Video Workflow

This guide describes the steps that have to be taken to create and ingest an annotated video recording into the language resource archive.

General Framework:

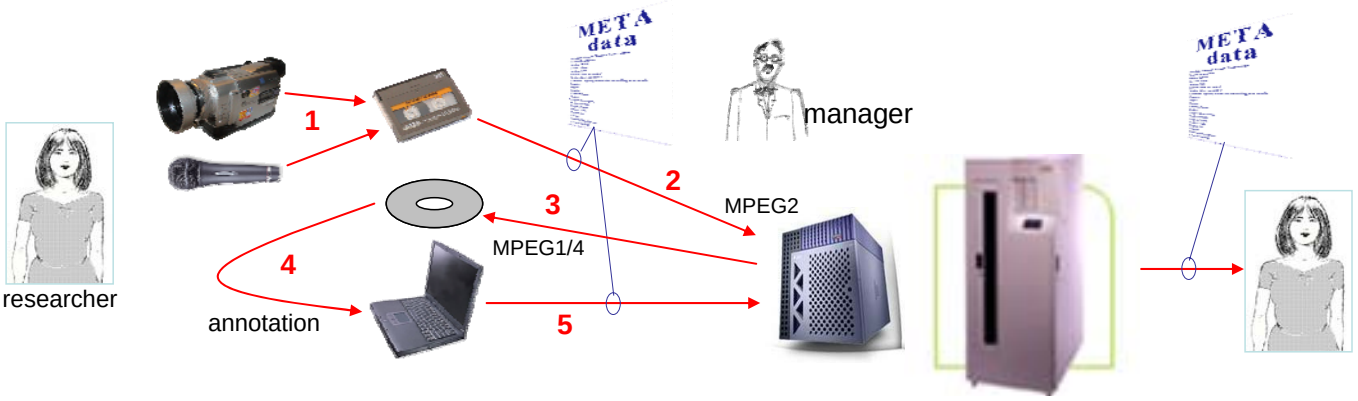
- Field researchers are creating many video recordings in the field. Often, but not all the times, digitization is done in the field as well to immediately create transcriptions, translations and other types of annotations. The workflow determines the interaction pattern that is necessary to create proper recordings, annotations and metadata descriptions and to ingest them into the archive.

1. Archive Ingestion

Until now archive managers carried out the ingestion of resources and took care that the consistency of the corpus is guaranteed as far as archivists can do it, i.e. we are not responsible for the content, the correctness of links etc. In future this role will be taken over by the LAMUS system (Language Archive Management and Upload System) which allows users to upload new versions and new resources with the help of web-based technology. LAMUS will check a number of formal aspects such as correctness of formats which will be described in the LAMUS documents. The user has (1) to specify the node in the archive hierarchy where the set of resources has to be integrated, (2) to provide correct metadata descriptions for the resources and (3) to upload formally correct resources. It is obvious that the donator has to take care that the relations between metadata descriptions and resources such as sound and annotation files is correct.

2. Central Workflow

The central workflow is the one where digitization is done centrally at the archiving institution. For video recordings this is still a major issue, since the storage requirements are extreme and since digitization of video is still a task not performed correctly by all available programs.

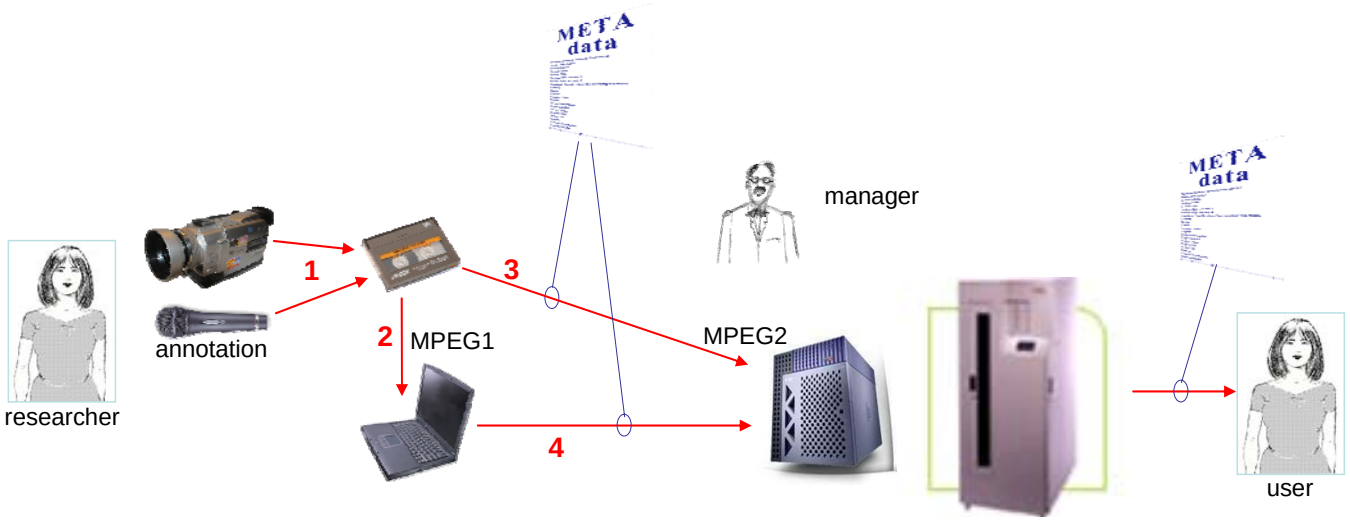


The user creates a video recording which includes audio information (1) and sends then a copy of the tape to the archive (2). The archivist creates a digital copy of the whole tape which is called a DMF and which contains an MPEG2 and a 16 bit linear PCM file containing the audio stream in high quality. From the MPEG2 version (which is the archive format) the archivist will also create an MPEG4 or MPEG1 stream, copy it on DVDROM and send that back to the donator (3). The donator will cut the material into sessions, create the annotations (4) and send back them to the archive (5). From the beginning metadata descriptions have to be provided by the donator to make sure that everyone knows exactly which resources belong together. Metadata descriptions now become very crucial, since the fieldworker will cut the material into sessions based on the MPEG1/4 stream and create the annotations time-linked with this stream. When the annotations are provided the archivist has to cut the existing MPEG2 stream at exactly the same frames to

maintain the time links with the annotations at the correct places. The correct times have to be included in the metadata descriptions so that the archivist can cut the stored MPEG2 file into the correct chunks.

3. Decentralized Workflow

The decentralized workflow has similar requirements than the central one which is due to the fact that MPEG2 cannot be dealt with in the field due to too high storage requirements.



The user creates a video recording which includes audio information (1), captures the video stream on the notebook, converts it to MPEG1 and immediately cuts it to sessions (2). With these MPEG1 files the donator can create the annotations and the metadata descriptions. When all work has been done a copy of the tape (3), the annotations and the metadata descriptions are sent to the archive (4). The archivist creates MPEG2 versions of the annotated sessions using the time code information provided in the metadata descriptions. All together is then uploaded into the archive.

4. LAMUS Usage

When LAMUS is used the donator can carry out all ingestion steps him/herself independent of the workflow if ready-made files are involved. In both the central and decentralized workflows the annotations can be uploaded once the donator has uploaded the proper metadata descriptions. It's the archivist that has to integrate the appropriate MPEG2 file which has to be done based on the proper metadata descriptions. This requires some exchange about proper file labels. Pre-assumption for uploading long media files is the availability of an Internet connection that is fast enough.