



perspectives of exploring a dialect language resource in a framework of European digital infrastructures

Amelie Dorn, Eveline Wandl-Vogt, Jack Bowers, Barbara Piringer & Melanie Seltmann

Lexicography Laboratory @ Austrian Center for Digital Humanities
Austrian Academy of Sciences (ACDH-OeAW)

1st International Congress on Sociolinguistics (ICS-1)
Budapest, Hungary

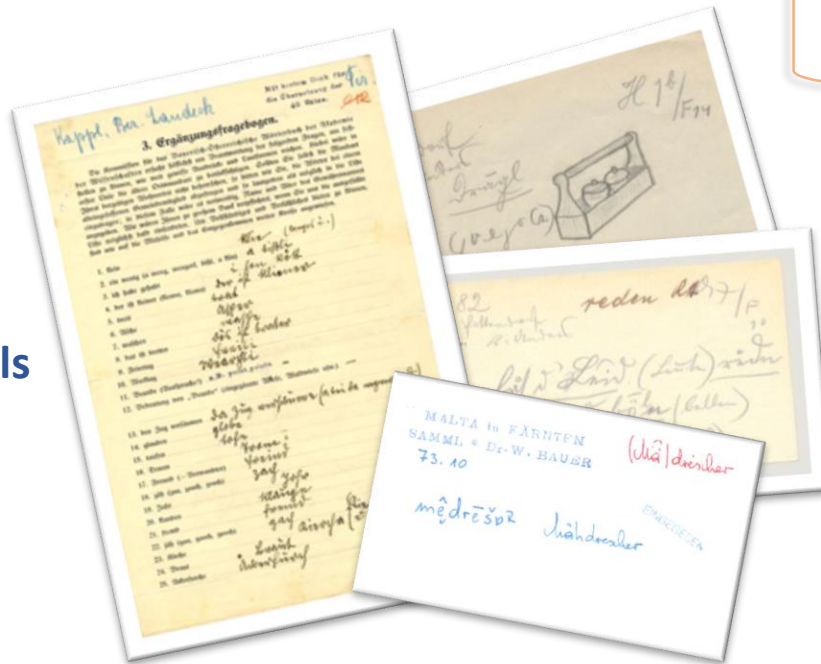


- exploring Austria's culture through the language glass

1911 - 1998

collection of the Bavarian Dialects in the region of the former Austro-Hungarian empire

Materials

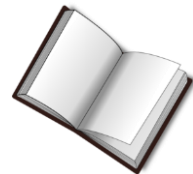


Question book

- ~ 20.000 questions
- ~ 3000 locations
- ~ 3.5 Mio records
- 100 years scientific lexicography



- exploring Austria's culture through the language glass



- **WBÖ** [Wörterbuch der Bairischen Mundarten in Österreich]
printed/digital dictionary
5 Volumes

