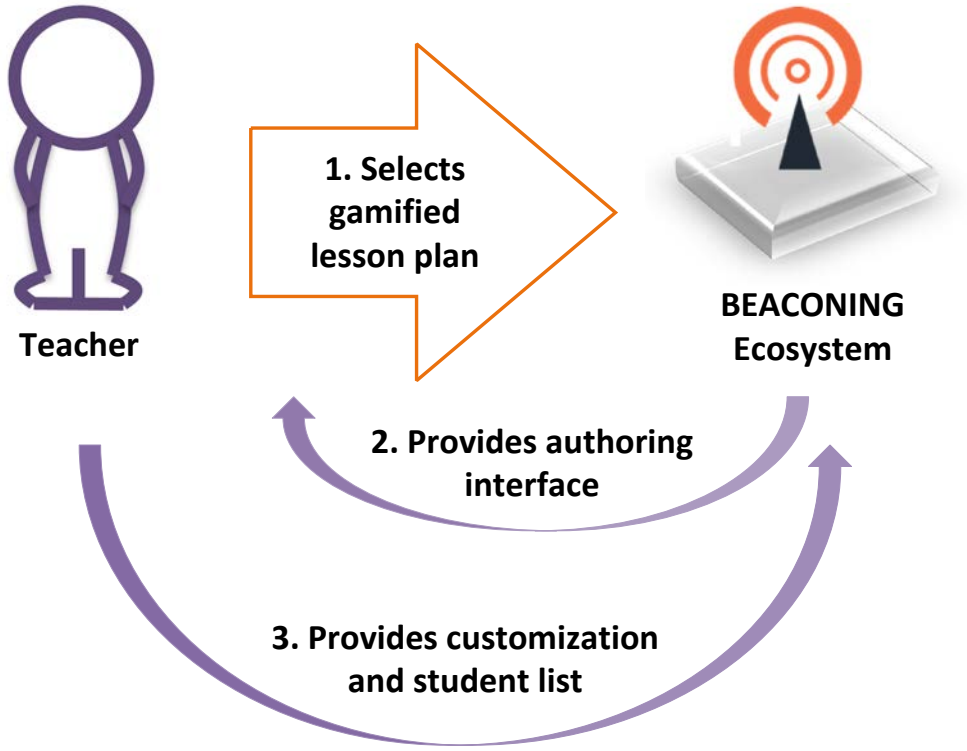

Figure 2. Scenario 1

Open Data Challenges:

The BEACONING platform must query game metadata to find out which customizations are available. Depending on how this metadata is stored and processed, there is a risk that customization queries lead to statistical disclosure of a group or particular person's disabilities.

The BEACONING consortium is aware of these possible risks, and therefore it will carry out a worst-case analysis identifying possible misuse. Based upon this analysis, we will develop actions for reducing the risks as well as decide upon the data storage and processing that gives the highest privacy protection, still allowing customisation

3.3.3 Data Flow Sequence- Scenario-1, Case- 1: BEACONING Integrated in The LMS

As it can be seen in Figure 3, whenever the teacher selects a gamified lesson path (GLP), the BEACONING ecosystem sends an enquiry to the BEACONING database to see what possible customization options are available at that moment. The list of available customization options is provided to the ecosystem and then another query for the class location and also for the student information is sent to the pilot specific Learning Management System (LMS) or to the BEACONING LMS (in case the Pilot does not have a proper LMS to run with). It should be noted that BEACONING will work with and without LMS integration, and that, in case a school would like to integrate BEACONING and their LMS, the LMS needs to provide standardised ports and gateways. In the current example, the LMS then provides the school ID (see chapter 2) and other required information about the students for the teacher's interface to customize according to the pilot specific needs. If necessary, the game developer can provide additional customization options as per the pilot specific needs, based on the assessment provided by the learning analytics component. The data generated by using the platform will be saved in the BEACONING vault and would be made available or not according to the customization selected during the setup of the lesson plan. The student's identification will be de-personalized before saving the data into the database and later in the vault.

