Submitting a proposal to Open Text Collection

Open Text Collections publishes curated and edited text collections. All text collections are peer-reviewed. In order to propose a collection to OTC, please include the following information in your proposal:

Required information

Metadata

- Language name: give the name of the language, with common alternative names
- Language area: in which area is the language spoken?
- Contributors: who should be identifed as author(s)/editor(s) of the collection? How was this determined?
- Quality of texts: which kinds of texts will be there in your collection? How were they selected, and how do they differ from other text collections, to the extent that such collections exist?
- Quantity of texts: how many texts will there be in your collection?
- Quantity of words: Give a rough estimate of words. Given that languages have words of different complexity and use different orthographic conventions, please give the number of words in the translation line. This is a proxy to compare the sizes of different collections.
- File formats: OTC will store your files in the CLDF format. Your files are most probably not in that format yet. Please let us know in what format you intend to submit (e.g. ELAN/eaf, FLEx, tex, xlsx). Submissions in docx are discouraged, submissions in pdf are not possible.
- Conversion assistance: do you require technical assistance in integrating/aligning/conforming your data? We can help with global replacements and similar operations to ensure consistency across texts.
- Archive: where can users find the primary data on which your texts are based (eg audio files, transcripts, facsimiles). If no such information can be given, please explain why.
- Collection procedure: how were the texts collected? How were the narrators/community informed about the intended uses (open access) and how did they express their agreement?
- Curation procedure:Which criteria were applied to include certain texts and exclude certain others?
- Editing procedure: What editing procedures were applied to arrive at the written form of the texts, i.e. how does the written form differ from the primary data? Common operations include: removal of hesitation markers, false starts, self-repairs; editing out back-channels by listeners; cutting out material irrelevant to the narrative (e.g. interruptions by random people); "language polishing", where certains words or constructions are replaced at the request of the original narrator ("Here, I said X, but this was a mistake, I should have said Y").

Contextualization

- Linguistic: where is the language spoken, by how many people, in which setting? What is the genealogical affiliation and what are the contact languages?
- Anthropological/sociological: how is the community structured? Which types of settlements/housing/food/work/beliefs/rituals/political organization do we find?
- Philological: is there a narrative tradition this collection relates to? Do we know of the narrative traditions of neighbouring/related communities and how they are similar/different? Can some general patterns be observed (e.g. trickster stories)
- Representative file: please send one representative file along with your proposal so that we can evaluate the technical quality of your work

Next steps

- if the proposal is accepted:
- send all files
- among the files sent, some will be reviewed in their entirety, while for others, a sample of sentences to be reviewed will be drawn

Evaluation criteria

- Coherence: is this a random set of narratives, or can the readers see what ties them together?
- Consistency: are the texts edited in a consistent fashion with regard to orthographic conventions, glossing, translation?
- Contribution to research and knowledge: in what sense does this collection provide novel data or insights?
- Accuracy: are the glosses and translations factually correct?
- Ethics: were the texts collected according to ethical standards?
- Presentation: how good is the contextualization of the texts?
- "Literary quality": this is a very vague criterion. The texts should have some intrinsic interest which engages readers/listeners. They should not be boring.