

This document is part of the Research and Innovation Action “European Live Translator (ELITR)”.
This project has received funding from the European Union’s Horizon 2020 Research and
Innovation Programme under Grant Agreement No 825460.



Deliverable D6.3

Intermediate Report on Integration

Adelheid Glott (AV), Armin Schweinfurth (AV), Chiara Canton (PV),
Rishu Kumar (CUNI), Ondřej Bojar (CUNI)

Dissemination Level: Public

Final (Version 1.0), 31st December, 2020





Grant agreement no.	825460
Project acronym	ELITR
Project full title	European Live Translator
Type of action	Research and Innovation Action
Coordinator	doc. RNDr. Ondřej Bojar, PhD. (CUNI)
Start date, duration	1 st January, 2019, 36 months
Dissemination level	Public
Contractual date of delivery	31 st December, 2020
Actual date of delivery	31 st December, 2020
Deliverable number	D6.3
Deliverable title	Intermediate Report on Integration
Type	Report
Status and version	Final (Version 1.0)
Number of pages	12
Contributing partners	AV, PV, CUNI
WP leader	PV
Author(s)	Adelheid Glott (AV), Armin Schweinfurth (AV), Chiara Canton (PV), Rishu Kumar (CUNI), Ondřej Bojar (CUNI)
EC project officer	Alexandru Ceausu
The partners in ELITR are:	<ul style="list-style-type: none"> ▪ Univerzita Karlova (CUNI), Czech Republic ▪ University of Edinburgh (UEDIN), United Kingdom ▪ Karlsruher Institut für Technologie (KIT), Germany ▪ PerVoice SPA (PV), Italy ▪ alfatraining Bildungszentrum GmbH (AV), Germany
Partially-participating party	<ul style="list-style-type: none"> ▪ Nejvyšší kontrolní úřad (SAO), Czech Republic

For copies of reports, updates on project activities and other ELITR-related information, contact:

doc. RNDr. Ondřej Bojar, PhD., ÚFAL MFF UK bojar@ufal.mff.cuni.cz
Malostranské náměstí 25 Phone: +420 951 554 276
118 00 Praha, Czech Republic Fax: +420 257 223 293

Copies of reports and other material can also be accessed via the project's homepage:

<http://www.elitr.eu/>

© 2020, The Individual Authors

No part of this document may be reproduced or transmitted in any form, or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.



Contents

1	Executive Summary	4
2	Collaboration and Communication First	5
3	Integration with PerVoice Service Architecture	5
4	Integration among Partner Services	5
5	Continuous Evaluation	6
6	Integration with Audio Sources and Presentation Layer	7
7	End-User Experience	7
8	Events	8
8.1	Student Firms Fair (Prague, March)	8
8.2	SAO WG VAT Workshop (Prague, June)	8
8.3	Liberec	8
8.4	PO Demo	8
8.5	UAB Demo	10
8.6	META-FORUM	10
8.7	SG1	11
9	Minuting Demonstrator	11
10	Conclusion and Future Works	12

1 Executive Summary

This deliverable reports on the integration of the services and systems developed by ELITR's parties into production processes to provide end-to-end workflows for the use cases of the project. It goes through many aspects of the integration task, starting from the human aspect of creating a relationship (Section 2) and a continuous exchange between people to test the technology both in laboratory (Section 5) and on the field during events (Section 8).

WP6 Integration is divided into four tasks and we report on their progress:

T6.1 Platform Updates (months 1-24): The existing PV and AV platform has been updated to be able to handle the new use cases, primarily the multi-target translation and display of subtitles. As a part of this task, a web-based “Publishing platform”, the Presentation Platform has been developed, to deliver the translations into the many target languages to the end users. With a view to making continuous improvements based on user feedback and due to the forced stop of face-to-face events, the activity is completed in its main component but requires further adjustments. See Section 7 for more details.

