



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

<b>Grant Agreement Number:</b>	761758
<b>Project Acronym:</b>	X5GON
<b>Project title:</b>	Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network
<b>Project Date:</b>	2017-09-01 to 2020-08-31
<b>Project Duration:</b>	36 months
<b>Deliverable Title:</b>	D5.1 – First report on piloting
<b>Lead beneficiary:</b>	UPV
<b>Type:</b>	Report
<b>Dissemination level:</b>	Public
<b>Due Date (in months):</b>	12 (August 2018)
<b>Date:</b>	
<b>Status (Draft/Final):</b>	Draft
<b>Contact persons:</b>	Álex Pérez, Joan Albert Silvestre and Alfons Juan

**Revision**

Date	Lead author(s)	Comments
25-Jul-2018	John Shawe-Taylor	Added some comments / suggestions
20-Aug-2018	Àlex Pérez and Alfons Juan	Final review after John's

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>VideoLectures.Net</b>	<b>5</b>
2.1	Initial plan . . . . .	5
2.2	Y1 update . . . . .	5
<b>3</b>	<b>poliMedia</b>	<b>6</b>
3.1	Initial plan . . . . .	6
3.2	Y1 update . . . . .	6
<b>4</b>	<b>virtUOS</b>	<b>7</b>
4.1	Initial plan . . . . .	7
4.2	Y1 update . . . . .	7
<b>5</b>	<b>Other pilots</b>	<b>8</b>
5.1	Y1 pilots . . . . .	8
<b>6</b>	<b>X5gon-TTP</b>	<b>8</b>
6.1	MLLP-TTP . . . . .	8
6.2	Y1 developments . . . . .	9
6.2.1	X5gon-TTP v1.0 . . . . .	9
6.2.2	OER collections from official pilots . . . . .	10
6.3	First results on transcription quality . . . . .	12
<b>7</b>	<b>Conclusions</b>	<b>13</b>
<b>A</b>	<b>virtUOS: Y1 proposal update</b>	<b>14</b>
A.1	Current state . . . . .	14
A.1.1	Privacy . . . . .	15
A.2	Pilot Idea . . . . .	15
<b>B</b>	<b>X5gon TTP: Detailed description</b>	<b>17</b>
B.1	Main features . . . . .	17
B.2	Main components . . . . .	18
B.2.1	Database . . . . .	18
B.2.2	Web Service . . . . .	19
B.2.3	Ingest Service . . . . .	20
B.2.4	Subtitle Editor . . . . .	20
B.3	Change log . . . . .	21
<b>C</b>	<b>X5gon TTP: Transcription and Translation Systems</b>	<b>22</b>
C.1	German ASR System . . . . .	23
C.1.1	Compilation of Resources . . . . .	23
C.1.2	System Description . . . . .	24
C.2	Slovenian ASR System . . . . .	25



## List of Figures

1	X5gon-TTP dashboard. . . . .	10
2	Three X5gon-TTP use cases . . . . .	11
3	Multilingual editor of transcriptions (top) and translations (bottom). . . . .	12
4	Use case: X5gon user accesses a new pilot's OER . . . . .	13
5	Progress for all languages in ASR given in terms of WER (the lower the better) . . . .	14
6	X5gon TTP Subtitle Editor . . . . .	20
7	ASR systems progress during transLectures . . . . .	22



## List of Tables

1	Basic statistics on VideoLectures.Net videos by their most common languages. . . . .	5
2	Basic statistics on poliMedias by their most common languages. . . . .	6
3	Main MLLP-TTP users from March 2014 to June 2018. . . . .	9
4	Pilot OER transcription and translation stats . . . . .	10
5	Details on ASR test sets. . . . .	13
6	German speech data . . . . .	23
7	German linguistic resources . . . . .	24
8	WER of the Slovenian ASR system. . . . .	26



## Abstract

The main objective of WP5 is to pilot successive versions of project components developed in WP1–WP4. D5.1 is to report the work done in Task 5.1 until August 2018 (M12). It is mainly focused on the scenarios posed by the three official pilots devised in the project, VideoLectures.Net, poliMedia and virtUOS. This report also includes a description of the first infrastructure and set of tools developed for piloting, X5gon-TTP.

## 1 Introduction

Work Package 5 (WP5) was planned to carry out a series of piloting studies to link the automatic analysis undertaken on the data with the experiences of groups of learners. Generally speaking, these studies are expected to contribute significantly in revealing the factors that hold user engagement, make learning enjoyable and rewarding, and help in developing a rounded understanding of different disciplines.

The main objective of WP5 is to pilot successive versions of project components developed in WP1–WP4, namely:

- **X5oerfeed:** *The project will deploy a technological pipeline for content understanding that is based on wikifier, dmoz and other services developed by JSI, and video transcription and translation services developed by UPV.*
- **X5analytics:** *The project will track data of users and their progress and use that to drive an analytics engine driven by state-of-the-art machine learning that can improve recommendations through better understanding of users, their progress and goals, and hence their match with knowledge resources of all types.*
- **X5recommend:** *Cross-site and cross-lingual recommendation.*

The above does not include evaluating the use of the X5gon platform in practice, or the coordination of a network of European OER repositories, which are main objectives for WP6 and WP7, respectively. Instead, WP5 can be seen as a primary source of feedback for WP1–WP4, and also for the non-technical WP6 and WP7.

WP5 runs from March 2018 (M6) to the end of the project (M36, August 2020), and consists of two main tasks:

1. **Task 5.1 Piloting on individual components (M6-M24, Leader UPV).**  
*Small in-house groups will be established to assess successive versions of project components developed in work packages 1 to 4. JSI will pilot individual components from WP2 and WP4, whereas UPV and Nantes will focus on WP3 components.*
2. **Task 5.2 Piloting on integrated components (M13-M36, Leader UCL).**  
*UCL, JSI, UPV, UOS and Nantes will pilot integrated components in the social network. In M13-M24, they will start piloting advanced analytics and social context meetings both virtual and physical. In M25-M36, advanced cross-lingual and cross-modal features will be piloted for the social network to be prepared for different cultures.*

This deliverable, D5.1 – First report on piloting, is to report the work done in Task 5.1 until August 2018 (M12). It is mainly focused on the scenarios posed by the three official pilots devised in the project, VideoLectures.Net, poliMedia and virtUOS, which are considered in Sections 2, 3 and 4, respectively. For each of them, the initial plan for the pilot is first summarized, and then a Year 1 (Y1) update is provided including relevant decisions and actions taken over Y1. These three Sections



are followed by a Section on additional pilots under study (Section 5). Apart from these Sections on pilots, this report also includes a description of the first infrastructure and set of tools developed for piloting, X5gon-TTP, which is provided in Section 6. The report ends with some conclusions and a brief account of ongoing and planned work.

## 2 VideoLectures.Net

### 2.1 Initial plan

VideoLectures.Net is a free and open access educational video lectures repository. The lectures are given by distinguished scholars and scientists at the most important and prominent events like conferences, summer schools, workshops and science promotional events from many fields of Science. The portal is aimed at promoting science, exchanging ideas and fostering knowledge sharing by providing high quality didactic contents not only to the scientific community but also to the general public. All lectures, accompanying documents, information and links are systematically selected and classified through the editorial process taking into account also users' comments. VideoLectures.Net serves more than 10,000 unique visitors per day, from all over the world.

Table 1 shows basic statistics on VideoLectures.Net lectures, by their most common languages.

Language	Videos
English	22335
Slovene	3540
Other	279
Totals	26154

Table 1: Basic statistics on VideoLectures.Net videos by their most common languages.

The initial plan of VideoLectures.Net is to maintain its position as one of the leading open scientific media and video services in the world by offering high quality multilingual service. As in X5gon, VideoLectures.Net has been included as a case study in the European projects transLectures (2011–2014) and traMOOC (2015–2018), where the use of advanced language technologies was studied to produce multilingual subtitles in a cost-effective way. VideoLectures.Net expects X5gon solutions to advance further in the application of language technologies to its content. More generally, X5gon is expected to address the language, personalisation and content convergence barriers of the diverse communities developing around VideoLectures.Net.

### 2.2 Y1 update

In the first year Videolectures.Net integrated the X5gon user activity tracker on pages containing the video content. When a user accesses such content the tracker sends user activity information to the X5gon platform where it is validated (e.g. checked if the sent data is in the correct format) and stored in the database for further analysis. We have acquired approximately **2.5 million records of user activity data** since its integration. The user activity tracker is described in D2.1 - Requirements and Architecture Report.