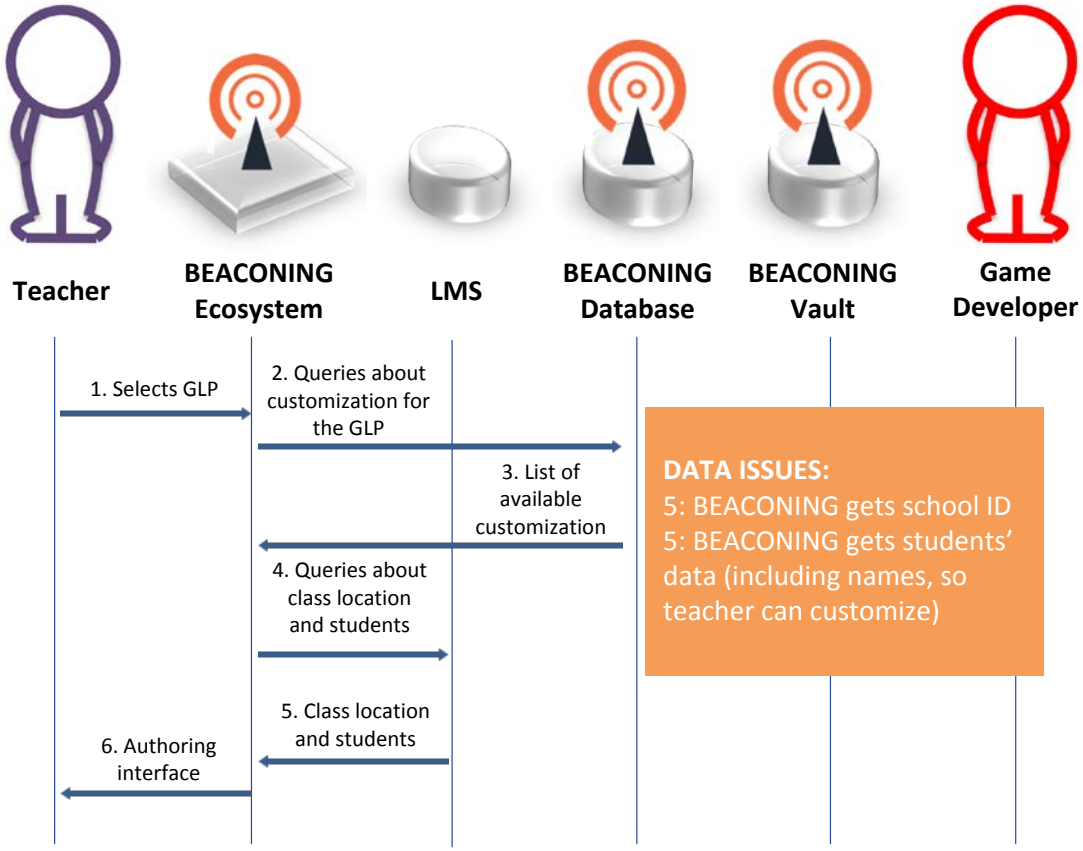

Figure 3. Data Flow of Scenario 1

3.3.4 Scenario-2: The Student Plays a Game

As the teacher sets up a lesson plan for the students and authorizes it, the students will be able to access a game with their identification credentials. Figure 4 shows that after logging in, when the student accesses the game from the BEACONING ecosystem, the game loads and the student plays it and learns through the process. The teacher has the option to customize the games

according to the student's specific needs. Consequently, if this option is selected by the teacher, the student is able to access the game with her/his own personal adapted elements. Throughout the process of play-learn, the generated data is sent to the BEACONING ecosystem and saved in the database. This data generation might include a numerous mix of numerical data, image files, audio clips, video clips, etc. according to the decisions the student has made in the place of dilemma or requirements. This data would be made FAIR (see Chapters 3.1 & 3.2) later on for re-use by the consortium or for external researchers under the Horizon 2020 Data Management Guidelines.