T6.2 Integration of ASR, SLT and multi-target SMT engines (months 3-36): All the components developed by the research partners have been integrated into the PV Platform and are ready to be used both at SAO events and in alfaview® conferencing platform. No deviation from plan. See Sections 3, 4 and 6.

T6.3 Minuting Demonstrator (months 1-36): The design and implementation of a demonstrator platform that provides live transcript and minuting for participants of online meetings, and its integrate with the ASR systems and the automatic minuting through the Pervoice Service architecture are in progress. No deviation from plan. See Section 9.

T6.4 Running SAO Events (months 7-36): Starting from the early beginning of the project all the interpreting and subtitling systems have been run at the various live events organized by SAO. No deviation from plan. See Section 8 for a brief report of the major events.

This report is concluded in Section 10.



2 Collaboration and Communication First

The key to a successful component integration are communication and collaboration. From the beginning of the project, regular meetings were organized to discuss the problems encountered and the successes achieved by the teams. Two years ago there were many separate entities, today there is a multi-disciplinary team working together to offer a service and tackling problems from many different perspectives.

To share the information and experience among the parties, have been prepared:

- A shared git repository (<https://github.com/ELITR>) for code and utilities,
- A Slack account (<http://elittr.slack.com>) for an improved communication among technicians during live test sessions.

In addition to this very many technical e-mails are being exchanged among partners in the consortium and we also run monthly project calls where we remind ourselves of the main outstanding issues.

3 Integration with PerVoice Service Architecture

Starting from the early beginning of the project, PerVoice has provided libraries and documentation to the parties in order to integrate their services into the PerVoice Service Architecture and an introductory lesson was held in February 2019.

The integration of services is made by using libraries that implement the communication protocol. PerVoice provided to partners the C library as well as the Java and .Net bindings. PerVoice also took into account new implementations and possible bindings in different languages, but most of the partners integrated well with the C library.

CUNI, KIT and UEDIN successfully integrated their components into the PerVoice Service Architecture. Integration has been successfully tested in 2019 and 2020 during the several events and demos.

alfaview® has successfully implemented the live transcription and translation functionality in the alfaview® video conferencing platform, and it has been successfully tested during several project remote meetings held on alfaview® platform.

4 Integration among Partner Services

Much of the integration work between services is done behind the scenes by PerVoice Service Architecture. However, while composing more complex services, we soon realized that some services can seem very good or close to perfect in isolated laboratory conditions but the performance seriously suffers in a real-world settings due to mismatch of tested and real inputs or due to errors in preceding steps of the pipeline. Serious amount of work thus has to be put not only into the technical integration of the services into the common architecture, but also to refinement and adaptation of data conventions at service interfaces.

For example, we soon noticed how ASR accuracy impacts both user experience and subsequent machine translation. During the events, we highlighted the importance of having ASR more robust to L2 speakers and speaker accents and, when possible, with domain adaptation on the topic covered by the event.

Another example is the output with partial hypotheses of ASR services. It doesn't accord well with the machine translation requirement to receive complete sentences. To solve this problem, KIT and CUNI introduced the segmenter and chopper components which are in charge of identifying and presenting complete sentences to machine translation. The work on segmenter tool is well described in D3.1 and still a matter of research.

CUNI is currently using UNIX pipe-based system to navigate audio and text to the ASR and MT workers and final presentation means via PerVoice Mediator and connectors. CUNI is



also researching possible implementations of a live dashboard to monitor the performance and latency of involved ASR and MT systems in real time to ease debugging of the components on-the-fly during live events in case of issues.

5 Continuous Evaluation

In order to ensure that the quality of our ASR and MT systems is growing or at least not decreasing, we want to continuously evaluate them using `elitr-testset`. We aim to make the evaluation automatic and reproducible. We also aim to periodically evaluate our systems on `elitr-testset` to visualize the progress we are making in improving our ASR and MT systems. In its current form, the continuous evaluation is a two step process, the first step consists of generating ASR outputs of files in `elitr-testset` and evaluating the generated ASR transcripts and the second step is to subsequently generate and evaluate the MT output for the ASR transcripts generated in the first step.

