Creation of Open Data Resources: Benefits of Cooperation

Kira Kovalenko
Institute for Linguistic Studies, Russia
Austrian Centre for Digital Humanities, Austria

Eveline Wandl-Vogt
Austrian Centre for Digital Humanities, Austria
Austrian Centre for Digital Humanities, Austrian Academy of Sciences (Vienna)
established in 2015
1 department
4 working groups,
director Dr. Karlheinz Mörth
50 researchers

Institute for Linguistic Studies, Russian Academy of Sciences (St. Petersburg)
established in 1921
6 departments
director Prof. N. Kazansky
120 researchers
JOINT PROJECTS OF THE ACDH AND ILS

main aims:

- increase accessibility
- increase interoperability
- increase reusability
- enrich dictionary data
- interlink dictionary data
- create new workspaces
- open up dictionaries for research process and public curiosity

ReferentIn: K. Kovalenko, E. Wanld-Vogt
JOINT PROJECTS OF THE ACDH AND ILS

A Dictionary of Russian Dialects: infrastructure for compilers and researchers
general information

since 1965
48 volumes
> 300,000 entries
chief editor
prof. Sergey Myznikov
8 compilers
A DICTIONARY OF RUSSIAN DIALECTS

Тычьяк, м., собир. Колья, палки и т.п. для ограды, изгороди. Загородили тычьяком. Калин., 1972.
A DICTIONARY OF RUSSIAN DIALECTS

online representation

https://iling.spb.ru/vocabula/srng/srng.html

ReferentIn: K. Kovalenko, E. Wanld-Vogt

Thema: Creation of Open Data Resources...
methods & standards:

- TEI P5
- Lemon model
- Ontolex core model
  with extensions:
    - synsem
    - decomposition
    - variation and translation
    - linguistic metadata
JOINT PROJECTS OF THE ACDH AND ILS

- A Dictionary of Russian Dialects: infrastructure for compilers and researchers
- Russian Manuscript Lexicons: infrastructure for researchers
created in the middle of the 16th century
alphabet arrangement
close connection to the text
more than 9 types
from 700 up to 16 000 word entries
foreign words from Greek, Latin, Hebrew, Church Slavonic, Ruthenian, Tatar, Arab, and German origin
about 150 lexicons
important source for historical lexicography
ReferentIn: K. Kovalenko, E. Wanld-Vogt
Thema: Creation of Open Data Resources

TEI P5

- *cr_xq*: a standards-based, fully configurable publication framework for XML data
- full-text search and field-specific searches
- synoptic view of facsimile and text
- computer-assisted collation and stemma creation
- facilitates creation of various indexes of tagged information in the ingested resources (lemma list, index of translations by language etc.)

- standards: METS, FCS-SRU, currently mostly used for TEI content (e.g. Austrian Baroque Corpus ABaC:us

http://acdh.oeaw.ac.at/abacus)
JOINT PROJECTS OF THE ACDH AND ILS

- A Dictionary of Russian Dialects: infrastructure for compilers and researchers
- Russian Manuscript Lexicons: infrastructure for researchers
- Russian Plant Names in the Diachronic Aspect (11-17 cc.): database with a search engine
general information

- starts in 2017
- time frame: 11-17 cc.
- team: researchers with different backgrounds
- sources:
  - primary (manuscripts, printed books of the 11-17 cc.)
  - secondary (historical dictionaries, modern researches, etc.)
- database with a search engine
- interoperability with the Austrian plant names database (exploreAT!)
- contribution to the project Biodiversity and Linguistic Diversity (https://reconcile.eos.arz.oeaw.ac.at)
- collaborative Knowledge Discovery Environment
**RUSSIAN PLANT NAMES IN THE DIACHRONIC ASPECT**

**information structure**

- Old Russian name
- modern Russian name
- foreign etymon (if loanword)
- type of representing a foreign phytonym (translation, transcription, transliteration, loan translation, hyperon)
- Latin name
- example of the use from text/dictionary
- bibliographical information (if printed: author, title, place of publication, date, page, genre; if manuscript: genre, author, name, location, page, etc.)
- use of the plant
- symbolical meaning
INTERNATIONAL RESOURCES

EUROPEANA

ReferentIn: K. Kovalenko, E. Wanld-Vogt

Thema: Creation of Open Data Resources...
EUROPEANA

Title
Taraxacum ceratophorum (Lede.) DC.

Classifications
Type: Specimen, http://dbpedia.org/resource/Specimen
Subject: Одуванчик рожистый, Одуванчик рогатый, Одуванчик латайский, Одуванчик камарский, Одуванчик короткорезкий, Одуванчик Шамесо, Одуванчик Хрьтена, Одуванчик клещевидный, Одуванчик садовский, Одуванчик Малеза, Одуванчик короткорезкий, Одуванчик караоксинский, http://opendata.nhm-wien.ac.at/common/Names/References/scientificName/141536

Time
Creation Date:

16.12.16
ReferentIn: K. Kovalenko, E. Wanld-Vogt
Thema: Creation of Open Data Resources...
BENEFITS OF COOPERATION

- collaborative approach, establish sustainable workflows
- shared use of unified or de facto standards and infrastructures; develop new standards
- experimenting with new methods and emerging technologies
- open up new data
  - dictionaries
  - manuscripts
  - lexical data
  - cultural data
- connection of data to the global resources
- more and better results for both cooperation partners
- sharing failures, risks and furthering learning and innovation
- improving competitiveness and visibility

SHARING DIGITAL TRANSFORMATION AND INNOVATION
Thank you for your attention!