

Audio Workflow

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

General Framework:

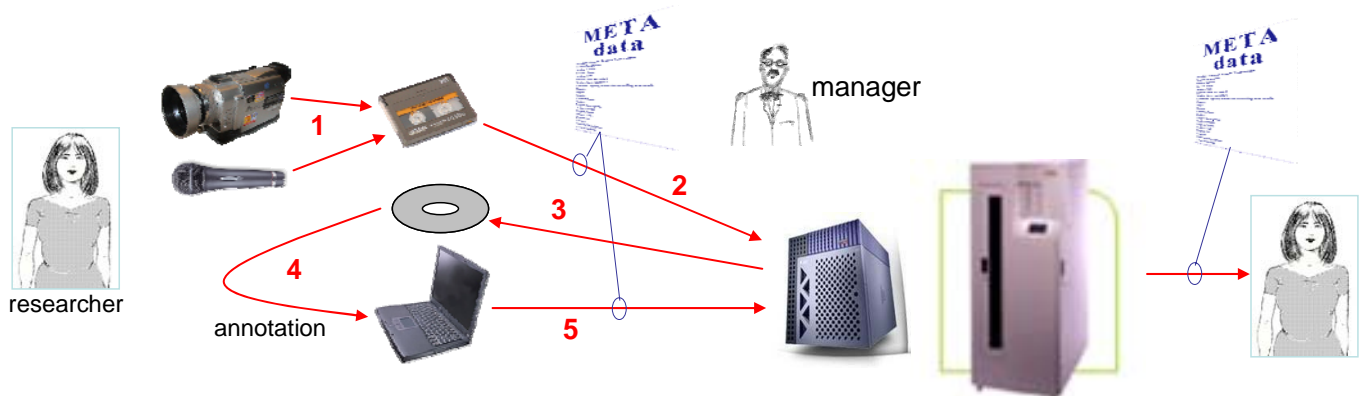
- Field researchers are creating many audio recordings in the field. Often, but not all the times, digitization is done in the field as well to immediately create transcriptions, translations and other types of annotations. The workflow determines the interaction pattern that is necessary to create proper 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 institutions. For audio recordings this is not done very often anymore, since the storage requirements are not that extreme anymore and since the digitization of audio is well-supported by most software products on the market.

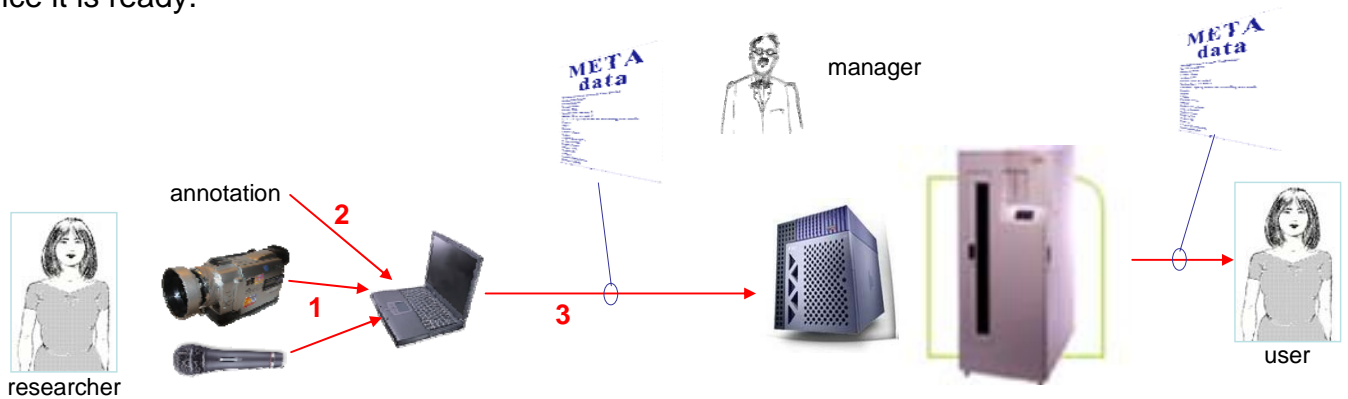


The user creates an audio recording (doing this with a video recorder will be dealt with under video) (1), send then a copy of the tape or disk to the archive (2), receives back the digitized version on CDROM/DVDROM (3), creates the annotations (4) and sends back the annotations to the archive (5). From the beginning metadata descriptions have to be provided by the donator to make sure that everyone knows exactly which resources belong together. If the whole recording will be split into several sessions (in case that there are for example several interviews on one tape) the original metadata description has to be replaced by descriptions that describe each individual session. In this case it is required to send both the audio files and the annotation files to be sure that the starting times are exactly the same.

Audio Workflow

3. Decentralized Workflow

The decentralized workflow is much more simple. It is the user who takes care of all aspects (recording, digitization, annotation, metadata description), creates the whole bundle of related resources in a workflow chosen by him/herself and provides the whole package to the archive once it is ready.



Important is that all resources that are specified in the metadata description as being the related media and annotation files are the correct ones. Only the donator can judge this.

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 the central workflow the donator should behave as if the archivist does not have any data yet, i.e. upload all files (recording, annotation, metadata) as a bundle which is exactly the case as in the decentralized workflow. This ensured that all data is synchronized properly.