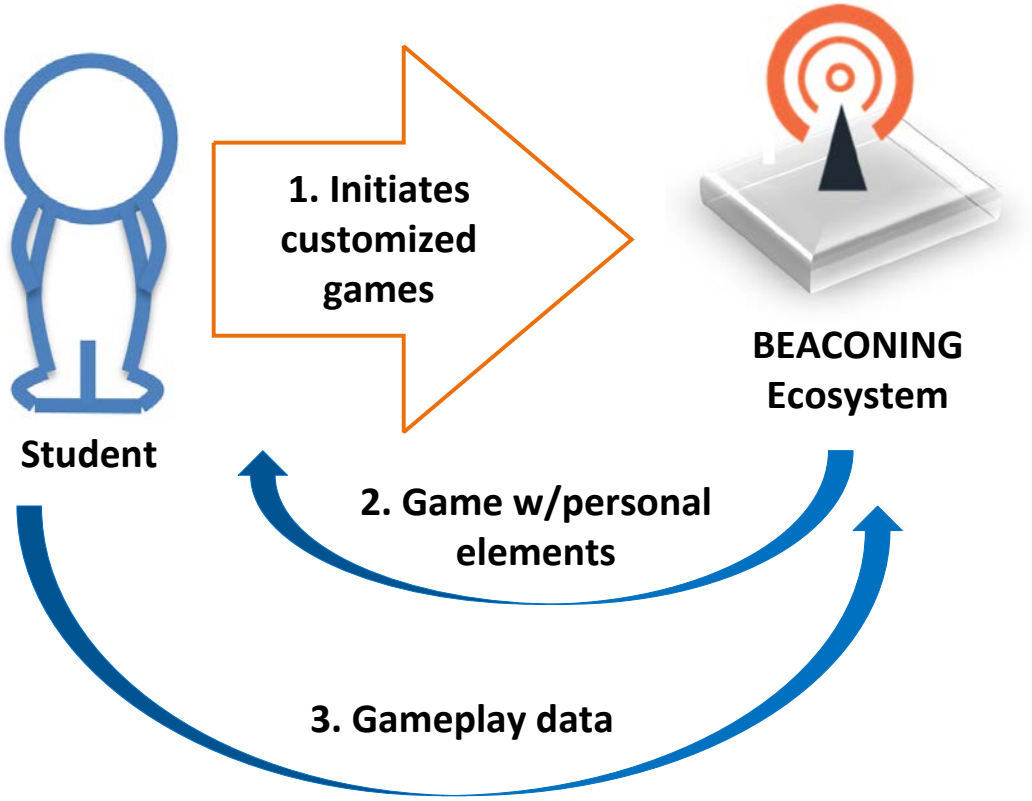

Figure 4. Scenario 2

Open data challenges:

1. The game may require the display of personal elements (name, photo, etc.);
2. Gameplay data may contain personal data (GPS location, choices taken in gameplay dilemmas, etc.). In such a case, pilot specific requirements will be followed (See section 4 for local requirements).

3.3.5 Data Flow Sequence- Scenario-2, Case- 1: Student Plays a Game

After the student launches the game, the application sends queries for gameplay configurations and for the content of the game to BEACONING database (see Figure 5). The relevant configuration and content are sent from the BEACONING database and the vault respectively. After that, the BEACONING ecosystem sends queries to the LMS for the student's personal information and those are provided to ecosystem and subsequently, the game interface is loaded for the student. During the gameplay, the student generates numerous data which are sent and saved in the BEACONING ecosystem's database for further use by the consortium members or by external researchers.

