

In this poster we present the CLLD (Cross-Linguistic Linked Data) project [1] - an MPG-internally funded project aimed at helping to preserve the world's language diversity heritage. The particular problem this project sets out to solve is bridging the gap between data collection and data sharing.

This problem can be framed with two observations from this year's ALT (Association of Linguistic Typology) 10 Conference [2]:

- Most talks about results reached with quantitative methods have been based on the WALS [3] dataset.
- Many linguists at ALT 10 work on and with "private" databases, often enriching or expanding the WALS dataset; but basically none of these are accessible or published.

A similar gap exists for dictionaries or wordlists for little known languages:

Collecting these is highly valued within the community, but publication via traditional channels is basically impossible because there are too few readers for such dictionaries to make publication as books worthwhile.

The CLLD project aims at solving this problem by providing sustainable, interoperable data publication structures, thereby offering incentives and lowering the barriers to publish linguistic databases. This strategy is implemented as follows:

1. The project develops a web application framework tailored for the publication of lexical or typological linguistic data - a CMS for linguistic databases [4].
2. Based on this framework, the databases of the department of linguistics at MPI EVA [5] are published. In addition, a language catalog (Glottolog [6]) is published, which allows cross-referencing language resources across distributed databases.
3. Two online database journals (one for dictionaries and one for typological databases) will be started, allowing for submission of databases which are too small to be published as stand-alone resources.
4. Data is published following Linked Data principles in order to allow for
 - uniform data access and interoperability of distributed resources at web-scale,
 - a well-defined lower boundary of service in a "graceful degradation of service" scenario (see below).
5. A strategy for "graceful degradation of service" is adopted, helping to describe service levels and requirements thereby making transitions of ownership/maintenance of services easier.

After the first year of the 4-year funding period, the following results have been obtained:

- The clld software package is used to publish the typological databases WALS, APiCS, eWAVE, PHOIBLE, AFBO as well as the lexical database WOLD and the language catalog Glottolog (see [1] and [7]); thus validating the concept and data models underlying its design.
- The process to gather initial submissions for the two database journals has been started.
- Applicability of the software and the process to linguistic resources outside the MPI EVA has been shown with the publication of the Phonetics Information Base and Lexicon (PHOIBLE).

[1] <http://clld.org>

[2] http://www.eva.mpg.de/lingua/conference/2013_ALT10/

[3] The World Atlas Of Language Structures Online. <http://wals.info>

[4] <https://github.com/clld/clld>

[5] <http://www.eva.mpg.de/linguistics/index.html>

[6] Glottolog. <http://glottolog.org>