A workflow for creating, analysing, and storingmulti-layer corpora:

Stephan Druskat, Thomas Krause, Carolin Odebrecht, Florian Zipser

Friedrich-Schiller-Universität Jena, Humboldt-Universität zu Berlin

corpus documentation task

- to document a corpus is a lot of work and usually pretty time consuming. You need:
  - a full list of word forms and lemmata
  - to state all annotation layer and annotation values for each document and the entire corpus
  - to know how often annotation values were annotated
- to find errors of annotations (semi-) autmatically
  - to find typos in a large corpus
  - to find tags, that were documented but never used in a corpus and vice versa
  - to document all formats of your corpus
  - to create consistent documentations for all formats
  - to create documentations of the different formattings