

TEI Guidelines in practice:

Using the **ISO Standard for Transcription of Spoken Language** in the Project **INEL**

Anne Ferger, Daniel Jettka



[@ProjektInel](https://twitter.com/ProjektInel)



Project INEL

- „*Grammatical Descriptions, Corpora and Language Technology for Indigenous Northern Eurasian Languages*“
(<https://inel.corpora.uni-hamburg.de/>)
- long-term research project (2016-2033)
- in the context of the [Academies' Programme](#), which is coordinated by the [Union of the German Academies of Sciences and Humanities](#)

ISO standard for Transcription of spoken language

- P5 guidelines <https://tei-c.org/guidelines/p5/> of Text Encoding Initiative <https://tei-c.org>
- standard ISO 24624:2016 Language resource management – **Transcription of spoken language**

ISO/TC 37/SC 4. 2016. Language resource management – Transcription of spoken language. Standard ISO24624:2016, International Organization for Standardization, Geneva, CH.

Challenge

Creating, storing and using data in a sustainable way

Creation of data: limited to certain formats because of tools

Sustainability: using standards (ISO standard TEI)

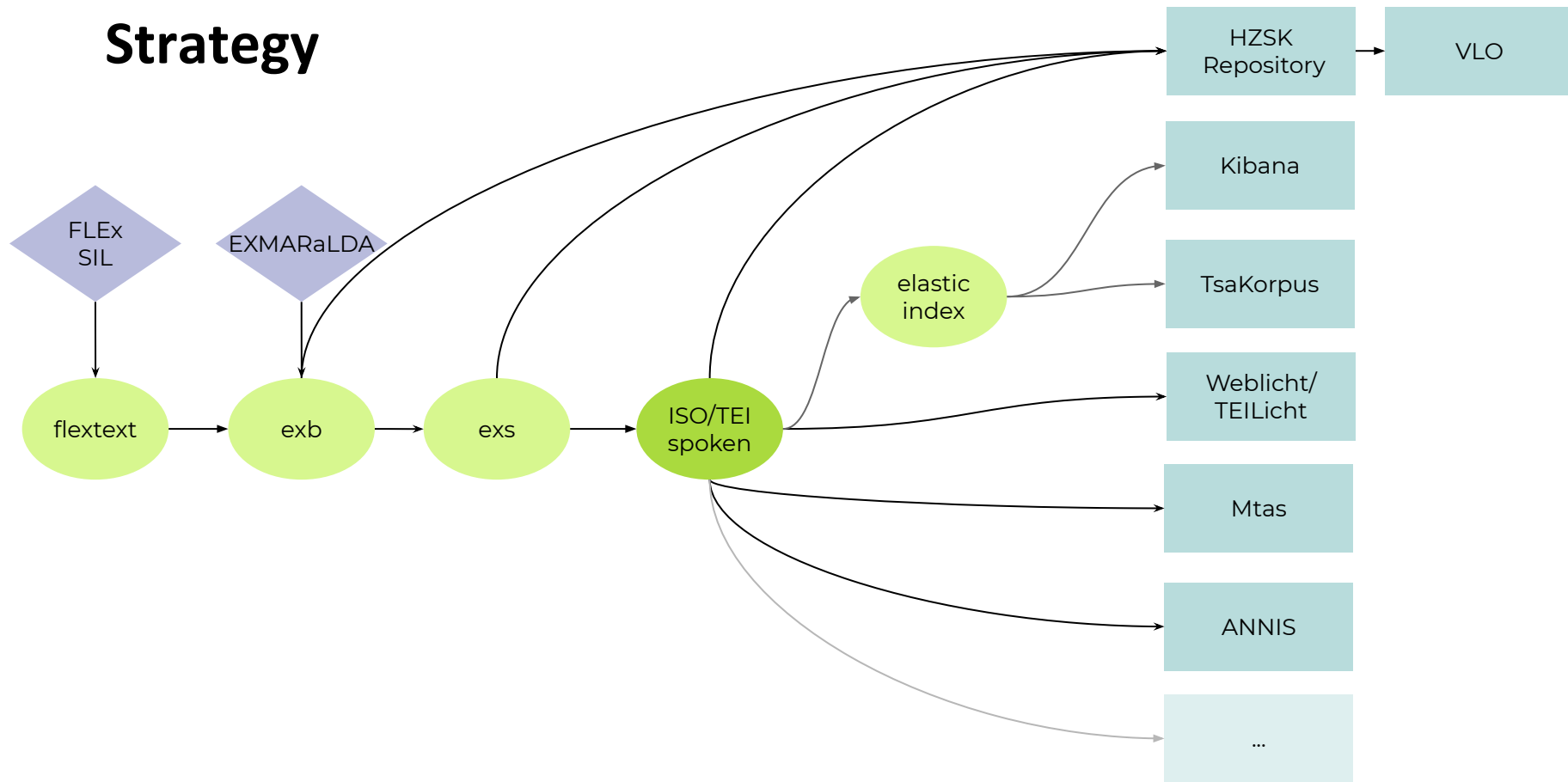
Using the standard: problems with applicability and modelling

- Missing: tool that creates ISO standard TEI in a convenient way
- Missing: tool that models the structure of the data created in the project appropriately

Using ISO/TEI-Spoken

- Adapting TEI ISO standard format and making it usable in different **tools** and **applications**
- Converting created data into the standard format
- Using or referencing ISO/TEI-Spoken in
 - Dissemination: [VLO](#), digital [repository](#)
 - Processing: [Weblicht](#) / [TEILicht](#)
 - Search: [Tsakorpus](#)
ANNIS/Mtas (future work)

Strategy



Strategy

- Accepting the difference between formats/tools during data creation and dissemination after work is finished
- Finding more use cases for the format
- Collaborating with projects using the same format and facing the same challenges