Materials

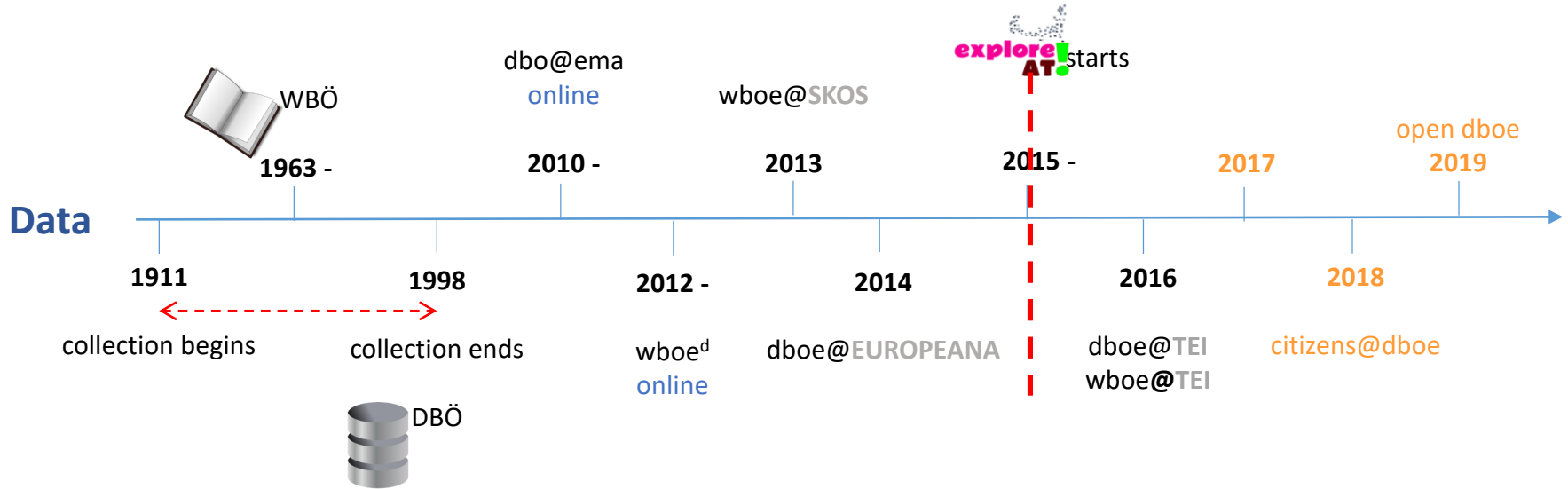


- **DBÖ** [Datenbank der Bairischen Mundarten in Österreich]
TUSTEP database ~ 2.3 Mio. single lexical data entries

- **dbo@ema**
[Datenbank der Bairischen Mundarten in Österreich
@ electronically mapped]
MySQL database ~ 200,000 entries



- exploring Austria's culture through the language glass





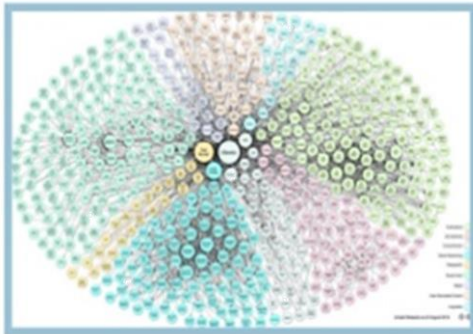
- exploring Austria's culture through the language glass

dbo@ema | wboe^donline |

	Abfrage	Projekt
	Index	
Index	Lemma	
Lemma		
Bibliographie		
Personen	Pusselein	
Orte		
Gemeinden		
Regionen		
Fragebuch		
	Details	
	<ul style="list-style-type: none"> WBÖ Lemma: Pusse Kategorie: Unbekannt Wortart: Unbekannt Wörterbücher: WBÖ 3, 1515 	

Data

wboe+dboe@LOD



wboe+dboe@TEI

TUSTEP entry

A HK 288, F288^#115.1 = F2880826.eck^#72.1

HL Fuchs:1

QU Schlagen b. Gmunden, Loitlesberger

QN {5.3a06} nōSkot. swTraunv. :OÖ *^@ Slg.LOITLESBERGER (19xx) [Slg."PIN+TierN.Gmunden"+Erg.] *O* Schlagen in Gm. Gmunden OÖ

====

NR 53C6: rot versch. Abstufg. (purpurn, der Purpur)

LT1 Fuks

BD/LT1 rotes oder rötliches Pferd

TEI (list)

TEI (fs> grammatical features)

TEI (listPlace)

XML XML

```

<item xml:id="d1e437" n="21">
  <ref type="archive">HK 288, F288^#115.1 = F2880826.eck^#72.1</ref>
  <cit type="etymon">
    <pRef>Fuchs</pRef>
    <gramGrp>
      <pos sameAs="lemma_wortat-tei-fs.xml#Sub">Sub</pos>
    </gramGrp>
  </cit>
  <ref type="source">Schlagen b. Gmunden, Loitlesberger</ref>
  <bibl>
    <ref>5.3a06</ref> nōSkgt. swTraunv.:OÖ
    Slg.LOITLESBERGER: (<date>19xx</date>)
    [Slg."PIN+TierN.Gmunden"+Erg.]</bibl>
    <placeName>Schlagen in Gm. Gmunden OÖ</placeName>
  <ref type="questionnaire">53C6: rot versch. Abstufg. (purpurn, der Purpur)</ref>
  <cit type="lautung" xml:id="d1e451" n="1">
    <pRef>Fuks</pRef>
  </cit>
  <cit type="translation" corresp="#d1e451">
    <oRef xml:lang="de">rotes oder rötliches Pferd</oRef>
  </cit>
</item>
  
```

BASIC XML

```

<record n="21">
  <A>HK 288, F288^#115.1 = F2880826.eck^#72.1</A>
  <HL>Fuchs:1</HL>
  <QU>Schlagen b. Gmunden, Loitlesberger</QU>
  <QN>
    <bibl>{5.3a06} nōSkgt. swTraunv.:OÖ *^@
    Slg.LOITLESBERGER: (19xx)
    [Slg."PIN+TierN.Gmunden"+Erg.]</bibl>
    <O>Schlagen in Gm. Gmunden OÖ</O>
  </QN>
  <NR>53C6: rot versch. Abstufg. (purpurn, der Purpur)</NR>
  <LT1>Fuks</LT1>
  <BD_LT1>rotes oder rötliches Pferd</BD_LT1>
  <orig>...</orig>
</record>
  