Furthermore, VideoLectures.Net is being prepared to integrate the X5gon recommendation list to the main portal. The list will contain OER material that are similar in content to the material the user is currently viewing. For more information on how the recommendations are produced, see D4.3 - Early Prototype of Recommendation Engine.



### 3 poliMedia

#### 3.1 Initial plan

UPV is among the top-ranked Spanish universities in Engineering and also a pioneering Spanish institution in the production of video lectures to support Education. A main effort in this regard is poliMedia [1]. Launched in 2007, poliMedia is a UPV service for the creation, storage, management and dissemination of video lectures in a professional setting. It was designed for UPV lecturers to record video lectures supplementing conventional face-to-face classes. After ten years since its creation, poliMedia now includes more than 28000 videos of around 10 minutes in length, recorded in dedicated studios by more than 1400 lecturers, mainly in Spanish, English and Catalan. Table 2 shows basic statistics on these videos, usually referred to as poliMedias, by their most common languages.

Language	Videos	Hours	Lecturers
Spanish	24549	5057	913
English	2543	374	341
Catalan	952	116	143
Other	438	54	52
Totals	28482	5601	1449

Table 2: Basic statistics on poliMedias by their most common languages.

Like VideoLectures.Net, poliMedia and its associated MOOC platform, UPV[X], have been already included as case studies in previous European projects studying how to produce multilingual subtitles cost-effectively by advanced language technologies. In particular, transLectures [2] included poliMedia as a case study from 2011 to 2014. Later, from 2014 to 2016, EMMA [3] included several UPV[X] MOOCs, which in turn are essentially based on poliMedias, to develop a pioneering multilingual European MOOC platform. As a result from these projects, the live poliMedia repository is being automatically transcribed and translated into many EU languages by state-of-the-art ASR and MT systems developed at the UPV. The resulting multilingual subtitles are generally accepted as accurate enough for educational purposes, though manual revision is also possible on the basis of a *Transcription and Translation Platform* which is indeed the seed of the X5gon-TTP platform described in Section 6. Besides lecturers' interest in having "perfect" subtitles for their own videos, manual revision is strengthened by the UPV's Docència en Xarxa (DeX) incentive plan, run every academic course to support the preparation of reusable educational resources. This is particularly true for poliMedias included in UPV[X] MOOCs, which are manually reviewed in an very efficient way by first reviewing automatic transcriptions, and then reviewing automatic translations produced from already reviewed transcriptions. A detailed analysis on this procedure can be found in [4]. In it, it is shown that a lecturer can save between 30% and 70% of the time devoted to review transcriptions, and between 25% and 75% of the translation review time, with respect to performing these tasks from scratch.

From poliMedia's perspective, X5gon is a natural continuation or, better said, an extension to the work done in transLectures and EMMA. In other words, poliMedia expects X5gon solutions to advance further in the application of language technologies to its content, and thus place it in a better position to fully engage with the X5gon network.

#### 3.2 Y1 update

During this first six months poliMedia has been focusing on two main tasks:





On the one hand, to identify and process the OER materials that will be shared with the project. The OER contents have been transcribed and translated, and the generated transcriptions and translations (accompanied by videos metadata) have been shared with all the partners in order to provide data to other models and tools being developed in the different work packages. More details on transcriptions and translations quality assessment for poliMedia will be found in 6.3.

On the other hand, to tackle the installation of the X5gon snippet into the official poliMedia repository. This issue has led to many technical, but mainly legal discussions regarding the accomplishment of the recent GDPR regulation changes. Since the snippet will be collecting and sending anonymized user data, poliMedia has been very concerned about sharing data that might violate any privacy policy or legal agreements of its students. Nonetheless, after checking that all the data being collected by the snippet is anonymized enough, poliMedia has finally agreed to integrate the snippet into its official repository. Some developments were needed to only integrate the snippet in the lectures having an open license (OER). Finally, in July 2018 poliMedia finished the installation of the snippet into all the open videos of the official repository.

## 4 virtUOS

### 4.1 Initial plan

The Center for Digital Teaching, Campus Management and University Didactics (virtUOS) is a central research and service facility of the University of Osnabrück with the task of researching and developing innovative technologies in the field of teaching and university organization and in cooperation with other institutions of the university process-based. In line with the university's goals, virtUOS has developed a differentiated research profile in recent years. The virtUOS media repository consists of 560,000 documents, 6,475 recordings, related into 470 series, 6 MOOCs and 19 OER courses.

The competitive advantages that X5gon solutions would bring to virtUOS over other monolingual or interlinking with well-resourced OER repositories or interlinking their own materials, would be an introduction and implementation of the X5gon services, especially the cross-lingual and learning analytics, over the Opencast (Open-Source Lecture Capture and Video Management for Education) and thus multiplying the impact creation as software contributor in the community and cross-link the University-wide Lecture Recording Service at the University of Osnabrück.

As a creator of OER production tools and a producer of OER, increasing the flexibility of OER uses is a key concern for virtUOS. Lecture Recordings produced with the Opencast tool and disseminated via Learning Management Systems and other public channels have to be presented in a more individualized and thus attractive format to develop their full potential. As a consequence of increased attractiveness, more lecturers and institutions will agree to have their lectures recorded and published as OER.

For the Osnabrück University similarly as to UPV, the X5gon project would allow to continue its innovation activities with MOOCs, Opencast Matterhorn penetration and research into pedagogical approaches and exploit its foreground from previous projects.

### 4.2 Y1 update

Following many technical discussions on X5gon progress and how it can better attend the needs of virtUOS, UOS has updated its pilot proposal so as to be more focused on a *OER Recommender for Lecturers*. The X5gon services will provide recommendations of additional OER resources to the teacher. The teacher can then decide whether to use these resources before making them available to students. This two-step procedure is necessary because teachers will most likely not accept a service that recommends additional material to the students without the teacher's consent.



A more detailed description of the updated proposal is given in Appendix A.

## 5 Other pilots

Although the work plan is focused on three official pilots described in previous Sections, other X5gon partners with OER resources (as is the case of UCL and Nantes) will also help testing X5gon services and tools. In addition to project partners, external OER repositories will be welcome to contribute on piloting of components and services once the project reaches a more mature development state.

### 5.1 Y1 pilots

Université de Nantes went through the installation of the X5gon snippet into their OER repository earlier this year. Their experience and the issues faced during the set up of the snippet helped to get the X5gon servers infrastructure and communication protocols up and running for the upcoming pilots.

University College London is helping in the integration of Moodle sites into the X5gon platform.

## 6 X5gon-TTP

### 6.1 MLLP-TTP

The UPV's Machine Learning and Language Processing (MLLP) research group started the development of a Transcription and Translation Platform (TTP) during the execution of the European project transLectures (2011–2014). In what follows, such a platform is referred to as the MLLP-TTP [5]. In transLectures, the MLLP-TTP was a clear need, not only for reviewing automatic transcriptions and translations, but also for trying innovative ideas on human-computer interaction. The main motivation to try these ideas was a strong belief that accurate enough results can hardly be obtained with fully automatic methods alone, and thus reaching higher accuracy levels inevitably requires some sort of intelligent interaction with users. Although this belief was, and still is, true when quality is critical, it is also true that the quality of automatic transcriptions and translations is rapidly improving in recent years, and thus widening the range of applications where automatic methods alone can deliver results of practical use. This is, briefly speaking, what we observed in transLectures for VideoLectures.Net and, especially, for poliMedia [6].

The MLLP-TTP has been extensively used after transLectures by many projects, (educational) organisations and casual users willing to try and use it for automatic transcription and translation of videos (text and web) on the basis of ever improving MLLP's ASR and MT systems. This includes poliTrans, a "stable" version of TTP run by the UPV's *Information and Communication Systems Area* mainly to produce multilingual subtitles for the live poliMedia repository [7]. It also includes the European project EMMA (2014–2016), its partners and multilingual MOOC platform [3]. Apart from poliTrans and EMMA, the MLLP-TTP has been used by 44 organisations from 14 countries from March 2014 to June 2018. In total, more than 5000 videos accounting for more than 1500 hours in 9 EU languages were processed by the platform (see Table 6.1).

As a result of an extensive use after transLectures, the MLLP-TTP and the ASR/MT systems behind it have undergone many improvements and extensions; and the MLLP has gained considerable experience on producing and managing multilingual subtitles for video lectures in many EU languages. As planned, the current MLLP-TTP and ASR/MT systems behind it are being ported and adapted to X5gon needs, particularly for the official pilots. This is described in what follows.

