

Lexicographic workflow

Metadata in the DWDS

Axel Herold

Berlin-Brandenburgische Akademie der Wissenschaften

February 9th, 2021

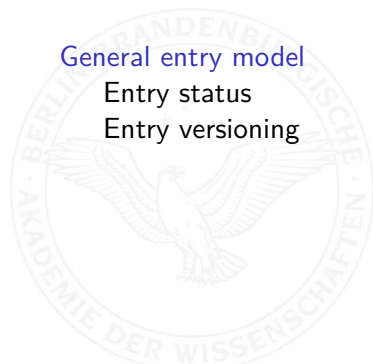


The DWDS – a lexical information system

General entry model

Entry status

Entry versioning



What is a lexical information system?

- ▶ representation of *lexical items* from different (linguistic) perspectives
- ▶ lexical items: bundles of properties (e. g. written/spoken form(s), semantics, morpho-syntax, etymology)
- ▶ relations across lexical items
- ▶ inspired by the mental lexicon
- ▶ this definition fits printed dictionaries, too!
- ▶ *digital* \equiv dynamic
- ▶ (synchronous) dictionary as core, additional resources

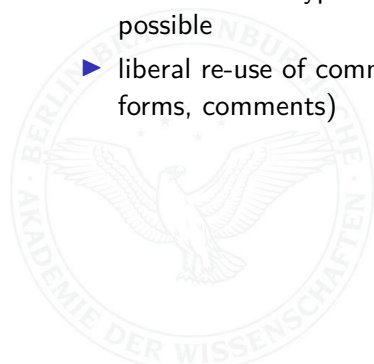
Klein (2004), Klein/Geyken (2010), Herold (2014)

Modeling framework

- ▶ historically (approx. 10 years ago): (pure) TEI representation of the „Wörterbuch der deutschen Gegenwartssprache“ (WDG, 1964–1977, see <https://www.dwds.de/d/wdg>)
- ▶ edition of the WDG by senior lexicographers (from Grimm's dictionary)
- ▶ slowly emerging target entry model (*ad hoc*, not *a priori*)
- ▶ switch to DWDS specific XML dialect:
 - ▶ swifter and unrestricted model changes
 - ▶ readability for senior staff

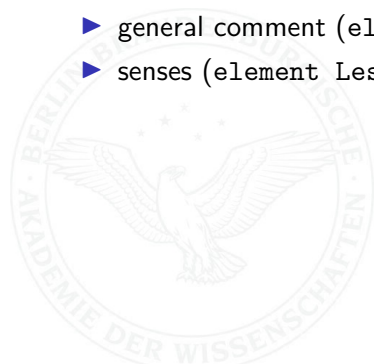
Major design decisions

- ▶ purely lexicographical view (see talk on TEI modeling)
- ▶ elements contain information to be presented directly
- ▶ attributes carry metadata
- ▶ restricted datatypes and extensional enumerations wherever possible
- ▶ liberal re-use of common structures (such as usage labels, forms, comments)



Entries

- ▶ metadata (attribute ...)
- ▶ forms (element Formangabe +)
- ▶ morphology (element morphologische_Verweise +)
- ▶ diachronic information (element Diachronie)
- ▶ general comment (element Kommentar ?)
- ▶ senses (element Lesartenangabe +)



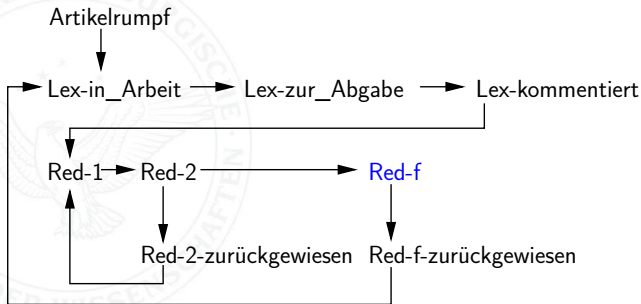
Entry status

- ▶ `//Artikel/@Status` models the editorial workflow
- ▶ cooperative approach

`Artikelrumpf` – basic entry template

`Lex-*` – stages in individual editing

`Red-*` – stages during clearance and publication



Entry versioning

two possible versioning schemes:

1. entry database (gitea) in the core dictionary writing system
 - ▶ versioning of *all* (internal) versions of an entry
 - ▶ possibly lots of „boring“ versions due to scripted editing and/or minimal changes
 - ▶ possibly extremely long changelogs
2. production database (MySQL-db) of the website backend
 - ▶ versioning of published entries only
 - ▶ much shorter and concise changelogs

For our current bibliography see

<https://www.dwds.de/d/publikationen>