```



- exploring Austria's culture through the language glass

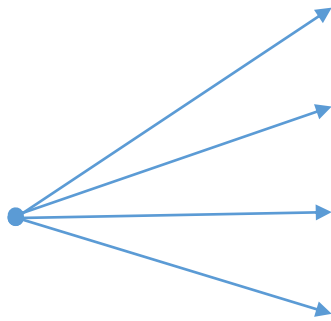
Focal topics

- plant names
- colours and patterns
- bread and pastries



- exploring Austria's culture through the language glass

Core themes



- e-lexicography
- visual analysis tools
- citizen science
- open science



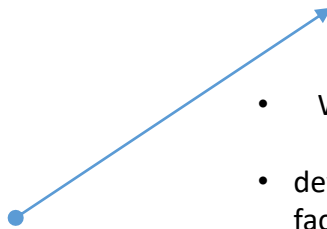
- exploring Austria's culture through the language glass



• e-lexicography

- WG4: LEXICOGRAPHY AND LEXICOLOGY FROM A PAN-EUROPEAN PERSPECTIVE
- develop a substantive theoretical insight from a European perspective into the conditions and factors surrounding **lexical variation**.
- develop a **framework** which allows for a 'neutral', i.e. non-stigmatizing, description of linguistic variants in dictionaries.
- examine the inclusion of frequency information of lexical variants in onomasiological dictionaries and its methods of presentation to the user.
- classify the diverse types and frequency of lexical variation found at phrase-level across common European languages.

Core themes





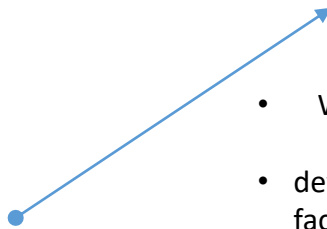
- exploring Austria's culture through the language glass



• e-lexicography

- WG4: LEXICOGRAPHY AND LEXICOLOGY FROM A PAN-EUROPEAN PERSPECTIVE
- develop a substantive theoretical insight from a European perspective into the conditions and factors surrounding **lexical variation**.
- develop a **framework** which allows for a 'neutral', i.e. non-stigmatizing, description of linguistic variants in dictionaries.
- examine the inclusion of frequency information of **lexical variants in onomasiological dictionaries** and its methods of presentation to the user.
- classify the diverse types and frequency of lexical variation found at phrase-level **across common European languages**.

Core themes





- exploring Austria's culture through the language glass

Core themes



• **visual analysis tools**



different forms of visual exploration of data

lemma map → geographical hits for specific headwords
 community graph
 relation plot
 timeline