Table 3: Main MLLP-TTP users from March 2014 to June 2018.

User	Country	Videos	Hours	Language
Universidad Carlos III de Madrid	Spain	2123	225	ES,EN
Jožef Stefan Institute	Slovenia	2038	1105	EN,SL,FR,ES,IT
ViJump	France	289	33	FR,EN
Hasso-Plattner Institute	Germany	118	30	EN,DE
Open Universiteit Nederland	Netherlands	112	8	EN
Translation Centre (CdT)	EU (Lux.)	110	39	EN,FR,DE,ES,IT
Universidade de Lisboa	Portugal	69	9	PT
ATiT	Belgium	42	20	EN,NL
UNED	Spain	38	32	ES,EN
TEDx UPV	Spain	28	6	ES
Altabox S.L.	Spain	14	1	ES
Coaching-Tecnologico.com	Spain	13	2	CA,ES
Vrije Universiteit Amsterdam	Netherlands	10	1	NL
31 other organisations	11	85	28	CA,EN,DE,FR,DE,ES,IT,PT
Totals: 44 organisations	14	5089	1539	9

## 6.2 Y1 developments

A large part of the work done in Task 5.1 during the second half of Y1 consisted in adapting the MLLP-TTP and ASR/MT systems behind it to X5gon official pilots. In Section 6.2.1, we first discuss our port of the MLLP-TTP infrastructure and set of tools for X5gon, named X5gon-TTP v1.0. Then, in Section 6.2.2, we describe the first OER collections from the official pilots uploaded to X5gon-TTP for piloting. Finally, in Section 6.3, we provide first results on the transcription quality we can expect from UPV's ASR systems adapted to the official pilots.

### 6.2.1 X5gon-TTP v1.0

The first version of X5gon-TTP is based on version 3.7 of the MLLP's *transLectures-UPV Platform (TLP)* which, apart from a few website-specific adjustments, is what constitutes the current MLLP-TTP. X5gon-TTP can be accessed at <https://ttp.mllp.upv.es/X5gon>. After authentication, the X5gon-TTP dashboard shows up with a control panel on the left and a video (object) listing in the main area, as in the example shown in Figure 1.

The X5gon-TTP dashboard is just an entry point to a high-level interface for interactive users. The main components supporting X5gon-TTP are: a PHP / HTML5 media *player / editor* for reviewing subtitles, a *web service* to integrate X5gon-TTP services into external repositories, an *ingest service* including core functions to manage multilingual media and a *database* to support the web and ingest services. The way in which these components are typically used is illustrated by the three use cases depicted in Fig. 2.

The first use case given in Fig. 2 (left) consists in adding a new recording. In this case, the OER repository uploads the new media to the X5gon-TTP server using the *ingest* interface of the web service. Then, the ingest service runs the appropriate ASR/MT systems and stores both the new media and its subtitles in the database. In the second use case shown (center), video lecture subtitles are manually reviewed by using the *player / editor*. All revisions, either modifications or simple validations, are added to the database through the *mod* interface of the web service. Also,

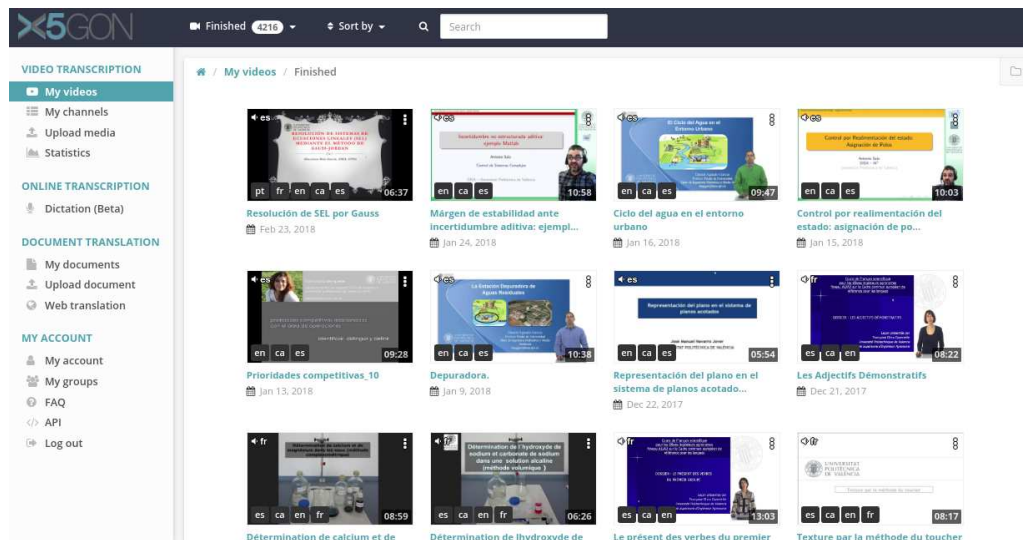


Figure 1: X5gon-TTP dashboard.

they are processed by the ingest service to improve ASR/MT models and update all non-reviewed subtitles. The third and final use case (right) is on retrieving multilingual subtitles. In this case, the OER repository retrieves the relevant multilingual subtitles from the database via the *subs* interface of the web service.

At this point, the importance of the X5gon-TTP player / editor is worth noting. It is for sure the software piece that has evolved the most in the last few years, from transLectures to X5gon. Indeed, it has been steadily improved over these years from the feedback of many users (see Section 6.1). The current X5gon-TTP player / editor works with different devices, layouts, editing modes and subtitles. For instance, Fig. 3 shows two basic examples of use. The example shown on top is as a transcription editor. Not only transcriptions can be edited in the boxes at the right, but also segments (including their timings) can be edited, inserted, deleted or merged either at the bottom left (above the signal) or right (at the edit boxes). The example of use shown at bottom is similar, though extended to multilingual subtitles, that is, transcriptions and translations.

### 6.2.2 OER collections from official pilots

In order to prepare the platform for piloting purposes, we ingested all the OER videos VideoLectures.Net, poliMedia and virtUOS have provided for initial piloting evaluations. Automatic transcriptions and translations have been generated for all this content and are accessible for every partner. Each one of the pilots has their own credentials to access their contents, where they can not only view the ingested contents, but also view and post-edit the transcriptions and translations generated by the ASR and MT systems. The current amount of transcribed and translated content for each pilot is described in detail in Table 4.

Table 4: Pilot OER transcription and translation stats

	en		es		de		sl	
	vids	hours	vids	hours	vids	hours	vids	hours
VL.NET	1629	819h	-	-	-	-	391	283h
poliMedia	165	24h	3870	559h	4	0.5h	-	-
virtUOS	26	34h	-	-	250	226h	-	-



