

# Notes on subtitles

by Maria Dimou / Academic Training & IT e-learning

Thanks are due to Vint Cerf, Ken Harrenstien, Dimitri Kanevsky (Google), Alex Manzoni (Red Cross), Andreas Hoecker, Manuella Vinciter, Thomas Baron, José Benito Gonzalez, Pete Jones, Jean-Yves LeMeur, René Fernandez, Lorys Lopez, Ruben Gaspar (CERN), Matthew Goodman (EPFL).

## Background

We have colleagues who can't hear. Elementary diversity awareness requires that we equip all CERN-made videos with subtitles. This note is about an investigation of:

1. A good transcription software.
2. A scalable way to display subtitles in CDS.

## Executive summary of conclusions

Scroll down for the latest news at the end of this page.

### About the transcription software - December 2019 status

The selected product should be used for *all* events, live and recorded (also the ones of the past). An Open and free-of-charge solution would be ideal, if quality corresponds.

The one used by *YouTube* seems to be the best quality-wise from the ones we've seen.

Concerns were raised by ATLAS and IT/CDA-DR management about potential legal aspects, the text having transited, via the API, through YouTube. The concern refers data ownership in general and data confidentiality in particular of *restricted* videos from important meetings.

A formal proposal can be presented to the HR *Diversity Office* requesting (co-)funding of the total audiovisual webcasting/recording process review, with systematic subtitles' inclusion in ALL cases. Manual review of the subtitles' quality is *indispensable* if we opt for quality results. This requires a permanent human effort provision that we may not afford. Hence, it has to be a priori agreed how much error margin we can accept in the output.

See Appendix for details.

### About displaying from CDS with subtitles - December 2019 status

We need a new *player* software to display simultaneously *camera* and *slides* .mp4 **with** subtitles. Lorys found the Open Source player [paella](https://paellaplayer.upv.es/) (<https://paellaplayer.upv.es/>) being now evaluated in the team of audiovisual experts in IT/CDA-IC. If proven as good as it looks, it will replace the current [Theo player](https://www.theoplayer.com/) (<https://www.theoplayer.com/>). *paella* does display the [CERN Document Server](https://cds.cern.ch/) (<https://cds.cern.ch/>) (CDS) and videos.cern.ch required **.vtt** format. Moreover, if the filename has the right format, *paella* displays the subtitles with no manual configuration needed (see [here today's required manual configuration](https://it-e-learning.docs.cern.ch/administration/#how-do-i-include-the-relevant-subtitlesvtt-file-in-a-published-video-record) (<https://it-e-learning.docs.cern.ch/administration/#how-do-i-include-the-relevant-subtitlesvtt-file-in-a-published-video-record>)).

A requirement of the audiovisual service experts is to find new/long-lasting operational solutions for the whole stack of webcast/recording/transcription equipment and software.

We are in a good path with the [opencast](https://opencast.org/) (<https://opencast.org/>) evaluation concerning this matter,

as *Opencast* seems able to replace both [Sorenson](https://en.wikipedia.org/wiki/Sorenson_Media) ([https://en.wikipedia.org/wiki/Sorenson\\_Media](https://en.wikipedia.org/wiki/Sorenson_Media)) for transcoding and [Micala](http://micala.sourceforge.net/) (<http://micala.sourceforge.net/>) for archiving.

For the few videos which contain subtitles at present, CDS displays well [the e-learning collection](https://cds.cern.ch/collection/E-learning%20modules?ln=en) (<https://cds.cern.ch/collection/E-learning%20modules?ln=en>) because it *does* invoke the player. For records containing several files .mp4 (e.g. camera and slides) and .vtt (i.e. even if subtitles are already present) CDS currently only offers to download, hence the subtitles are not merged.