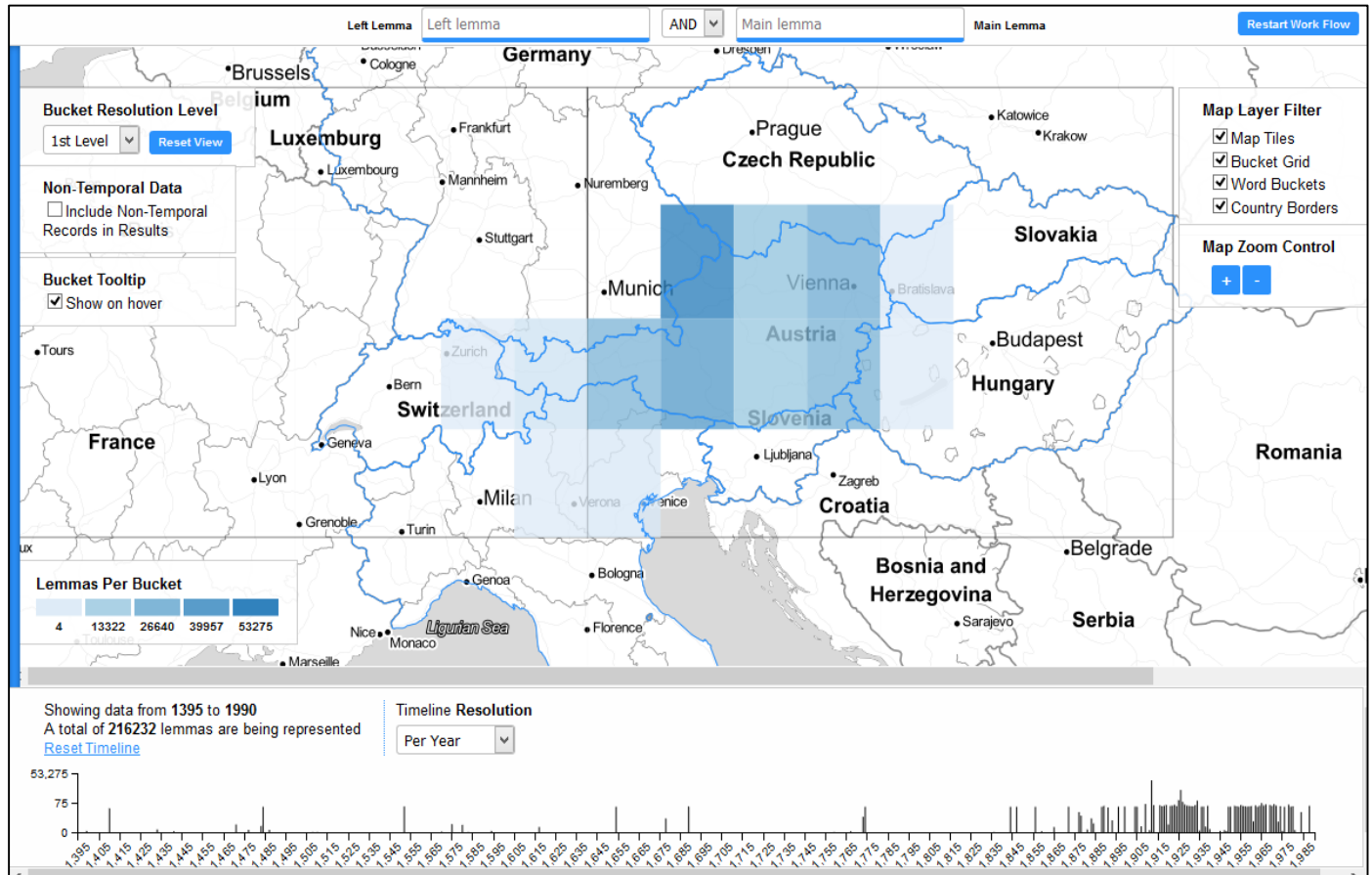
- exploring Austria's culture through the language glass



UNIVERSIDAD SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

Core themes





- exploring Austria's culture through the language glass



VNIVERSIDAD
B SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

Core themes

Left Lemma AND Main Lemma [Restart Work Flow](#)

Showing community graph for the selected bucket

31 members

Toggle Tree Showing relations for "gold" as main lemma

dukäten (1) ○ — 1 Relation — ○ gold
fein (1) ○

Showing data from 1909 to 1967
A total of 289 lemmas are being represented
[Reset Timeline](#)

Timeline Resolution: Per Year

Lemmas Per Bucket

1	9	16	24	31
---	---	----	----	----

Map Layer Filter:
 Map Tiles
 Bucket Grid
 Country Borders
 Word Buckets

Map Zoom Control: + -

Bucket's Top Main Lemmas:
 1. grufo
 2. gelb
 3. faden
 4. wäge

* The list may include lemmas without registered year



- exploring Austria's culture through the language glass



socientize
citizen science projects



ibercivis



Core themes



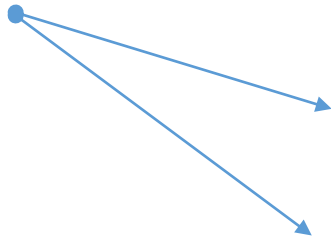
• **citizen science**

- citizen science (CS) survey
- development of specialised focus groups for the 3 focal topics
- CS bread tool -



- exploring Austria's culture through the language glass

Core themes



- **social innovation**
- **open science**



- exploring Austria's culture through the language glass



APIS [Austrian Prosopographical Information System]

geo-tagged place names

Linked projects



BioLing

plant names





- exploring Austria's culture through the language glass



include linked knowledge from

European and global infrastructures →

embed in existing infrastructures



Key objectives

Semantic search engine





- exploring Austria's culture through the language glass

Key steps

- Web-based, collaborative, multilingual, **domain-tuned IT infrastructure**
- exploitation and linking of data through **semantic web**
- **archiving, editing, analysing** non-standard data and lexicographic output
- facilitate **scholarly and amateur inquiry**
- explore the data with **interactive visual tools**

- interoperability
- standards
- workflows





- exploring Austria's culture through the language glass

Thank you for your attention!



- exploring Austria's culture through the language glass

References

- Linked Open Data (LOD) cloud: lod-cloud.net
- Cliparts: www.clker.com
- Data, data modelling: Eveline Wandl-Vogt
- Interactive visual tools: Roberto Theron