To simplify the presentation of results, including overview of performance across various domains of interests, we summarize the results according to the index files, as described in D1.5. The evaluation tools used in this process, namely ASRev (for evaluationg ASR) and SLTev (for evaluating MT) evaluate the scores against a manually transcribed (for ASR) or translated (for MT) file and generate standard scores for individual files. These scores are then averaged for a given index file and are reported in Table 1.

`elitr-testset` as well as our systems are still in development, so not all files can be currently processed by our ASR systems or are missing their human annotated translation files as required by SLTev. Gradually, this compatibility will be improved and we will include all the files in `elitr-testset` in subsequent runs of continuous evaluation.

The current scores reported in Table 1 is generated using KIT's sequence-to-sequence ASR worker (Fingerprint: `en-EU-lecture_KIT-s2s`) and UEDIN's Rainbow MT worker for English input (Fingerprint: `rb-EU_fromEN-en_to_41_all`). More systems will be gradually added for comparison. We are including the commit-id of `elitr-testset` and SLTev/ASRev tool in the evaluation process for better reproducibility of the results as our systems and datasets will evolve. Numbers reported in Table 1 are averaged over the given index file, excluding the files for which ASR/MT were not available.

Our long-term goal with continuous evaluation process is to include a component wise comparison of our MT and ASR workers as well as different pipelining approaches. Regular evaluation of multiple system variants on domain-specific indices will also allow us to immediately choose the current best configuration for a particular subtitling session.



Index Name	Worker	Source Lang	Target Lang	sacreBLEU	WER
auto-langtools-workshop	en-EU-lecture_KIT-s2s	EN	–	–	0.41
asr-english-any-domain	en-EU-lecture_KIT-s2s	EN	–	–	0.45
asr-english-any-domain	rb-EU_fromEN-en_to_41_all	EN	CS	14.39	–
asr-english-any-domain	rb-EU_fromEN-en_to_41_all	EN	DE	18.38	–
auto-iwslt2020-antrecorp	en-EU-lecture_KIT-s2s	EN	–	–	0.46
auto-iwslt2020-antrecorp	rb-EU_fromEN-en_to_41_all	EN	CS	13.66	–
auto-iwslt2020-antrecorp	rb-EU_fromEN-en_to_41_all	EN	DE	17.95	–
auto-iwslt2020-devset	en-EU-lecture_KIT-s2s	EN	–	–	0.56
auto-linguistic-mondays	en-EU-lecture_KIT-s2s	EN	–	–	0.42
auto-asr-english-any-domain	en-EU-lecture_KIT-s2s	EN	–	–	0.46
auto-asr-english-any-domain	rb-EU_fromEN-en_to_41_all	EN	CS	14.39	–
auto-asr-english-any-domain	rb-EU_fromEN-en_to_41_all	EN	DE	18.38	–
auto-iwslt2020-wgvat	en-EU-lecture_KIT-s2s	EN	–	–	0.38
auto-iwslt2020-wgvat	rb-EU_fromEN-en_to_41_all	EN	CS	15.94	–
auto-iwslt2020-wgvat	rb-EU_fromEN-en_to_41_all	EN	DE	18.41	–
auto-asr-english-auditing	en-EU-lecture_KIT-s2s	EN	–	–	0.37
auto-asr-english-auditing	rb-EU_fromEN-en_to_41_all	EN	CS	16.45	–
auto-asr-english-auditing	rb-EU_fromEN-en_to_41_all	EN	DE	19.60	–
auto-iwslt2020-consecutive	en-EU-lecture_KIT-s2s	EN	–	–	0.36
auto-iwslt2020-consecutive	rb-EU_fromEN-en_to_41_all	EN	CS	18.49	–
auto-iwslt2020-consecutive	rb-EU_fromEN-en_to_41_all	EN	DE	24.35	–
auto-iwslt2020-khanacademy	en-EU-lecture_KIT-s2s	EN	–	–	0.55

