Open Linguistics Working Group (OWLG)

Christian Chiarcos
chiarcos@uni-frankfurt.de
Open Knowledge Foundation (OKFN, http://okfn.org)

- non-profit organization
- founded in 2004
- promote open knowledge in all its forms
  - e.g., publication of government data (UK, US)
- provide infrastructural support for several working groups
OKFN Open Linguistics Working Group (OWLG)

- founded in Oct 2010 in Berlin, Germany
- open network of individuals interested in
  - linguistic resources and/or
  - their publication under open licenses
- multi-disciplinary
  - NLP/CL, typology/language documentation, SW, ...
- infrastructure
  - mailing list, web site/blog, wiki
  - http://linguistics.okfn.org
OWLG goals
(http://linguistics.okfn.org)

1. **Promote open data** in relation to language data
2. Point of **reference and support** for open linguistic data
3. **Facilitate communication** between researchers that use, distribute, or maintain open linguistic data
4. **Mediate between providers and users** of technical infrastructures
5. Build and maintain an **index of open linguistic data sources**
6. Assemble **best-practice guidelines and use cases** concerning creating, using and distributing data
7. Gather **information on legal issues**
OWLG goals
(http://linguistics.okfn.org)

1. **Promote open data** in relation to language data
2. **Point of reference and support** for open linguistic data
3. **Facilitate communication** between researchers that use, distribute, or maintain open linguistic data
4. **Mediate between providers and users** of technical infrastructures
5. Build and maintain an **index of open linguistic data sources**
6. Assemble **best-practice guidelines and use cases** concerning creating, using and distributing data
7. Gather **information on legal issues** these aspects are specifically well developed
OWLG activities

- mostly point-to-point cooperations between individual members
- regular telcos/meetings
- workshops -> building an interdisciplinary community
  - collocated with larger events of different communities
  - Linguistics Track of the OKCon, June 2011, Berlin, Germany
  - Linked Data in Linguistics
    - March 2012, Frankfurt/M., Germany
    - Sep 2013, Pisa, Italy
    - May 2014, Reykjavik, Iceland
  - MLODE-2012, Sep 2012, Leipzig, Germany
  - Linked Data in Linguistic Typology, Sep 2013, Leipzig, Germany
OWLG activities

- point-to-point cooperations between individual members
- regular telcos/meetings
- workshops -> building an interdisciplinary community
  - keeping ties with other communities & projects
  - e.g., Cyberling, W3C OntoLex, ACL SIGANN/SIGLEX
  - e.g., MPI-EVA, LOD2, LIDER, QTLeap
- joint publications and presentations
- building and maintaining the Linguistic Linked Open Data (LLOD) [sub-]cloud
a collection of linguistic resources
- published under open licenses
- as linked data
- decentralized developed and maintained
- meta data at [http://datahub.io](http://datahub.io)
  => cloud diagram
- developed as a community effort in the context of the Open Linguistics Working Group of the Open Knowledge Foundation

next:
LLOD 2011-2014
Building the Cloud: 2011
A sketch from a table napkin

- initially, we maintained a list of open or representative resources
  - in Jan 2011, we marked possible synergies
- merely a vision
  - includes non-open resources as placeholders for other resources to come
  - not physically realized
- a strong metaphor brought to a new community

http://nlp2rdf.lod2.eu/OWLG/llod/2011/01/llod.png

Closing chapter of the LDL-2012 companion volume

- **hypothetical** linking for selected data sets from NLP, SW and typology described in the book
Drafting the Cloud: LREC-2012

„draft status“
hand-crafted, including resources whose RDF conversion and linking was suggested, not yet performed at the time
Building the Cloud: MLODE-2012

- Multilingual Linked Open Data for Enterprises
  - goal: build the first instance of the LLOD cloud
  - workshop & hackathon
    - authors were encouraged to provide data
    - data conversion, metadata update at http://datahub.io

- automatically generated diagram
  - Richard Cyganiak’s converter scripts

http://sabre2012.infai.org/mlode
Building the Cloud: 2013+

- MLODE data post-proceedings
  - Special issue of the Semantic Web Journal
  - Preparation of additional data sets in the process
    - e.g., lemonUby (Eckle-Kohler et al., accepted)
- Linked Data in Linguistic Typology, Aug 2013
  - additional potential datasets
    - lexical databases of Austronesian languages
    - a database of syllable structures
- Intensified community work
Building the Cloud: Sep 2013

- More data sets not fully linked, yet
- New drawing script
  - By John McCrae & Christian Chiarcos
- Manually categorized and colored
  - GraphML
Building the Cloud: May 2014

- more data sets
- more rigid criteria
  - linked & accessible
- two-layered resource taxonomy
- this (<=) version is merely to elicit feedback
  - new diagram end of May 2014
Recent developments

- finalizing LLOD diagram revision
  - for LDL-2014, May 27th, 2014
- harmonizing linguistic resource categories
  - synchronization with MetaShare categories
- adding new resources
  - relevant LREC „Share your resources“ datasets?
- subsequently enforce further constraints on LLOD „bubbles“
  - open licenses (currently: accessible ~ LOD diagram)
  - well-formedness / meta data check