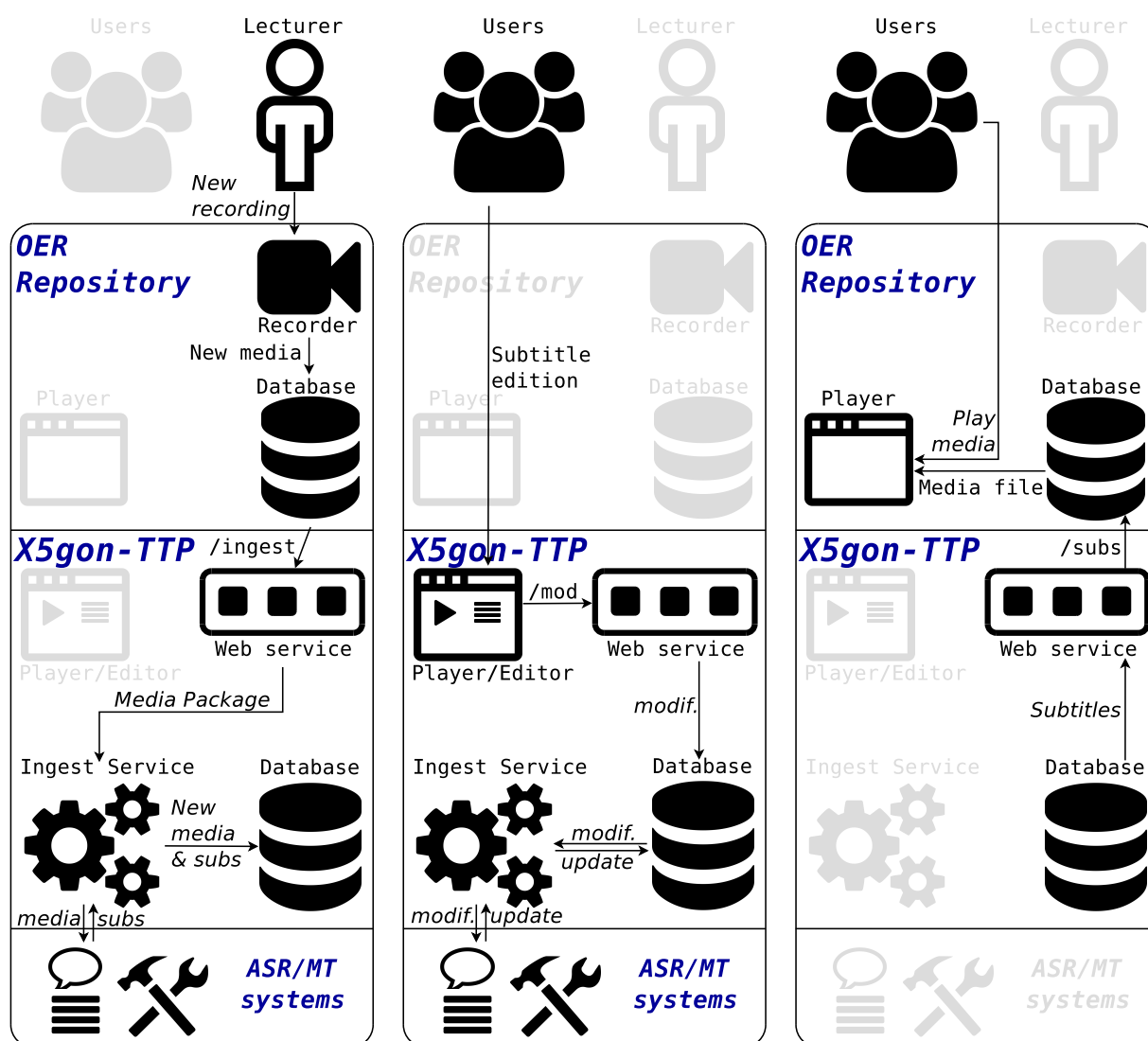


Figure 2: Three X5gon-TTP use cases: adding a new recording (left), reviewing subtitles (centre), and retrieving multilingual subtitles (right).



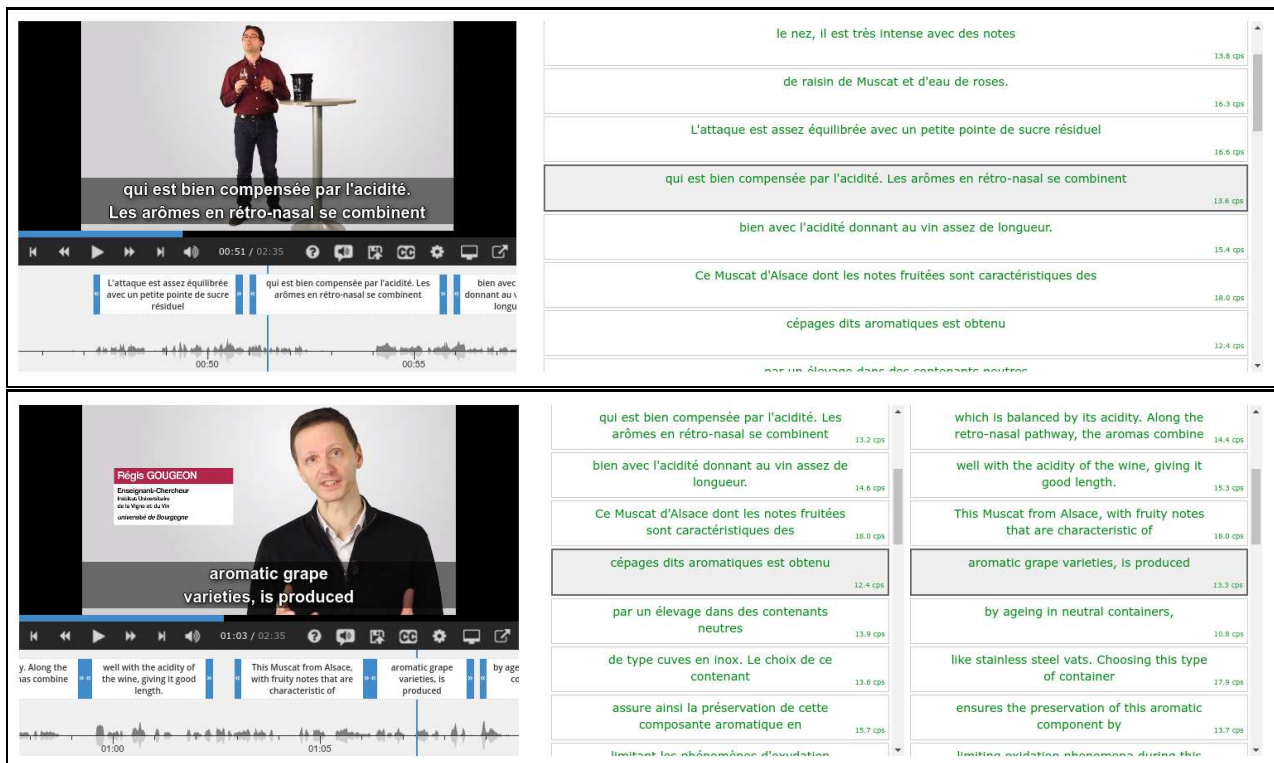


Figure 3: Multilingual editor of transcriptions (top) and translations (bottom).

In the case of poliMedia, the collection corresponds to all repository OER up to the time writing. In the case of virtUOS, the collection corresponds to the piloting videos shared by UOS up to July 2018.

As pilot sites finalize the integration of the X5gon snippet, the platform is being set up to automatically transcribe and translate the new OER contents accessed by users for which subtitles are not yet available. Figure 4 depicts the connection between the pilot site, the X5gon processing pipeline and X5gon TTP. When a student (or user) accesses an OER in any of the X5gon pilots, the X5gon snippet sends a query to the X5gon server with some data like the content URL, an anonymous user ID, etc. This data goes directly into the processing pipeline, which first queries the server database to check whether this content has already been previously processed (transcribed, translated, *wikified*, etc.). If it is a new OER, the server will send a query to the X5gon TTP platform in order to automatically transcribe and translate the OER contents. Once available, the generated subtitles will be automatically downloaded by the X5gon server and stored in the database.

### 6.3 First results on transcription quality

For evaluating the ASR systems we defined test sets for the different languages considered in X5gon consisting of representative (manually transcribed) videos from the official pilot sites. Test sets for VideoLectures.Net and poliMedia were already defined in the transLectures project. For VirtUOS we considered best to define two different test sets according to the two main type of media objects being recorded in the platform: one for the screencast recordings, and one for the classroom recordings (with lapel microphone). More details on these sets are given in Table 5.

The word error rates (WERs) for each system on the corresponding test sets are shown in Figure 5. The WER is the ratio, expressed as a percentage, of the number of basic word editing operations (sub-

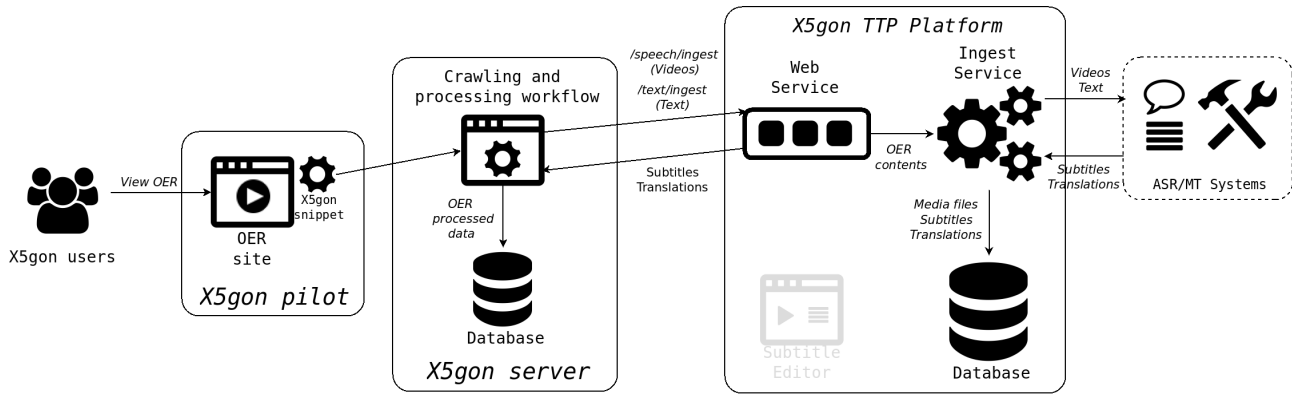


Figure 4: Use case: X5gon user accesses a new pilot's OER

Language	Set name	Videos	Hours
Spanish	poliMedia	23	3.0
English	VL.NET	4	3.4
Slovene	VL.NET	4	3.4
German	VirtUOS (screencast)	19	5.3
German	VirtUOS (classroom)	2	2.4

Table 5: Details on ASR test sets.

stitutions, deletions and insertions) required to convert the automatic transcription into the corrected transcription and the total number of words in the reference transcription.