Table 1: This is an auto-generated table to report the ASR and MT evaluation on ELITR-testset commit-id **8ee98c622159afa38de994201eb0fc7c2c6d4297** and SLTev commit-id **764d52a1158d37d569c9454c4c43cbbf3fd57a8f**. The prefix “auto-” in index names indicates that the index itself is generated automatically by scanning `elitr-testset` documents.

6 Integration with Audio Sources and Presentation Layer

In a real use case it is also important to consider the integration with the input source and the output destination. Client software takes audio in input as raw PCM 16KHz, therefore basically, the input can be taken from a common audio jack connected to the conference control room as long as it has the required characteristics. Otherwise, a trans-coding step must be introduced which transforms the audio signal into the required format. We also tested our capacity to follow several audio input channels (main stage speaker audio and interpreter’s booths audio) using different client and input streams.

Audio quality and volume has arisen as a key point to improve ASR output accuracy. The possibility to check the audio input from a centralized point must be taken into consideration from the outset when organizing events for the project.

Presenting the subtitles is the second half of the problem. We tried overlaying subtitles over the main slides presented at main stages using a video-mixing software and also an hardware mixer.

The opposite problem is to grab slides from the presenter computer to deliver them to the Publishing Platform running at end-user devices such as laptops or cell phones. Two software solutions thus have been developed: screen capture on the presenter laptop followed by streaming in the form of a video, and regular screen shots and their publication upon screen shot change.

Now that the events take place exclusively online, a way has been studied to also make an HLS streaming of a video recording.

7 End-User Experience

Last but not least is the user experience, because the user who uses the service is the last element in the chain of integration. PerVoice and alfaview® successfully extended their platforms to deliver live transcription and translations to end users.

The alfaview® platform has been extended and integrated with the PerVoice Service Architecture to deliver live transcription and translation into multiple languages to remote meeting



participants. The feature has already been tested successfully in the field by alfatraining, an educational provider who uses alfaview®, and also during ELITR periodic remote meetings.

PerVoice has developed a web-based subtitle streaming platform, the Presentation Platform, which has been used during the Annual meeting of the EUROSAT Core Group 2, in June 2019 in Prague. During the event it delivered live and automatic subtitles to roughly 30 participants in 10 different languages.

CUNI proposed another approach to present the transcribed and translated text, the Online Text Flow web application focusing on providing the user with the complete history of the running text and improved translation into multiple languages at the sentence level. The tool has been deployed during the Linguistic Mondays seminar series at CUNI (15–30 attendees weekly), as well as at the preparatory run of the EUROSAT 2020 Congress workshop on Using Language IT Technologies in Audits held at the SAO Czech Republic in February 2020 (21 participants).

For a more complete overview of presentation methods and reasoning on user experience put in place by the ELITR project, please refer to D6.1.

Due to Covid-19, now events take place exclusively online and a new and unexpected challenge has arisen to make the contents of the conferences usable also in streaming.

8 Events

The driving force towards integration was the commitment, from the beginning of the project, to participate in events and demos. Below are mentioned the main events but it is important to note that we kept running also others minor events during these months.

8.1 Student Firms Fair (Prague, March)

From the 20th to the 22nd of March 2019, ELITR attended the first dry-run event: The Student Firms Fair in Prague.

During this event, the first integrated parties' services (the KIT and CUNI ones) have been used to provide live subtitles on different languages which were spoken on the presentation podium, see Figure 1.

