



X Modal
X Cultural
X Lingual
X Domain
X Site
Global OER Network

Grant Agreement Number:	761758
Project Acronym:	X5GON
Project title:	Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network
Project Date:	2017-09-01 to 2020-08-31
Project Duration:	36 months
Deliverable Title:	D3.6 – Copyright and privacy issues
Lead beneficiary:	Nantes
Type:	Report
Dissemination level:	Public
Due Date (in months):	24 (August 2019)
Date:	
Status (Draft/Final):	Final
Contact persons:	Colin de la Higuera

Revision

Date	Lead author(s)	Comments

Abstract

Project X5GON is to deal with copyright issues through the access to educational content and with privacy issues due (essentially) to the fact that learners leave traces which are used to build models but also to propose recommendations. More issues have been analyzed with care in the deliverable D9.1.

In this deliverable we discuss the research and technical choices which were made in order to comply with the copyright and privacy issues identified and scrutinized with the help of the ELT (Ethics and Legal team).

Summarizing, the copyright issues have been taken very seriously with a number of initiatives linked with the Connect-Service, the crawling, the database. The goal is to avoid any non licensed material in the system. Models are being build based on the collected material but we are careful that only material we have the relevant (copyright/privacy) consent is being used.

On the other hand the privacy question has been treated in full compliance with the GDPR: a user is anonymously tracked during his learning sessions.

1 Introduction

Education and artificial intelligence are the core of X5GON. In both cases the issues of copyright and education are of great importance. Together, they make policy makers nervous, users cautious and researchers very much aware of the risks and consequences of their actions.

Education is concerned in two ways. First, because educational resources (textbooks, articles, other pedagogic material) represent a sizable part of household and communities budgets. This is one of the reasons why, for more than 10 years, there has been a strong movement in favour of open educational practices, advocating the use of open educational resources (OER). Second, because education often concerns children or adolescents, in many ways a more vulnerable population, whose data has to be processed with great care.

Artificial intelligence is also very much concerned with these issues. New questions appear: the copyright issue for artifacts (for example course-ware) automatically built from machine learning is not yet settled. And machine learning makes ample use of data, which leads to many questions concerning privacy, handling the anonymization process, etc.

2 Task 3.4 in the X5gon project

In the original project we wrote:

Task 3.4 Privacy and user oriented guidelines (M12-30, Leader NAN)

Development of user oriented approaches allowing to check the pedagogical correctness of the implementations, and of discussions concerning the numerous privacy issues to be addressed. Discussions concerning these ethical matters will be promoted both within the project (and therefore in Liaison with the other WPs) and with representative groups who are engaged in studying these questions. The result of the task will be a report on multicultural, pedagogical, juridical, ethical (privacy) issues developed as a webpage, elaborated and enriched continuously during the lifespan of the project.

Task 3.4 can thus be seen as the one where researchers become aware of the privacy issues, identify (or get help to identify) the different risks, and include solutions addressing these questions in the development of the tools delivered/to be delivered in the project.

3 The main risks and issues as identified in Deliverable 9.1

We have met the Ethics and Legal team (ELT) for X5GON in several occasions. In particular the team was present at the launching seminar in Ljubljana in November 2017.



Deliverable 9.1 was prepared in August 2018 to address the concerns about these questions, from the legal point of view. From the report (D9.1) we have identified the key following issues:

1. Legal consequences of data processing if a commercial exploitation is envisaged (sec 1.3)
2. Ethical consequences of Machine Learning algorithms (sec 1.4)
3. Security of the conservation and transmission of Data (sec. 2.5)
4. Long term preservation and Archiving (sec. 2.4.3)
5. Reuse by other of collected data (sec. 2.4.2 and 2.4.4 B)
6. Use of copyrighted works (sec. 2.4.4 A)

We will explain how the X5GON researchers and developers tried to take into account these questions.

4 The questions raised and the responses through research

4.1 Legal consequences of data processing if a commercial exploitation is envisaged

During the review meeting of the project (Nov. 2018) it was suggested by the review panel that X5GON should offer opportunities to companies to work on the material produced so far and then to offer services to the public. Different discussions took place with companies interested in a collaboration. The proposed framework consisted in offering access to all the content we had collected, with the intermediate and rich models which we had built over this content. But no user data was to be shared.

