

Glottolog/Langdoc

Providing references for the world's lesser known languages as Linked Data

Sebastian Nordhoff
Max-Planck-Institute for Evolutionary Anthropology, Leipzig

OKCon 2012, June 30, Berlin

Scope of Glottolog/Langdoc

- ▶ coverage of descriptive literature of the worlds languages
- ▶ dictionaries
- ▶ grammars
- ▶ texts
- ▶ collation of existing literature databases

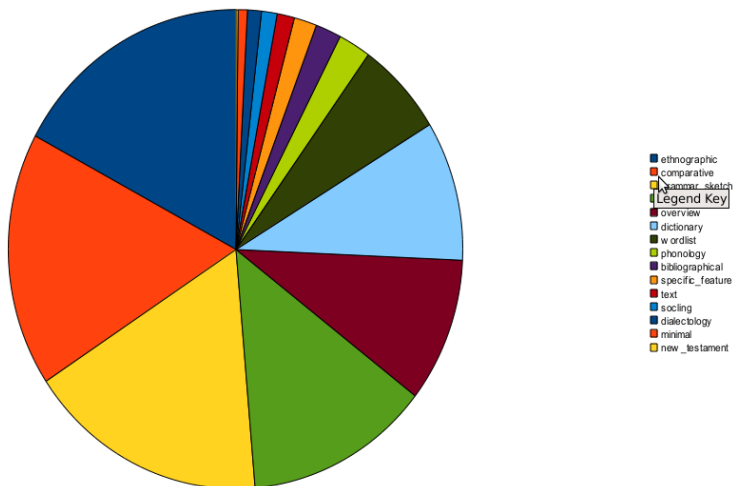
Scope of Glottolog/Langdoc

- ▶ what glottolog does not do
 - ▶ no structural information about lgs
 - ▶ no lexical information about language
 - ▶ no actual documents
- ▶ what glottolog does provide
 - ▶ references
 - ▶ languoids
 - ▶ dialects
 - ▶ languages
 - ▶ families

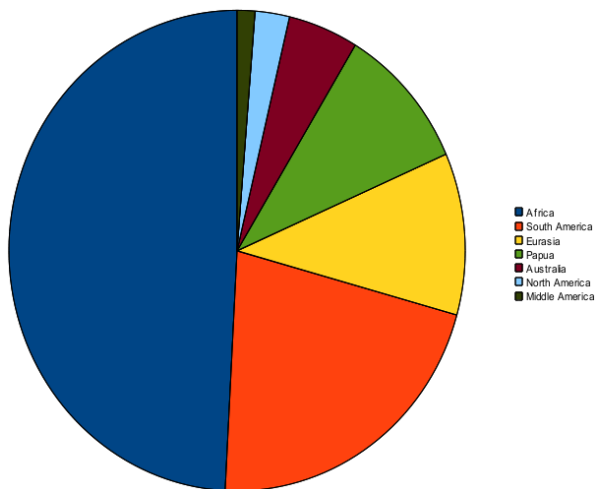
Some stats

- ▶ 61 488 languoids in 103 classifications
- ▶ 193 338 references, of which 161388 distinct
- ▶ 110 914 references assigned to one or more languoids
- ▶ 5 405 languoids with at least one reference
 - ▶ most references treat languages, not dialects or families, hence the low number
 - ▶ there are about 7 100 languages with an ISO-Code
 - ▶ 76% coverage
- ▶ 138458 total links between languoids and references

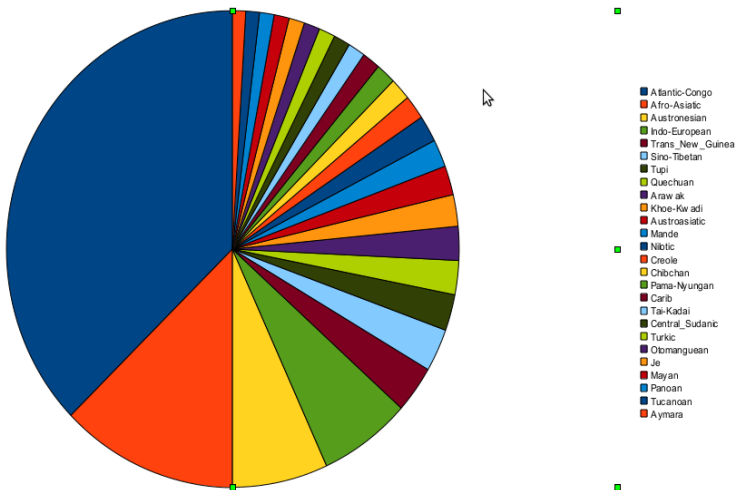
Document types



Areas



Language Families



Use cases

- ▶ find references
 - ▶ Give me all dictionaries of any Chadic language
- ▶ draw samples
 - ▶ Give me a sample of grammars which is stratified for continent and language families
- ▶ evaluate documentation status of families
 - ▶ What is the average number of works per language in a family? Which families are in most need of documentation before they die out.

References

- ▶ data about references
 - ▶ document type
 - ▶ languoid
- ▶ links
 - ▶ GoogleBooks
 - ▶ WorldCat

Languoids

- ▶ data about languoids:
 - ▶ references
 - ▶ supernodes
 - ▶ siblings
 - ▶ subnodes
 - ▶ author
 - ▶ names
 - ▶ codes
 - ▶ persistent ids for **all** languoids
- ▶ links: LL-Map, Multitree, Ethnologue, OLAC, ODIN, WALS, Wikipedia

Todo

- ▶ Choose semantic framework: RDF, microdata
- ▶ choose vocabs
 - ▶ refs: DC, TEI
 - ▶ languoids: GOLD, graph theoretic vocab???
- ▶ mint URIs
 - ▶ <http://glottolog.livingsources.org/languoid?id=12345>
 - ▶ <http://glottolog.livingsources.org/languoid/id/12345>
 - ▶ <http://glottolog.livingsources.org/languoid/id/12345.rdf>

Thanks

- ▶ Harald Hammarström
- ▶ Alain Fabre
- ▶ Youni Maho
- ▶ Guillaume Ségerer