8.2 SAO WG VAT Workshop (Prague, June)

The 28th-29th of June 2019, ELITR has tested transcribed and translated subtitles into 10 different languages at WG VAT workshop, a closed event run by SAO for 30 participants (Figure 2). The annual meeting of representatives of the Core Group 2 of the EUROSAT working group on VAT WG issues was held in Prague. The subtitles have been live delivered to the participants through the Presentation Platform on the provided on their own laptops. Also, the interpreters' input has been tested: During the event, some interpreters provided additional version of English (only "shadowing" or "re-speaking" the original speech) and German and Czech interpretation.

8.3 Liberec

The event was run by SAO in the city of Liberec, as a regular meeting of the EUROSAT organizers. This workshop took place on October 9th, speakers spoke in English and we tested both several translation services (Czech, German, Hungarian, Polish, Spanish) and subtitle usability (Figure 3).

8.4 PO Demo

Unfortunately, due to Covid-19 restrictions, EUROSAT Congress, planned to took place in May 2020, was postponed. In order to achieve the project milestone, we planned an online demo with

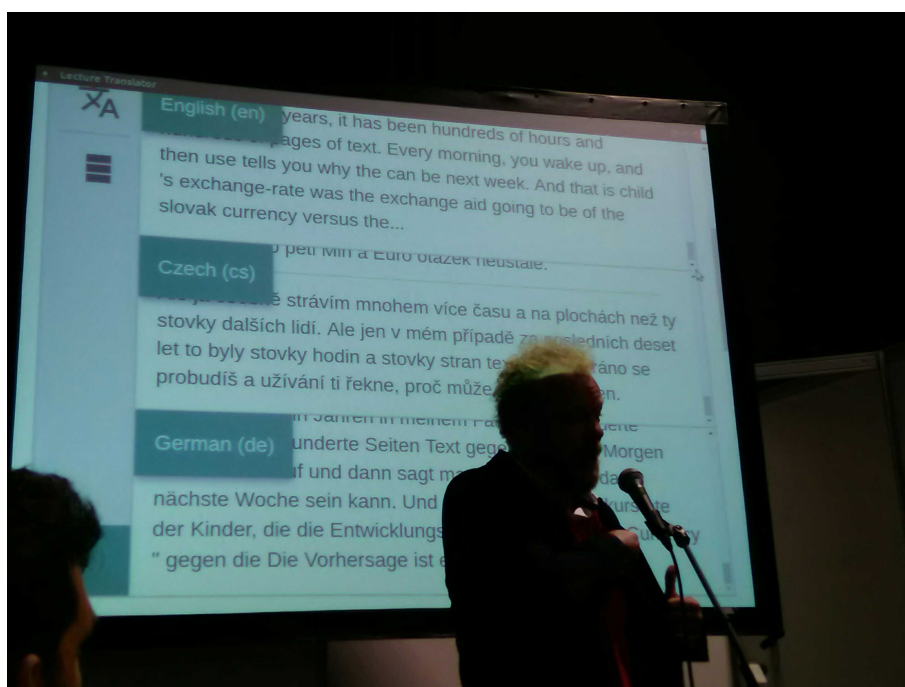


Figure 1: ELITR live subtitles in March event complementing the speech of Tomáš Sedláček, a renowned Czech economist.