## Details about the transcription software

- 1. amara** For the IT e-learning short videos we do the subtitles by hand using [amara](https://amara.org/en/) (<https://amara.org/en/>). This is affordable because the videos are ≤5 minutes-long & the script is pre-written. Still the process making the subtitles\_en.vtt file and configuring CDS is entirely manual and very time consuming (>3 hours for a 5 mins' video). All relevant links here:
  - [Video library sorted by topic](https://twiki.cern.ch/EduTech/VideoLibrary) (<https://twiki.cern.ch/EduTech/VideoLibrary>)
  - which are sorted views of [the relevant CDS collection](https://cds.cern.ch/collection/E-learning%20modules?ln=en) (<https://cds.cern.ch/collection/E-learning%20modules?ln=en>).
  - Documentation on [using amara](https://it-e-learning.docs.cern.ch/video/faq/#q5-how-do-i-introduce-subtitles) (<https://it-e-learning.docs.cern.ch/video/faq/#q5-how-do-i-introduce-subtitles>)
  - Documentation on [configuring subtitles' display in CDS](https://it-e-learning.docs.cern.ch/administration/#how-do-i-include-the-relevant-subtitlesvtt-file-in-a-published-video-record) (<https://it-e-learning.docs.cern.ch/administration/#how-do-i-include-the-relevant-subtitlesvtt-file-in-a-published-video-record>)
  - How we did [mass subtitling of existing videos](https://it-student-projects.web.cern.ch/projects/e-learning-it-collaboration-devices-applications-insert-subtitles-video-tutorials) (<https://it-student-projects.web.cern.ch/projects/e-learning-it-collaboration-devices-applications-insert-subtitles-video-tutorials>).
- 2. YouTube** For selected videos from the [Academic Training recordings in CDS](https://cds.cern.ch/collection/Academic%20Training%20Lectures?ln=en) (<https://cds.cern.ch/collection/Academic%20Training%20Lectures?ln=en>) we rely on *YouTube* transcription with impressively good quality results. Nevertheless, family names and particle names are not always right), examples in [the CERN Lectures YouTube channel](https://www.youtube.com/channel/UCwXkOx0EuKBR5m_OOiaZRUA) ([https://www.youtube.com/channel/UCwXkOx0EuKBR5m\\_OOiaZRUA](https://www.youtube.com/channel/UCwXkOx0EuKBR5m_OOiaZRUA)). Input from Jean-Yves on the quality: *Domage, lorsque le discours devient technique, c'est quand même approximatif. Par exemple, regarde* <https://www.youtube.com/watch?v=8Hl5e5Y5Tho> (<https://www.youtube.com/watch?v=8Hl5e5Y5Tho>) *autour de 42mn10s. At 42mn59s, "neutrino" is transcribed into "student Reno's" :-).*  
The [current process \(aveditor\)](https://it-e-learning.docs.cern.ch/aveditor/#the-academic-training-use-case) (<https://it-e-learning.docs.cern.ch/aveditor/#the-academic-training-use-case>) to publish a CDS record in YouTube doesn't scale (manual, very time consuming and not covering recordings with access restricted). Notes on tests done by Pete and Alex:
  - Some [help for manual download](https://www.4kdownload.com/howto/howto-download-youtube-subtitles) (<https://www.4kdownload.com/howto/howto-download-youtube-subtitles>). Manual and temporary if YouTube policy changes.
  - For Linux users, [the youtube-dl script](https://www.ostechnix.com/download-youtube-videos-with-subtitles-using-youtube-dl) (<https://www.ostechnix.com/download-youtube-videos-with-subtitles-using-youtube-dl>).
  - Other [downloader](https://youtubedownload.video/en1/) (<https://youtubedownload.video/en1/>).
  - Transcription [process in our e-learning collection](https://twiki.cern.ch/EduTech/TranscribeYourVideo) (<https://twiki.cern.ch/EduTech/TranscribeYourVideo>).
- 1. S2T** The Digital Memory project evaluated [S2T tool by WIPO](https://www.wipo.int/s2t/) (<https://www.wipo.int/s2t/>) (World Intellectual Property Organisation): It has been trained with 60 hours of training data from HEP so far (after we signed a use agreement). [First comparative report](#)

(<https://cds.cern.ch/record/2304470>). It makes reference to the Google Cloud Speech API. It was written in 2017. Conclusion was that human review was indispensable. Examples:

- Reviewed by human [here \(http://digital-memory.web.cern.ch/digital-memory/media-archive/video/open/subtitles/script/displaySubtitle.php?videoid=Video-2293460-a\)](http://digital-memory.web.cern.ch/digital-memory/media-archive/video/open/subtitles/script/displaySubtitle.php?videoid=Video-2293460-a).
  - Not reviewed [here \(http://digital-memory.web.cern.ch/digital-memory/media-archive/video/open/subtitles/script/displaySubtitle.php?videoid=Video-423243\)](http://digital-memory.web.cern.ch/digital-memory/media-archive/video/open/subtitles/script/displaySubtitle.php?videoid=Video-423243)
2. **ai-media** Australian commercial company [Ai-Media \(https://www.ai-media.tv/\)](https://www.ai-media.tv/), also in the UK. Was hired by Thomas to do some live captioning of Vidyo meetings. Contact James.Ward@ai-media.tv. The service involves paying money. Input by Jean-Yves: *I think the only service tested with AI-Media was live captioning, which turned out to be below the expectations of the user (Thomas can provide the details as I think it was on a very specialized topic).*
  3. **3playmedia** USA commercial company [3playmedia \(https://www.3playmedia.com/\)](https://www.3playmedia.com/) was listed as candidate, evaluation not yet done. The service involves paying money.
  4. **MLLP service** (<https://ttp.mllp.upv.es/index.php?page=fag>) from the [Universitat Politècnica de València \(UPV\) \(http://www.upv.es/\)](http://www.upv.es/), a "service" by their automatic transcription/translation/interpretation research unit, sponsored by the EU for [the EMMA project \(http://project.europeanmoocs.eu/about/\)](http://project.europeanmoocs.eu/about/).
  5. **EPFL input** Maria contacted MOOC makers (Matthew Goodman) at EPFL. They mostly use *amara* for the MOOCs, like us for the e-learning. For all-day events like [the EPFL Open Science day \(https://www.epfl.ch/campus/events/celebration-en/open-science-day/\)](https://www.epfl.ch/campus/events/celebration-en/open-science-day/) they just rely on *YouTube*.

## Details about the CDS display

Investigation started in June 2018 and is managed by SNOW tickets. On-going as per the Executive summary above.

- [2018 analysis and conclusion \(https://cern.service-now.com/service-portal/view-request.do?n=RQF1041572\)](https://cern.service-now.com/service-portal/view-request.do?n=RQF1041572)
- [November 2019 investigation \(https://cern.service-now.com/service-portal/view-request.do?n=RQF1460345\)](https://cern.service-now.com/service-portal/view-request.do?n=RQF1460345) of possible new products and workflow.

## Appendix

Maria contacted Vint Cerf, Internet father, now at Google, who also has a hearing impairment. Replies below led to the recommendation in the Executive Summary above. Comments by Ken Harrenstien (Google): Human review always required. One advantage of YouTube is that you can solicit volunteers world-wide to help fix and translate caption tracks. Suggested APIs:

1. <https://developers.google.com/youtube/v3> allows to upload a private video, wait a while until the ASR track is created, download that track, then delete the video. Free.
2. [Cloud Speech to Text API \(https://cloud.google.com/speech-to-text/\)](https://cloud.google.com/speech-to-text/) almost the same as what YouTube uses.
3. [Google Live Transcribe code now Open Source \(https://github.com/google/live-transcribe-speech-engine\)](https://github.com/google/live-transcribe-speech-engine). It can be tried via Android devices [here](#)

<https://www.android.com/accessibility/live-transcribe/>).

## Subtitles' quality review

For reviewing the transcription we could see if the CERN host-states' office can advise us on contracts like:

- [Place d'apprentissage \(https://junior.gateway.one/apprentissages/lieu-geneve?region=fr-CH\)](https://junior.gateway.one/apprentissages/lieu-geneve?region=fr-CH) which involves small amount of money contributed by the state or
- [The swiss army civil service \(https://www.zivi.admin.ch/zivi/fr/home.html\)](https://www.zivi.admin.ch/zivi/fr/home.html)

Nevertheless such contract solutions can't be available without interruption. ATLAS doesn't mind the imperfections of the automatic transcription, as the community is aware of the expected terms. In any case, it is important to have an agreement on the quality of the expected transcription result from the beginning of the project.

## Actions' log

### 2020/01/09 CDA Internal meeting

Present: Tim, Jose, Nicola, Jean-Yves, Thomas, Rene, Maria D., Ruben.

[Meeting agenda \(https://indico.cern.ch/event/868274/\)](https://indico.cern.ch/event/868274/). One can see, on this event, 2 lectures with subtitles made while Ruben tested the [MLLP service \(https://tp.mllp.upv.es/index.php?page=faq\)](https://tp.mllp.upv.es/index.php?page=faq) from the UPV University, a "service" by their automatic transcription/translation/interpretation research unit, sponsored by the EU for [the EMMA project \(http://project.europeanmoocs.eu/about/\)](http://project.europeanmoocs.eu/about/).