During Y1 there has been an effort on building a first ASR system for German, especially adapted for the VirtUOS settings. It is worth mentioning that the VirtUOS (screencast) test set is only composed of 2 videos from 2 different speakers: one of them with a lot of reverberation (resulting in 61.3% WER), and the other one with better acoustic conditions, registering 23.8% WER. Word error rates are still not as good as those for English or Spanish, but there is still room for improvement, particularly as we enlarge the available in-domain training corpora. More details on the German ASR system can be found in Annex C.

The Slovene ASR system has also been improved during this period, both technically and in terms of available training data. More details on this work can be found in Annex C.

English and Spanish systems have not been modified during this period. Our effort to improve the ASR systems quality has been focused on *De* and *Sl*, which are currently achieving the poorest results.

## 7 Conclusions

During the second half of Y1 VideoLectures.Net and poliMedia have finally set up the snippet into their official repositories. The installation of the snippet into the official pilots was a key target since user and content data needs to be collected from the pilots in order to develop the different models and tools in WP1-4.

Additionally, a large part of the work done consisted in adapting the MLLP-TTP and ASR/MT systems behind it to X5gon official pilots. The X5gon-TTP platform was deployed in order a) to transcribe and translate all the content from the pilots and b) to evaluate the performance of the ASR/MT systems.

Future work in piloting will be focused on the evaluation of the X5gon components (analytics,

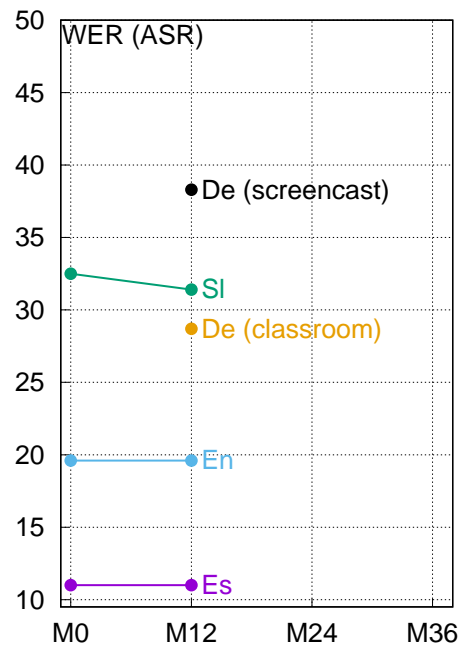


Figure 5: Progress for all languages in ASR given in terms of WER (the lower the better)

recommendation) in the pilot sites, in addition to the continuous assessment of the transcription and translation tools.

## A virtUOS: Y1 proposal update

Following many technical discussions on X5gon progress and how it can better attend the needs of virtUOS, UOS has updated its pilot proposal so as to be more focused a *OER Recommender for Lecturers*. In what follows, we briefly review the current state of virtUOS, and then the pilot idea is described.

### A.1 Current state

The University of Osnabrueck uses the open-source Learning Management System (LMS) Stud.IP (<http://www.studip.de/>). The University of Osnabrueck is a contributor to the Stud.IP software and has several developers for this software.

Within the LMS every course of the university is represented by a corresponding “learning space”. Within these course a course description, a files module, a bibliography module, an agenda module and in some cases even a courseware module are available.

Usually lecturers provide material (e. g. slides, lecture notes, reading lists) in the files module based on §60a UHRG, which means the material might be copyrighted. The lecturer is not forced to provide license information while uploading a file. So there may also be open licensed documents in the files area.

In the bibliography module the lecturers can provide references to the recommended literature, as regular bibliographic references or even with links to library catalogs.

In the agenda module the lecturers can provide a short description of each course session. Each session may have its own folder in the files module.





The courseware module may contain structured texts, videos and quizzes, similar to the edX courseware.

There can be several other optional plugins.

### A.1.1 Privacy

Additional to the usual privacy regulations based on GDPR, students at a german university have additional rights. Their legal rights during their studies are regulated in laws. The university cannot ask them to sign any mandatory Terms of Service (ToS). If there are any optional ToS, students who do not sign them must not be discriminated in any way. There can be exceptions from the usually privacy regulations for research purposes, but these have to be checked carefully with the privacy officer and they are only provided for a limited time.

The teacher on the other hand may be asked to sign ToS to use a certain service.

## A.2 Pilot Idea

We think that a OER recommendation module for a course would be a good idea that would be beneficial to teachers and learners.

- *Teacher*

First of all it is the teacher who activates and uses the module. The module provides as much (meta-)data from the course as possible as input for the X5gon recommendation service. Maybe even a teacher profile can be created in X5gon to provide more information about the knowledge and the preferences of the teacher and the domain. As teachers have often recurring courses, reusing recommendations would also be a good option. The data provided about the course should only be made publicly available by X5gon services on explicit permission of the teacher.

The X5gon service provides recommendations of additional OER resources to the teacher. The teacher can decide to use these resources, delete some sources, keep some of them “in private” for his own use and to arrange them for the students before making it available to them. This two-step procedure is necessary because teachers will most likely not accept a service that recommends additional material to the students without the teacher’s consent. Apart from this acceptance problem a direct recommendation to the students from within a specified course could be problematic from a pedagogical standpoint as the additional material might distract and confuse the students. The resources may be assigned to a certain session in the agenda and optionally are only shown shortly before or after the session takes place.

The teacher should also have a search box to find additional OER content based on the current course background information, so that he can find additional links for his students or resources that he might want to use in his presentations.

- *Teacher Privacy*

As teachers are free to decide if they want to use the X5gon service or not they can be asked to accept some ToS. But we have to keep in regard that they might be scared away if they have the feeling that are kept under surveillance.

- *Students*

The students only see the OER recommendations provided by the teacher. They can also decide to hide certain resources if they want to. They probably would need an indicator which of the resources they have viewed already.



### – *Student Privacy*

The students must be free to not accept any ToS but still get all the necessary information. But they might get added benefit if they accept the ToS. In our scenario: they would need to see the recommendations from the teacher in any case. But with the ToS signed they would get the indicator which OERs they have already watched. Using a cookie to get at least pseudonymised tracking is probably okay. Providing generalized usage data (how many students have watched this OER) to the lecturer might also be okay.

### • *Challenges*

The first barrier is probably to convince teachers to use such a tool and to notice its added value. We are quite confident that with some marketing within the university we can find a critical mass of teachers to use this tool.

We don't think that marketing activities aimed at the students will be very important or even necessary as they can not activate the recommendation plugin themselves. If a teacher has chosen to use this plugin in one of his courses, the students will instantly notice that an additional plugin has been activated for their course because they typically use the LMS on a daily basis. Also, it is to be expected that teachers using this plugin will inform their students about it in one of the first sessions anyway.

### • *Research and Evaluation*

Within this tool we would get a lot of interaction data of the lecturer that can be used by X5gon to improve recommendations. The interaction data from the students will not be as valuable, as we can probably not notice what OERs have been opened .

Teachers and students can be surveyed on their impression of the X5gon system and the quality of the recommendations. These insights will allow us to derive acceptance criteria that can be useful for other universities, e. g. regarding their OER policies.

### • *Generalization*

If the pilot runs well, the idea can be ported to other LMS systems, like Moodle, Illias, Sakai, Blackboard, etc.