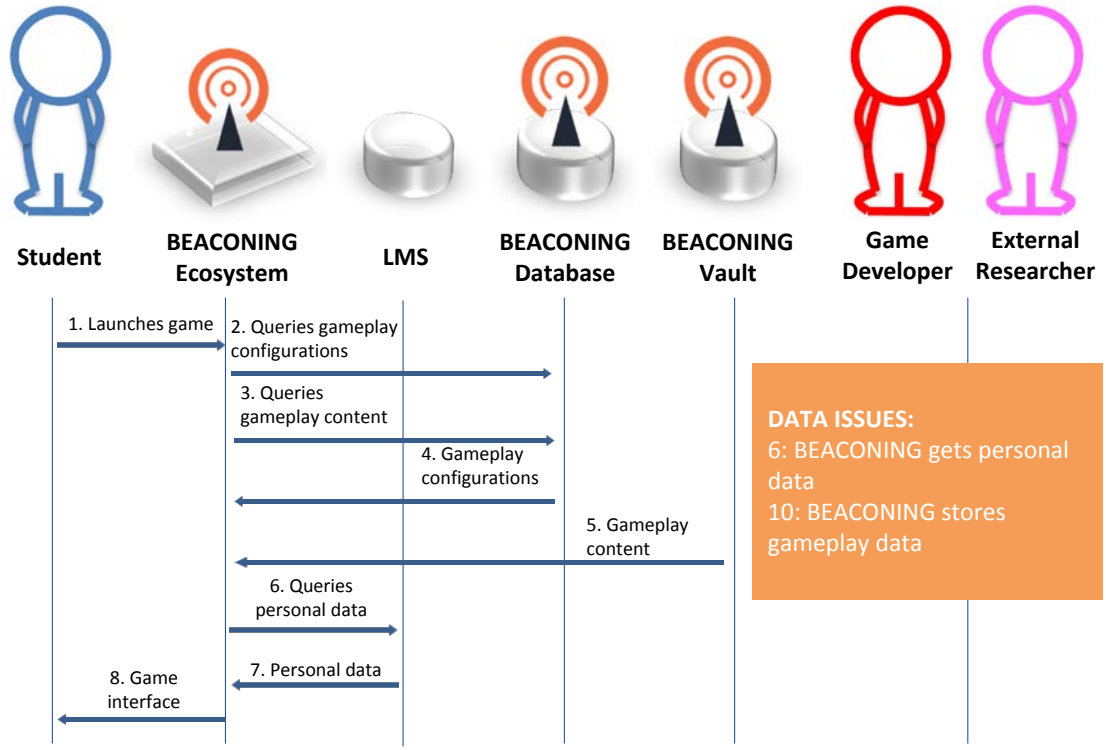

Figure 5. Data Flow of Scenario 2

3.3.6 BEACONING Open Data Architecture: School Side

As it can be seen in Figure 6, games or web pages used by students and teachers get personal data either directly from the LMS services or from a local BEACONING server managed by local authorities (school, district, region, country, etc.). Only the local server and local app know how to associate a BEACONING ID with a local LMS ID.

The International BEACONING databases never hold personal identification data.