Figure 2: WG VAT interpreters booth view



Figure 3: End-user using Presentation Platform at Liberec's event

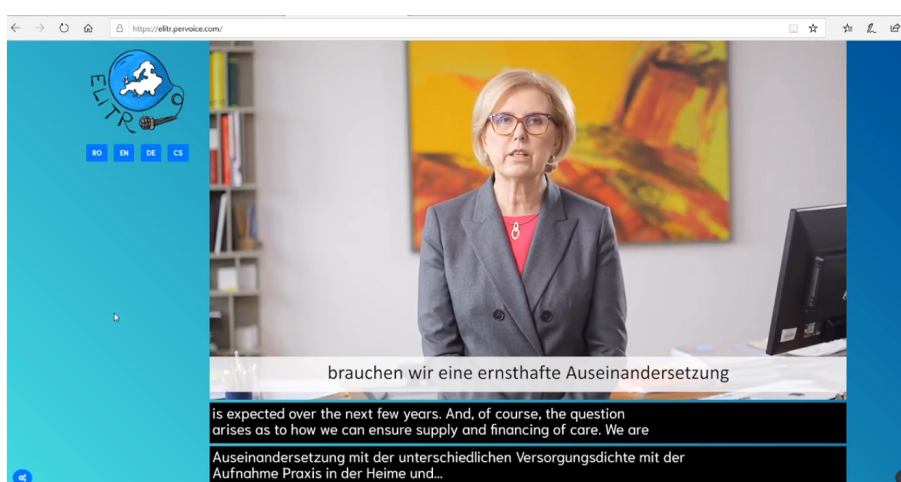


Figure 4: Demo presented in Presentation Platform during UAB demo event

the ELITR Project Officer which took place on June 17th, 2020. For the first time the UEDIN's multi-lingual MT service has been used in an official demo. The transition from a face-to-face to an online scenario gave us the opportunity to explore new possibilities and solutions.

8.5 UAB Demo

On August 19th, 2020, we had the demo with the User Advisory Board of the project (Figure 4). The UAB session was useful also for other parts of the project, e.g. we were discussing data privacy concerns when collecting data for WP5 Minuting.

8.6 META-FORUM

The META-FORUM 2020 conference took place from 1st to 3rd of December (Figure 5). We had the opportunity to participate to the poster session and during the main event we subtitled informally the event in 8 languages (English, German, Romanian, Czech, French, Hungarian, Polish and Dutch). Moreover the KIT's new S2S ASR service has been put online for the first

European community to use.	Tastaturhörung, die nicht die Zeit hat.	Comunități Europene.	krajiny, a chcemo umohat evropckému společenství využít.	335. Un euro est donc l'audition du clavier qui n'a pas le temps.	számára, hogy felhasználasson. Tehát az euró az, hogy a billentyűzethallgatónak.	umožní vám vykorzovat spolupráci evropské.	
335. So an Euro is the keyboard hearing do not have the time.	336. Leider sprechen wir über alle Sprachen und über alle Länder, die wir in Ägypten vertreten haben, und über das Netzwerk des Nationalen Kompetenzzentrums, das wir um ja arbeiten oder die wir als Eugene eingerichtet haben.	336. Din referință, să vorbim despre toate limbile, despre toate țările pe care le-am reprezentat în Egipt, despre rețeaua Centrului Național Competent pe care o desfășurăm, da, sau pe care le-am înființat în calitate de Eugene.	336. Takže euro je slyšení klávesnice nemají čas. Bohužel, mluvit o všech jazycích a o všech zemích, které jsme zastupovali v Egyptě, a o síti národního kompetenčního centra, kterou provozujeme, ano, nebo které jsme založili jako Eugene.	336. Malheureusement, pour parler de toutes les langues, de tous les pays que nous représentons en Égypte, et du réseau du Centre national compétent que nous travaillons ou que nous avons mis en place comme Eugene.	336. Sajnos minden nyelvről és minden országról, amelyeket képviselünk Egyiptomban, és a nemzeti hálózatról rendelőző központ háttérrel, amelyet mi, igen, működünk, vagy amelyet Eugene-ként hoztunk létre.	336. Také bychom mohli mluvit o některých z těchto jazyků a o některých z těchto zemí.	
337. But we can at least talk about some of these Yassos.	337. Aber wir können wenigstens über einige dieser Yassos sprechen.	337. Dar putem măcar vorbi despre unele dintre aceste Yassos.	337. Ale můžeme alespoň mluvit o některých z těchto jazyků a o některých z těchto zemí.	337. Mais on peut au moins parler de quelques-uns de ces Yassos.	337. De legalább beszélhetünk o néhány ezekről a Yassokról.	337. Ale możemy przynajmniej porozmawiać o niektórych z tych Yassos.	
338. And a couple of selected examples.	338. Und ein paar ausgewählte Beispiele.	338. Și câteva exemple selectate.	338. A několik vybraných příkladů.	338. Et quelques exemples sélectionnés.	338. És néhány kiválasztott példa.	338. I kilka wybranych przykładów.	
339. And so just for those of you who are not aware of what the are set up in 32 National Competent Center.	339. Und so nur für diejenigen von euch, die sich nicht dessen bewusst sind, was sie im 32. Nationalen Kompetenzzentrum geschaffen sind.	339. Și astfel doar pentru cei dintre voi care nu sunt conștienți de ceea ce sunt înființate în 32 Centrul Național Competent.	339. A tak jen pro ty z vás, kteří si nevšimají, co jsou vytvořeny v 32 Národním kompetenčním centru.	339. Et donc juste pour ceux d'entre vous qui ne savent pas ce qu'ils sont mis en place dans 32 Centre National Competent.	339. Tehát csak azoknak, akik nem tudják, hogy mit hoztak létre a 32 Nemzeti Kompetens Központban.	339. Tak więc tylko dla tych z was, którzy nie wiedzą, co to zostało założone w 32 Krajowym Kompetencyjnym Centrum.	