### • *Technical Requirements in X5gon*

#### **REST-Interface**

- to provide general course metadata (description, bibliography, study program, ... )
- to provide session descriptions
- to provide privacy settings
- to provide teacher feedback
- to get general recommendations incl link previews
- to get session related recommendations incl link previews
- to get usage data

#### **Getting usage data from students**

- Alternative 1: Use a Javascript snippet, that just watches which links have been clicked (but does not grab the site!)
- Alternative 2: X5gon referrer links that forward to the requested resource from an x5gon server (like Google or Facebook do it)

## B X5gon TTP: Detailed description

### B.1 Main features

The main features of the X5gon TTP platform are the following:

- **Easy integration of ASR, MT and TTS technologies:** TLP provides a framework via the Ingest Service in which custom ASR, MT and TTS systems can be easily integrated into the platform to provide automatic subtitles, document translations, and synthesised text-to- speech audio tracks.
- **Exploitation of media-related resources:** Topic adaptation techniques for ASR are supported by TLP (i.e. exploitation of text extracted from supplied media slides, related documents, etc.)
- **Support to grid computing:** TLP has been developed expressly to take advantage of computer clusters in order to parallelize at maximum transcription, translation, and text-to-speech processes.
- **Ergonomic solution for subtitle and translation editing:** TLP includes two web-based software tools, the TLP Transcription Editor and the TLP Media Player, that can be used to review transcriptions and subtitles respectively for audio/video objects with ease and comfort; as well as a third web-based tool called the TLP Translation Editor to review text document translations.
- **Support to collaborative subtitling/translation:** TLP enables collaborative video subtitling and document translation by managing edit sessions and keeping track of all user editions made on subtitle and translation files. Changes can be postponed to be reviewed by authors, who can later review these changes to accept or reject them.
- **Automated postprocess triggered by user interactions:** Whenever a transcription or translation is edited, time alignments at the word level for transcriptions, automatic translations, and automatic synthesised audio tracks are re-generated; among other actions. This feature is directly managed by the Postprocessor module.
- **Support to multilingual web translation:** MT technology can be used as to translate static and dynamic web pages or web sites in a very straightforward manner trough the Web Translation Plugin and the Web Translation Editor.
- **Extensive and powerful public API:** TLP offers a complete API that makes very easy the integration process between TLP and remote video repositories, MOOC platforms, institutions, companies, etc.
- **Based on open-source tools and libraries:** All software tools included in TLP are entirely based on open-source languages and libraries.
- **Publicly released as open-source software:** Successive versions of TLP have been publicly released under the open-source Apache License 2.0.

## B.2 Main components

### B.2.1 Database

The Database is a SQL-based relational database which stores all the data required for the Web Service and the Ingest Service. The main entities stored in the Database are the following:

- **Videos/Audios:** All the information related to a specific audio or video is stored in the database, including language, duration, title, keywords and category. An external ID, provided by the client is used to identify the media object in all transactions performed between the client and the API.
- **Speakers:** Information about the speaker/lecturer can be used by the ASR system to adapt the underlying models to the unique characteristics of the given speaker and, therefore, improve the quality of the resulting subtitles.
- **Subtitles:** All subtitles automatically generated by the Ingest Service are stored in DFXP format into the database and retrieved by the client via the Web Service.
- **Audiotracks:** As in the case of subtitles, automatically synthesized audio tracks from translated subtitles are also stored in the database.
- **Documents:** Similarly to video and audio entities, text documents are stored as well along with metadata such as document format, sentences, words, language, etc.
- **Translations:** Again, similarly to subtitles objects, translations in the original document format, as well as in an internal XML format (DTLX format), are stored in the database.
- **Websites:** Metadata of every website or webpage to be translated, as well as all source and translated texts.
- **Uploads:** Every time an /ingest operation is performed, a new upload entry is stored in the database to track its progress.

The platform also needs to store data outside the relational database (in the filesystem of the server). There are defined several data sources, and their root directories or mount points have to be registered in the `mount_points` table of the database:

- **Language Processing modules:** Root directory of all language processing modules to integrated into the Ingest Service (systems key).
- **Video/audio objects:**
  - **Media files:** Root directory of media files, such as video, audio, slides, thumbnail files (media key).
  - **Subtitle files:** Root directory of subtitle files (transcriptions key).
  - **Uploads:** Root directory for the uploaded objects and temporal files (uploads key).
  - **Alignments:** Root directory for the temporal files of text-to-speech alignment processes (valign\_uploads key).
- **Text objects:**
  - **Documents:** Root directory of all text document files (text\_documents key).

- Uploads: Root directory for the uploaded objects and temporal files (text\_uploads key).
- Web objects:
  - Uploads: Root directory for temporal files (web\_uploads key).

### B.2.2 Web Service

The Web Service is the API interface for exchanging information and data between the client and the platform. It also enables the viewing and editing capabilities of the Subtitle Editor. The Web Service defines a wide set of API HTTP interfaces to allow for the full integration of a remote media repository, OER Platform, or website.

The Web Service implements three sub-APIs for the three main classes of objects than can be processed by the platform:

- **/speech** sub-API: is aimed to deal with video and audio files.
- **/text** sub-API: is intended to deal with text documents and contents.
- **/web** sub-API: is intended to specifically deal with text contents extracted from html documents from the client (web browser) side.

There are available several tools and libraries to interact with this API in different programming languages (PHP, Python).

Here we describe the most important /speech API interfaces for X5gon needs:

- Interfaces for media upload and management:
  - /ingest: Upload media (audio/video) files and any attachments and metadata to the TLP Server for automatic multilingual subtitling and speech synthesis.
  - /uploadslst: Get a list of all the user's uploads.
  - /status: Check the current status of a specific upload ID.
  - /systems: Get a list of all available Speech Recognition, Machine Translation, and Text-To-Speech Systems that can be applied to transcribe, translate, and synthesize a media file.
  - /list: Get a list of all existing audio and videos in the Database.
- Interfaces for downloading media and subtitle files:
  - /metadata: Get metadata and media file locations for a given media ID.
  - /langs: Get a list of all subtitle and audiotrack languages available for a given media ID.
  - /get: Download the current subtitle file for a given media ID and language.
  - /getmod: Get subtitle modifications in JSON format of a given media and session ID.

### B.2.3 Ingest Service

The Ingest Service is the service devoted to process video/audio and text/document files uploaded via the `/speech/ingest` and the `/text/ingest` interfaces of the Web Service, respectively. It checks periodically (typically every minute) whether new uploads are awaiting to be processed, also checking if the ongoing uploads are progressing correctly or have failed. The uploads table of the Database is used to keep track of the status of every upload.

The process of the media contents is made in several steps:

1. **Media Package Processing:** The media files are processed for the first time, performing several security, data integrity and data format checks, and, if all checks are correct, the upload status moves to the next processing step.
2. **Transcription Generation:** In this step, a transcription file in DFXP format is generated from the main media file (video, audio) using an Automatic Speech Recognition (ASR) system.
3. **Translation(s) Generation:** In this step, one or more translation files in DFXP format are generated from a transcription file (either automatically generated in the previous step, or provided in the media package file), using the appropriate Machine Translation systems.
4. **Media Conversion:** In this step, the main media file is converted into the media formats required by the Subtitle Editor in order to maximize browser compatibility.
5. **Store Data:** The data contained in the media package and the data automatically generated by the Ingest Service are stored in the Database.

### B.2.4 Subtitle Editor

The Subtitle Editor is an HTML5 media player which allows users to review and modify media subtitles of different languages with ease. It provides a highly ergonomic editing interface, optimized to reduce user effort.

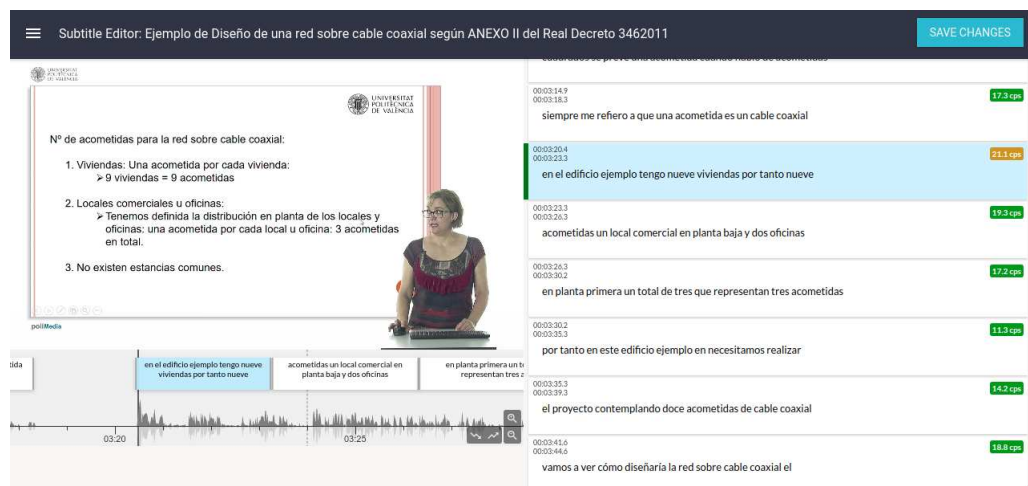


Figure 6: X5gon TTP Subtitle Editor

### B.3 Change log

Since the beginning of the project, the platform is being adapted to be suitable for X5gon needs. We are receiving feedback from the project partners involved in piloting to improve transcription and translation reviewing efforts. Here we add a brief changelog of the developments made along the project:

#### **Version 3.6.2** (2018-01-10)

- Added new export format (IBM JSON)
- Decreased Subtitle Editor inactivity time from 5 to 2 minutes

#### **Version 3.7.0** (2018-02-26)

- New Subtitle Editor set as the official one. The old subtitle editor has been deprecated.
- Subtitle Editor:
  - Huge memory and speed improvements.
  - Added a new transcription edition interface. Former transcription editor has been deprecated.
  - New features: word-based loop playing, improved segment merging, etc.
- Web Service (API):
  - Changes on /speech/get:
    - \* new option pending revs: apply to the current subtitle file those revisions that are still pending to be reviewed.
    - \* implemented true history recovering with "session\_id" option: apply to the original subtitle file all confirmed subtitle modifications upon the given session ID.
  - New /speech/getmod: Get subtitle modifications in JSON format of a given media and session ID.
- Web Interface:
  - Ability to move folders.
- Several bugfixes and improvements.

#### **Version 3.7.4** (2018-05-02)

- PHP Library:
  - Updated library to support PHP 5.6 and up (ingest/new)
- Subtitle Editor:
  - Added support for WebVTT format
  - Added autoconfirm option.
  - Show error when media not supported.
  - New features: undo, shortcut list.
  - Other bugfixes and improvements.





- Old Subtitle Editor:
  - Fixed communication with Web Service.
- Ingest Service:
  - On ffmpeg: use aac library instead of libfdk\_aac
  - Support rescheduled tasks from SGE.

## C X5gon TTP: Transcription and Translation Systems

The development of the current state-of-the-art X5gon ASR and MT systems started in 2011 with the transLectures project. Ever since then, there has been a continuous work on improving these systems so as to provide the most accurate results in a wide range of scenarios. As the evaluations carried out in transLectures included performance metrics of these systems for VideoLectures.Net and poliMedia, it is worth showing what was the starting point and what was the progress of these systems along the project (2011-2014). There were ASR systems for English, Slovene (VL.NET); Spanish and Catalan (pM) . Figure 7 shows the WER figures for all these systems up to October 2014.

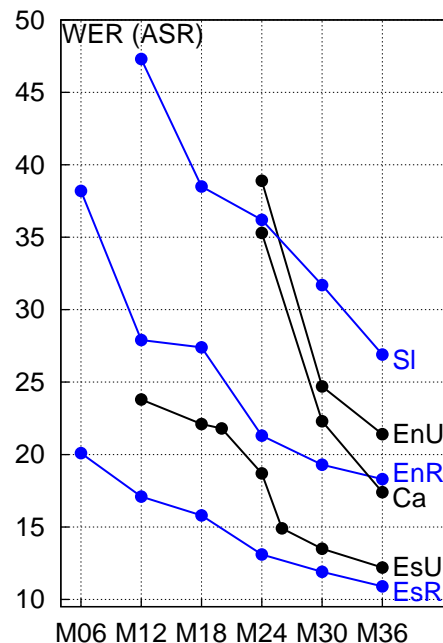


Figure 7: ASR systems progress during transLectures (November 2011 to October 2014). Progress for all languages in ASR given in terms of WER (the lower the better). Each partner's best performing adapted system is plotted for each language, in different colours: UPVLC (black), RWTH (blue). For duplicate system, there is a suffix character to indicate which partner is responsible for which duplicate system: R (RWTH), U (UPVLC).

According to the requirements of the three main pilots involved in the project (VL.NET, poliMedia and VirtUOS), our efforts to improve ASR systems have been focused on four languages: English, Spanish, Slovene and German. For both English and Spanish we are currently using systems that were built previously to the beginning of the X5gon project. Both of them are based on the systems developed for the EMMA project [8]. In particular, the English system remains unchanged, while the Spanish system presents some major improvements: additional training data and the use of



Bidirectional LSTM RNNs for acoustic modelling in the second decoding pass [9]. During the last months of Y1, we devoted some efforts to improve our existing Slovene ASR system (Section C.2), and to build a new German ASR system from scratch (Section C.1).

## C.1 German ASR System

This section describes the ASR system developed for German during Y1. It has been specifically developed to transcribe German videos from the virtUOS pilot. The system is composed of two main separated models: the language model (based on words) and the acoustic model (based on phonemes). Additionally, a third (simpler) model is required to map words into phonemes. This last model, usually called *lexical model*, is used to join both acoustic and language models in order to obtain an automatic transcription.

In what follows we first describe the resources collected for training the different models, and then we introduce more technical details for each of them.

### C.1.1 Compilation of Resources

Resources employed to train the German ASR system are described in this section. They are organised into those needed to train the acoustic model (transcribed speech), those needed to train the language model (plain text), and those for the lexical model (pronunciation dictionaries).

#### Resources for Acoustic Modelling

There are not many freely available speech corpora for training acoustic models. Indeed, for German we only found one corpus: *German Speech Corpus by Technische Universität Darmstadt* (GSC-TUDa) [10]. This corpus only contains 30 hours of annotated speech recorded using 5 different devices (about 158h in total), which is not enough to train a high quality acoustic model. Moreover, the recordings that can be found in GSC-TUDa are quite different from the recordings found in the virtUOS pilot. For these reasons we decided to increase our resources by crawling publicly available data from the web. We collected all types of content: user generated videos, broadcast news, etc. The only requirement was to collect publicly available multimedia objects with subtitles.

Most of the subtitles found around the web can not be considered as phonetic transcriptions, which is required to train acoustic models. Moreover, in most cases the subtitles contain mistakes. In order to overcome these problems, the crawled data was forced-aligned using an initial acoustic model that was trained using the GSC-TUDa corpus. The resulted aligned data was then filtered using some heuristics based on the output of the alignment. As a result we managed to create a speech corpus containing about 716 hours of publicly available data crawled from the web. Table 6 sums up the basic statistics of all audio resources.

Table 6: Statistics of annotated speech resources employed in the estimation of the German acoustic model.

Corpus	Duration(h)	Words(K)	Vocabulary(K)
Crawled Data	716	6150	180
GSC-TUDa	158	1161	4

## Resources for Language Modelling

In contrast to what happens with acoustic resources, there are several monolingual and parallel text corpora which can be used to estimate a German LM. In the case of parallel text corpora, only the German part is selected to estimate the model. Statistics of the LM resources are shown in Table 7. IWSLT De SMT refers to the monolingual part of the *International Workshop on Spoken Language Translation* (IWSLT) English-German statistical machine translation (SMT) task, while IWSLT De ASR 2013 Dev refers to a subset of the IWSLT 2013 German evaluation task used for development purposes.

Table 7: Statistics of resources employed to estimate German LMs

Corpus	Sentences(M)	Words(M)	Vocabulary(K)
Wikipedia [11]	65.2	642.1	8036.1
Europarl [12]	2.2	45.9	354.7
Common Crawl [13]	2.4	44.7	1313.6
News-Crawl [14]	2.0	29.6	661.6
Reuters [15]	0.5	17.6	280.5
Tatoeba [16]	0.3	2.6	86.5
IWSLT De SMT	0.2	3.2	120.9
IWLST De ASR 2013 Dev	0.001	0.02	3.6

## Resources for Lexical Modelling

Instead of using a rule based approach, we decided to build the German lexical model using an automatic approach based on statistical models. For that purpose an initial supervised German pronunciation dictionary is needed. In our case, we used *The CELEX Lexical Database* (WEBCELEX) for German, which contains more than 310K words with their corresponding pronunciations [17].

### C.1.2 System Description

In the following subsections, the models comprising the German ASR system (acoustic, language and lexical models) are described in detail.

#### Acoustic Model

The German ASR system performs speaker adaptation based on fCMLLR features [18]. A direct consequence of this is that it is composed by two hybrid HMM/NNs acoustic models [19]: a standard model and a fCMLLR model. In this setup the standard model is used in a first recognition pass to obtain a initial transcription, which is then used to perform fCMLLR normalisation over input features. The final transcription is obtained recognising the normalised features with the fCMLLR model. For the standard model a conventional feed-forward DNN is used, however for the fCMLLR model a BLSTM was used as in the Spanish system. Estimation of the HMMs and DNNs was carried out using the transLectures UPV toolkit (TLK) [20], while BLSTMs were training using TensorFlow [21].