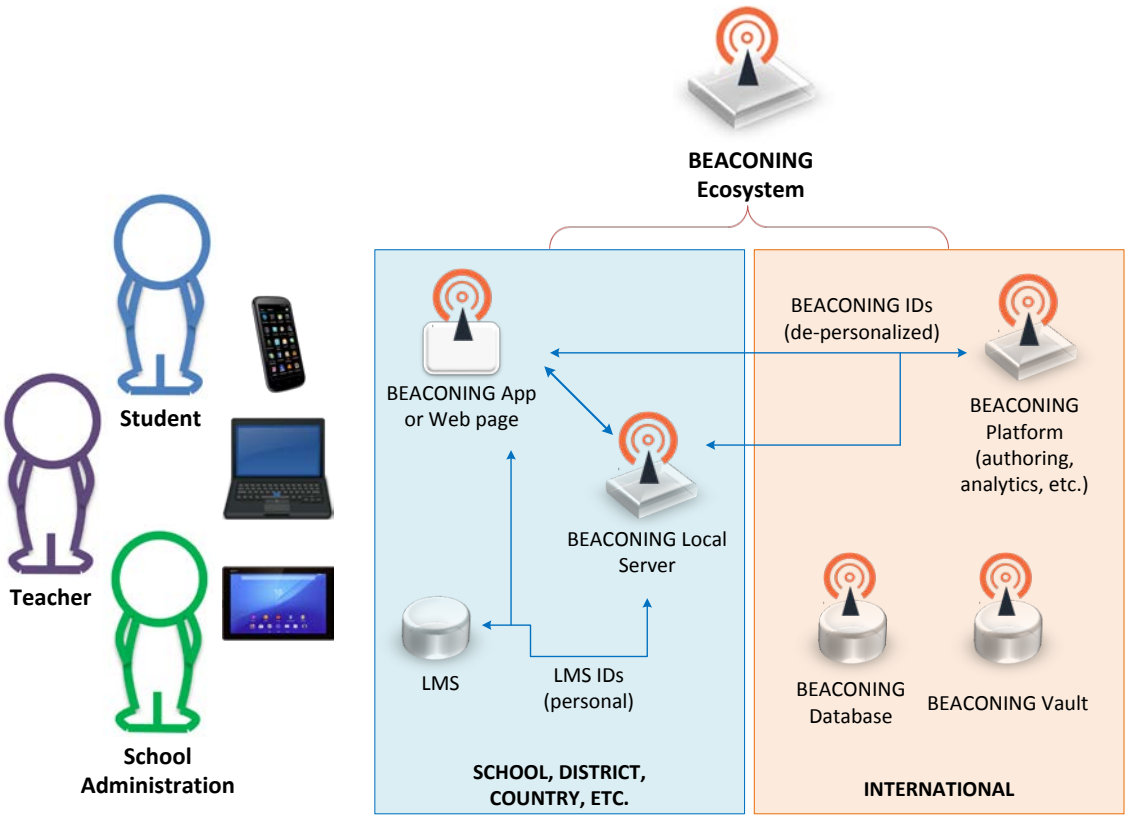

Figure 6. Data Architecture: School

3.3.7 BEACONING Open Data Architecture: International Side

Figure 7 shows that tools or web pages used by game developers, BEACONING participants, or external researchers, only work with BEACONING IDs, they have no access to local LMS IDs.

Special local needs (country/organization/purpose) can be catered for by a dedicated web services proxy.