In doing so, we believe that the copyright issues identified in D9.1 and the privacy questions raised were receiving an adequate answer.

A specific question is being studied now (July 2019): we are aiming to hold some hackathons in the autumn and an envisaged idea is to get the participants to make use of the (anonymized) user data to build new visualizations of learners, learning paths, learning recommendations. We are currently seeking an updated advice.

4.2 Ethical consequences of Machine Learning algorithms

Several members of X5GON are researchers in the area of machine learning, we are aware of these questions so we are conscious of the risks. In particular specialists are worried about the bias questions that can arise.

We believe this issue does arise here with a consequence for learning. When recommending educational resources, if we are based on collaborative filtering techniques, or on features which may relate quality with popularity of reputation, we could end up by only recommending resources from the better known organizations (most important universities, wealthiest institutions). As a consequence, great resources from under-valued institutions or countries would fail to be recommended. This would not only be of little interest to the community -one of the key ideas behind opening education is that everybody can play the role of consumer and of producer of knowledge!- but cause political problems: it would be very quickly apparent that some countries (or just one country) would be producing all the educational resources which get recommended and the whole open education alternative would be endangered.



This relates with the issue of *serendipity*. It is essential that resources which would be positively evaluated *per se* get a fair chance to be viewed by the users. Even if the author doesn't come from a prestigious University or from a leading country. This may mean studying the effects of bias during learning or introducing specific serendipitous scenarios.

4.3 Security of the conservation and transmission of Data

These questions have been taken very seriously. The data is stored in a database run by our expert partner Posta. And the different APIs have been protected.

4.4 Long term preservation and Archiving

This question is being dealt with. We are aware that the legacy issue is crucial. The alternative arrangements are being discussed regularly:

- Contacts with companies have been made, with the idea that they may propose products using the data and participate toward the maintenance of the database
- Partner Posta is very interested in remaining the hub for the material proposed by X5GON.
- Discussions with the Unesco take place on a regular basis: the Unesco sees X5GON as a great opportunity for them to support their own ambitious OER agenda.

4.5 Reuse by other of collected data

One side goal of X5GON is to produce data for research. The intention is that this data is open.

Concerning the content data As all the data is licensed with Creative Commons licenses consistent with Open Education (no ND clause), it is possible to share the transcriptions and translations of all learning resources which have been collected.

Concerning the user data The user data has been anonymized. As such, we believe that the user data can be shared. But at this point we are not sure.

4.6 Use of copyrighted works

One key obstacle for the development of open educational resources is the fact that most producers of educational material simply ignore that by not licensing their work, this one is not reusable by others. Through X5GON by scrupulously following the rules concerning the material we hope to convince teachers and other authors of educational resources to use correctly the open licenses available. The Creative Commons CC-BY with or without the SA option is the easiest to use and the one which will make reuse simplest.

It is not in the missions of this project to verify the conformity of the different OER which are proposed on the web in the different partner OER libraries, collections and repositories. But it is our duty to check that each resource which we are transcribing and thus could possibly recommend, is correctly licensed. Which we do.



5 Dealing with these questions at management level

Both questions of privacy and of copyright were taken very seriously at all moments. This resulted in specific decisions being made:

- When launching strategies to bring onboard more websites, the importance of being able to get hold of the copyright has been essential. In the meta-data, the license is essential. It has been adopted as a policy that only licensed materials would be recommended by X5GON.
- The systems chosen to track users across sites have been kept as simple as possible and always withholding the identity of the user:
 1. A cookie is created when it doesn't exist.
 2. A user's choices (what page/resource she chooses to use **inside** the set of web-pages identified for X5GON) is kept for further treatment and specifically to build the models.
 3. The cookie remains active for 10 years.
- The partner websites are encouraged to warn users with a banner that their anonymized data is being collected.
- Solutions allowing a user who makes the request to be let out and not be used for further models have been found.

6 Conclusion

Although the questions of copyright and privacy remain a challenge and a lot of energy is spent not just in complying with the texts but in finding the better ways to reassure all stakeholders, the X5GON project participants remain fully aware of their responsibility and will continue searching for solutions which comply with the highest standards for copyright and privacy.