The characteristics for the final German acoustic models are the following:

- Standard model (1-pass)
  - 18867 tiedphoneme 3-state HMM with a 64-mixture component Gaussian per state.



- A seven hidden layer DNN with the following architecture: 528 ( $48 \times 11$ ) input cells, 18867 output cells and 2048 cells in each internal layer.
- fCMLLR model (2-pass)
  - Same HMM topology than the standard model.
  - A five layer BLSTM with the following architecture: 48 input cells, 18867 output cells and 600 cells in each internal layer for each direction ( $600 \times 2$ ).

## Language Model

The language model for German corresponds to a linear interpolation of several 4-gram models. Specifically, the six corpora in the upper part of Table 7 were used to estimate six independent 4-gram language models which were then interpolated. Interpolation weights were optimised using the *IWSLT De ASR 2013 Dev* set shown in the bottom part of Table 7. When the German ASR system was built, the amount of supervised data from the virtUOS pilot was scarce, so it was decided to use all of them for evaluation purposes. The choice of a subset of the IWSLT task as development, was based on the fact that the IWSLT task is composed of videos from TED talks [22], which have some similarities with educational videos. Therefore, it made sense to select this data for development purposes.

Apart from the conventional n-gram language model, an additional Recurrent Neural Network Language Model (RNNLM) was also trained for lattice rescoring using the Mikolov's toolkit [23]. That is, during the second decoding step (fCMLLR step) the decoder generates lattices containing the most promising hypothesis. The language model probabilities of these lattices are then rescored (replaced) by the ones obtained by interpolating the original 4-gram with the RNNLM. Most probable hypothesis is then selected as the final transcription. For training the RNNLM the IWSLT De SMT data set was used (see Table 7 for more details).

## Lexical Model

The lexical model is a combination of the German dictionary described in Section C.1.1 and a automatic grapheme to phoneme model estimated using the same dictionary. More precisely, the dictionary is first consulted to obtain the pronunciation(s) of a word. If no pronunciation is found, then the grapheme to phoneme model is used to obtain the most reliable pronunciation. The grapheme to phoneme model was estimated using the Sequitur G2P software [24].

## C.2 Slovenian ASR System

The Slovenian ASR System has been developed to transcribe Slovenian videos from the VideoLectures.Net pilot. At the beginning of the project, the Slovenian ASR system developed by UPV within the transLectures project was used. The main characteristics of this system are the following:

- **Acoustic Model:** as in the case of the German ASR system, the Slovenian ASR system is composed by the two hybrid HMM/NNs acoustic models. Both standard and fCMLLR models are multilingual feed-forward DNNs trained using Spanish and Slovene speech data. In particular, 800 hours of Spanish speech data and 66 hours of Slovenian speech data were used to train the acoustic models using the transLectures UPV toolkit (TLK) [20]. The main characteristics for these acoustic models are the following:
  - Standard model (1-pass)
    - \* 9921 tiedphoneme 3-state HMM with a 64-mixture component Gaussian per state.



- \* A six hidden layer DNN with the following architecture: 528 ( $48 \times 11$ ) input cells, 9921 output cells and 2048 cells in each internal layer.
- fCMLLR model (2-pass): same topology than the standard model.
- **Language Model:** The language model corresponded to a liner interpolation of several 4-gram models built within the transLectures project.
- **Lexical Model:** The lexical model is composed by the grapheme representation of the vocabulary words.

Throughout this first period, this initial Slovenian ASR system has been improved by applying Recurrent Neural Network language models (RNNLMs) using the CUED-RNNLM toolkit [25]. These improvements are summarized in Table 8.

Table 8: WER of the Slovenian ASR system.

Corpus	VideoLectures.Net	SI-TEDx-UM
Initial ASR System	32.5	27.6
+RNNLM	31.4	26.3



## References

- [1] poliMedia. <https://media.upv.es/#/catalog>.
- [2] transLectures. <http://www.translectures.eu/web>.
- [3] EMMA. <http://project.europeanmoocs.eu/>.
- [4] Juan Daniel Valor Miró, Pau Baquero-Arnal, Jorge Civera, Carlos Turró, and Alfons Juan. Multilingual Videos for MOOCs and OER. *Educational Technology & Society*, 21(2):1–12, 2018.
- [5] The MLLP Transcription and Translation Platform (MLLP-TTP). <https://ttp.mllp.upv.es>.
- [6] Jorge Civera and Alfons Juan. T36: Final report. Technical report, UPV, 2014.
- [7] poliMedia. <https://politrans.upv.es/>.
- [8] UPVLC. D2.3.2: Report on final transcription and translation models. Technical report, EMMA, 2015.
- [9] Albert Zeyer, Patrick Doetsch, Paul Voigtlaender, Ralf Schlüter, and Hermann Ney. A comprehensive study of deep bidirectional lstm rnns for acoustic modeling in speech recognition. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2462–2466, New Orleans, LA, USA, March 2017.
- [10] GSC-TUDa: German Speech Corpus by Technische Universität Darmstadt. <https://www.lt.informatik.tu-darmstadt.de/de/data/open-acoustic-models>.
- [11] Wikipedia. <https://www.wikipedia.org/>.
- [12] Europarl Corpus: European Parliament Proceedings Parallel Corpus v7. <http://www.statmt.org/europarl/>.
- [13] commoncrawl 2014. <http://commoncrawl.org/>.
- [14] News Crawl corpus (from WMT workshop) 2015. <http://www.statmt.org/wmt15/translation-task.html>.
- [15] REUTERS: Reuters Corpora (RCV1, RCV2, TRC2). <http://trec.nist.gov/data/reuters/reuters.html>.
- [16] Tatoeba. <https://tatoeba.org/eng/downloads>.
- [17] WEBCELEX: The CELEX Lexical Database (English, Dutch and German word features). <http://celex.mpi.nl/>.
- [18] M.J.F. Gales. Maximum likelihood linear transformations for hmm-based speech recognition. *Computer Speech and Language*, 12(2):75 – 98, 1998.
- [19] Geoffrey Hinton, Li Deng, Dong Yu, Abdel rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath George Dahl, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29(6):82–97, November 2012.

- [20] M.A. del Agua, A. Giménez, N. Serrano, J. Andrés-Ferrer, J. Civera, A. Sanchis, and A. Juan. The translectures-upv toolkit. In JuanLuis Navarro Mesa, Alfonso Ortega, António Teixeira, Eduardo Hernández Pérez, Pedro Quintana Morales, Antonio Ravelo García, Iván Guerra Moreno, and DoroteoT. Toledano, editors, *Advances in Speech and Language Technologies for Iberian Languages*, volume 8854 of *Lecture Notes in Computer Science*, pages 269–278. Springer International Publishing, 2014.
- [21] TensorFlow. <https://www.tensorflow.org/>.
- [22] TED. <https://www.ted.com/talks>.
- [23] The RNNLM Toolkit . <http://www.rnnlm.org/>.
- [24] Sequitur G2P. <http://www-i6.informatik.rwth-aachen.de/web/Software/g2p.html>.
- [25] Xi Chen, Xin Liu, Y. Qian, Mark J. F. Gales, and Philip C. Woodland. Cued-rnnlm — an open-source toolkit for efficient training and evaluation of recurrent neural network language models. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6000–6004, 2016.
- [26] Davor Orlic. D7.1: Website. Technical report, X5gon project, M1, 2018.
- [27] Erik Novak. D2.1: Requirements & Architecture Report. Technical report, X5gon project, M6, 2018.
- [28] Stefan Kreitmayer. D4.1: Initial prototype of user modelling architecture. Technical report, X5gon project, M6, 2018.
- [29] D1.1: Quality assurance models. Technical report, X5gon project, M12, 2018.
- [30] D1.2: Report on selected and evaluated quality assurance models. Technical report, X5gon project, M12, 2018.
- [31] D3.1: Learning Analytic Engine 2.0. Technical report, X5gon project, M12, 2018.
- [32] D6.1: Report of the OER network model and interface design evaluation. Technical report, X5gon project, M12, 2018.
- [33] D7.2: First real-world and online community engagement plan. Technical report, X5gon project, M12, 2018.
- [34] D8.1: Detailed market analysis. Technical report, X5gon project, M12, 2018.
- [35] D9.1: Ethical Data. Management and Data. Management Pla: year 1. Technical report, X5gon project, M12, 2018.
- [36] D9.4: First year report. Technical report, X5gon project, M12, 2018.