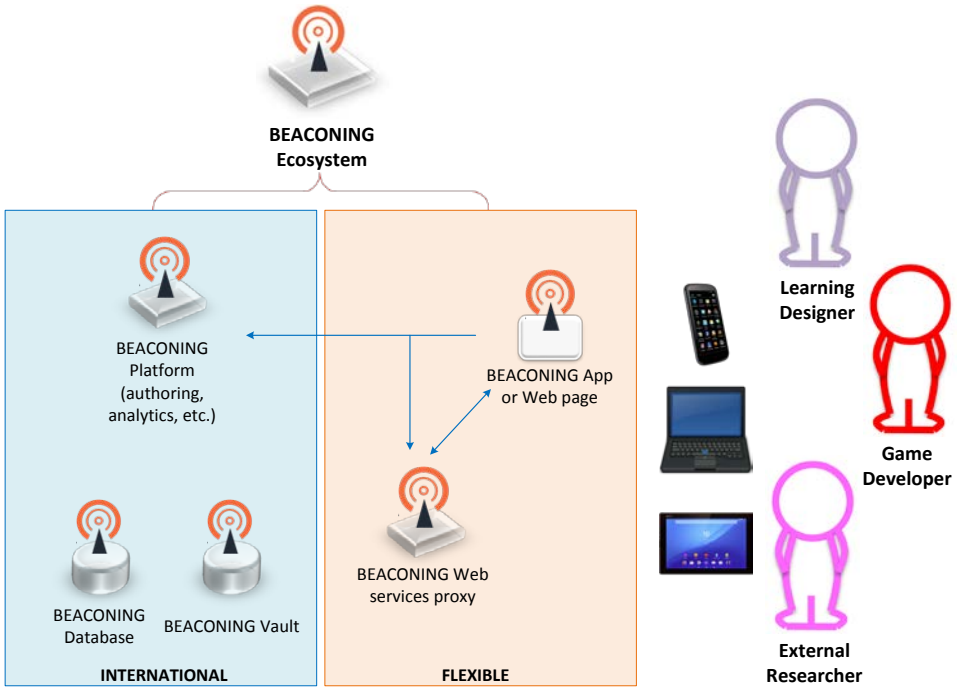

Figure 7. Data Architecture: International

4 ETHICS

In Deliverable 1.7 it was mentioned that all ethical and privacy regulations of the participating countries will be respected and will conform to the current legislation and regulations in the countries where the research takes place, to European legislation, as well as International conventions and declarations. To ensure this transparency, our consortium members have provided their pilot specific ethical aspects according to their country specific rules and regulations. In order to receive formal consent from the parents and/or guardians, the consortium has decided to use template of forms adapted to the national regulations (see D1.7), when there will be debrief session between the BEACONING platform personnel, the participants, their parents and/or guardians. The pilot specific needs are as follows:

1. **United Kingdom:** As decided in D1.7;

2. **Germany:** The German pilot is carried out in a higher education institution. Therefore, besides the consent of the student, no specific requirements are needed for this site. The consents declaration will be stored at BIBA. The students will decide whether they want to be a part of the research process or not;

3. **France:** The French pilot will abide by the ethical procedures as described in section 2.2.3 of Deliverable D1.7. In particular, in France, according to the legislation on data privacy, the CNIL states that there should be a declaration sent to the CNIL for collecting data on students. Exception: if the collection of the data concerns only administrative and pedagogical information related to objective data strictly necessary to the management of the schooling of the students that can be easily accessed by each individual student in case of a request.
   As the BEACONING platform can be installed in the schools, there should be also some specific forms that inform teachers and students that, the area they are entering now is covered by the usage of BEACONING.
   Regarding the mobile devices usage, which is not allowed in schools during the class, there should be mentioned in the BEACONING charter that for particular usage of the BEACONING in school, during precise timeframe, the usage of the mobile could be allowed;

4. **Greece:** In Greece, there should be an agreement on the software, in particular the BEACONING software shall be registered in a list of accredited software by the Ministry of Education to be used by the teachers.
   Also, as specific requirements, in case the BEACONING Software is using specific network infrastructures with specific ports that need to be opened in a firewall to enable access, there should be a form declaring the open accesses and Network;

5. **Turkey:** As decided in D1.7, they are using a consent forms to receive formal consent from the students' parents;

6. **Portugal:** Portuguese pilot will work with students in higher education. They do not have any special requirements, as the students will decide whether they want to be a part of the research process or not;

7. **Romania:** The pilot in Romania will be held with high school students. The students will be supervised by their teachers, so the activities will be performed through the schools. In order to involve the students younger than 16 years, BEACONING needs to receive a written consent from parents/guardians for processing personal information and to get involved in extra-school projects.
   For the students with special needs, the supervising teacher will obtain the acceptance from the parent/guardian of each student. It is the teacher's responsibility to obtain this acceptance in order to enrol a student in the project.
   Also, regarding the BEACONING installation in the schools, a specific form will be needed that informs teachers and students about where they are localized and what they are used for;

5 CONCLUSION

This deliverable pays attention to ensuring the advancement of the Open Data Architecture Plan. This document is an indicator that in terms of ethical and data management issues, all of the users of the BEACONING platform observe the same principles as the project advances into the next steps of realisation. In addition to the decisions taken in D1.7, the consortium is working to ensure data security, user privacy and the ethical aspects of the project.

This deliverable indicates advances in the work of privacy, security and ethical issues that has been carried out during the last 4 months. This work has been carried out according to the guidelines established in D1.7 taking the development of data security, ensuring privacy for the participants, and the chain of control, in order to control access to the stored data and at the same time, making the generated and collected data FAIR and abide by the Horizon 2020 guideline.

The next deliverable will be provided in month 24, after the small-scale piloting. However, an internal update will be made as soon as it is clear how we implement the learning analytics (see D4.6) in specific pilots and have more information on the data we have to collect. This will be ready before the small scale testing starts and the first prototypes are ready at month 18.