Figure 5: Demo presented in OnlineTextFlow during META-FORUM test event

time.

8.7 SG1

The annual meeting of the SG1 group of the SAO also took place on 3 December 2020. The configuration was the same as the META-FORUM 2020 conference, however we provided to the participants only the languages of interests (English, Czech and German) through OnlineTextFlow platform.

9 Minuting Demonstrator

During 2020 alfaview® platform has been extended with the possibility to export created transcription and translations. As already written in the proposal, ‘The full development of such an online editor with concurrent access is beyond the capacity of this proposal.’ In order to enable the integration of a collection of additional meeting functions, alfaview® has developed a toolbox with integrated link service. This means that in the backend of alfaview®, the moderator can attach individual links and buttons to an alfaview® room, which can be accessed by all participants with a single mouse click. For example, it is possible to link a GoogleDoc, call it up from alfaview® and a stenographer can directly log the meeting and create short minutes. PerVoice instead will investigate the possibility to natively publish transcriptions on an online editor like GoogleDocs or Teletype for Atom.

For the Minuting Demonstrator, English will be the only supported language.

Our first valuable goal is to develop a nice and usable interface. The user interface will be a two-sections text editor. In the first section there will be the automatic transcription of the meeting and in the second the meeting agenda. At first stage, a human user will take the automatic transcription complete sentences, copy-and-paste them into the agenda structure (the agenda is part of the input) and then he will summarize them. This “wizard-of-oz” approach to minuting will help in understanding good strategies of minuting and it can also lead to gathering additional minuting data for the purposes of automation of the process.

After having a stable manual workflow, we will start introducing the automations presented in *D5.1: Initial Report on Summarization*, but always keeping the user interface as flexible as possible. Automatic Minuting is an open field research topic, but we agreed that everything we will be able to provide, will be an advantage both for us and for the next researchers working on this topic.



10 Conclusion and Future Works

On the integration the consortium moved with great decision from the very first stages of the project. The partners integrated without difficulty from the early beginning, but the real challenges came by working on the field and putting into practice our researches. There are some open questions that will be investigated over the next year. For example the optimized integration of multi-target translation services in the PerVoice Service Architecture and the most suitable segmentation of ASR output. alfaview® has found a good solution for linking a minuting demonstrator with the implementation of the toolbox. Depending on the development and the time required, alfaview® can imagine to continue to look more deeply into the integration of a minuting demonstrator. There are also open points on less technical aspects such as the coordination of many people in different locations during the live events but we will keep working on these aspects with the dedication and constancy that has characterized the ELITR project till today.