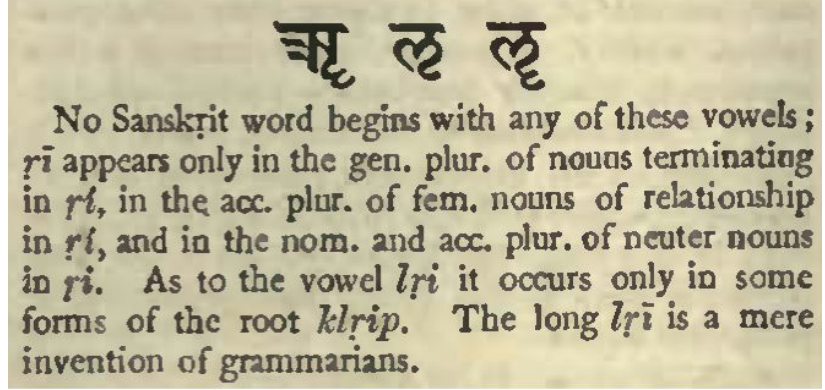


# What I learned as a field linguist helping with the Cologne Digital Sanskrit Dictionaries

Felix Rau (University of Cologne)

Although it is a retro digitisation project for lexical resources in a major classical language, lessons from the last 19 years of the Cologne Digital Sanskrit Dictionaries can inform our treatment of lexical resources from language documentation and linguistic fieldwork.



## Conclusions

1. Data is stable, formats less so, applications are short-lived
  - a. The applications of the CDSL has changed several times as have the formats, but the underlying data is relatively stable (carefully enhanced)
2. The format doesn't matter (as long as it is well documented)
  - a. The transformation from the older XML to TEI is in itself not important
3. A consistent (simple) data model is/would be an enormous advantage
  - a. The inconsistencies in the data model, categories, structures were the main issue for a unified interface and query
4. Implicit hierarchies are evil.
  - a. Implicit hierarchies are the main obstacle in creating a better markup for the CDSL
5. A good dictionary does not necessarily make a good lexical database
  - a. Despite being excellent dictionaries, they are not easily transformed into an

## Links

Website: <http://www.sanskrit-lexicon.uni-koeln.de/>  
Code and Data: <https://github.com/sanskrit-lexicon/>  
LAZARUS-Project: <http://www.cceh.uni-koeln.de/lazarus>  
Slides: <http://tr.im/dictionaria>  
Contact: [f.rau@uni-koeln.de](mailto:f.rau@uni-koeln.de)