### Summary and Conclusions

#### Transcoding tools

Ruben reminds us of the legacy components: CES (Central Encoding System), [Micala \(http://micala.sourceforge.net/\)](http://micala.sourceforge.net/) and [Sorenson \(https://en.wikipedia.org/wiki/Sorenson\\_Media\)](https://en.wikipedia.org/wiki/Sorenson_Media) (company no more exists). Affected services are CDS, Indico and SNOW.

**Transcoding** The, originally Zurich University, Open Source product [Opencast \(https://opencast.org/\)](https://opencast.org/) is chosen, as *Opencast* is well maintained by a large community and covers both our transcoding and archiving needs.

**Player** The new Open Source [paella \(https://paellaplayer.upv.es/\)](https://paellaplayer.upv.es/) is chosen. It comes with *Opencast*, displays simultaneously *camera* and *slides* .mp4 **with** subtitles in .vtt format. Any user can edit the paella subtitles as a file and send it back to us (maintainers in IC) for validation and integration in the subtitles. It is validated also in webcast and mobile environment. It will replace the current [Theo player \(https://www.theoplayer.com/\)](https://www.theoplayer.com/).

- Timeline for Opencast: Collaboration with Universitat Politècnica de València (UPV) signed. PJAS comes in July 2020. Consultancy contract can be signed earlier. Ruben to communicate his optimal plan with the order of services affected. Get agreement from DR section.
- About the actual videos' filesystem: Opencast has NFS as default. We shall use CEPH (ask the service managers to mirror the current mediaarchive now in DFS). Opencast's access to storage is done via a web service.

- About restricted videos: Nicola finds this is an opportunity to improve the current workflow, where the user has to login twice, once for the restricted event and one for viewing the video itself.

**Running of the services** Responsibility remains with Ruben, will need DR participation on the CDS integration part of the new solutions.

## Transcription tools

List of products reminded and enhanced with the [MLLP service \(https://ttp.mllp.upv.es/index.php?page=fag\)](https://ttp.mllp.upv.es/index.php?page=fag) from the [Universitat Politècnica de València \(UPV\) \(http://www.upv.es/\)](http://www.upv.es/), by their automatic transcription/translation/interpretation research unit, sponsored by the EU for [the EMMA project \(http://project.europeanmoocs.eu/about/\)](http://project.europeanmoocs.eu/about/).

## Scope

We should provide subtitles for *all pre-recorded* videos, not live. Subtitles should appear by default, unless the info is important and the event manager asks not to include the automatically-produced ones but check/make them manually, to guarantee transcription quality. When we reach operation status, the one who checks the automatic transcription's quality should be the **speaker** (Tim).

## Preferred products

- [MLLP \(https://ttp.mllp.upv.es/index.php?page=fag\)](https://ttp.mllp.upv.es/index.php?page=fag) can take slides and article/script/notes accompanying a video and use them via their AI algorithms to fix the vocabulary of the automatic transcription. They also provide a solution for live speech. We could offer to run the service here as they lack computing power. This solves the problem of "black box in Valencia".
- [WIPO S2T \(https://www.wipo.int/s2t/\)](https://www.wipo.int/s2t/) is also free-for-CERN and Open Source. It requires GPUs. We have an evaluation license. We shall have to go through the CERN Legal service if we wish to run it as a service.

**After the meeting** Jean-Yves circulated [the requirements' list \(https://codimd.web.cern.ch/s/Hk1edZjIU#\)](https://codimd.web.cern.ch/s/Hk1edZjIU#/) he has for the Digital Memory project. Ruben and Jean-Yves will extend the table to select one of the 2 above products best suiting the criteria best, also in terms of functionality, OS and infrastructure. Thomas suggests to add the optional requirement for the selected *automated transcription* product for our offline base of videos, to *also* do live transcription. Maria to write a TECH student project proposal and request *Diversity Office* funding. The student will set-up the *service* for the pre-selected product.

## A.O.B.

- There was a question of impact of the transcoding infrastructure change to the timescale of videos' move from cds.cern.ch to videos.cern.ch. This is not obvious and should be discussed at the next checkpoint meeting.
- Suggested date **Thursday March 12th @ 4PM (???)**
- Discussion items include:
  - Storage issues, i.e. EOS vs CEPH.
  - Where to display the final video.