

# Hidden Markup – The Digital Work Environment of the “Digital Dictionary of Surnames in Germany”

Franziska Horn, Jörg Hambuch, Sandra Reker

Balisage: The Markup Conference

2-5 August 2016





# English surnames published in the Dictionary



English surnames published in the DFD (version from July, the 27<sup>th</sup> 2016), word cloud is based on frequencies (logarithmized).





# Outline



- Project Presentation
- General Description of the Work Environment Onodi
- New Onodi Features in Detail
  - Copying of UUIDs
  - Managing Bibliographical References
  - Publishing Workflow
- Conclusion





# Project Presentation



## Partners and Aim

- Academy of Sciences and Literature Mainz, Technische Universität Darmstadt and Johannes Gutenberg University Mainz
- preparation of the first-ever comprehensive digital dictionary of surnames in Germany

Data basis with  
all 850 000  
surnames

2012



DFD with 200 000  
published articles

2036





# Project Presentation



## Features of DFD Articles

- etymology and origin
- geographical distribution of the surname
- the occurrence of the name in countries other than Germany
- morphological or semantic variants
- further reading on the topic

The screenshot shows the website 'namenforschung.net' with the title 'DIGITALES FAMILIENNAMENWÖRTERBUCH DEUTSCHLANDS (DFD)'. Below the title is a navigation bar with links: WÖRTERBUCH, PROJEKTVORSTELLUNG, MITARBEITER/INNEN, and AKTUELLES. The main content area is titled 'Miller' and includes a search result link 'Zurück zum Suchergebnis'. Below the title are tabs for 'Allgemeines', 'Etymologie', 'Verbreitung', 'Verwandte Artikel', 'Literatur', and 'Metadaten'. The 'Allgemeines' tab is active, showing a table with the following data:

Allgemeines	
Häufigkeit	7268
Rang	312
Sprachvorkommen	deutsch englisch
Herkunftsland	Deutschland Großbritannien USA

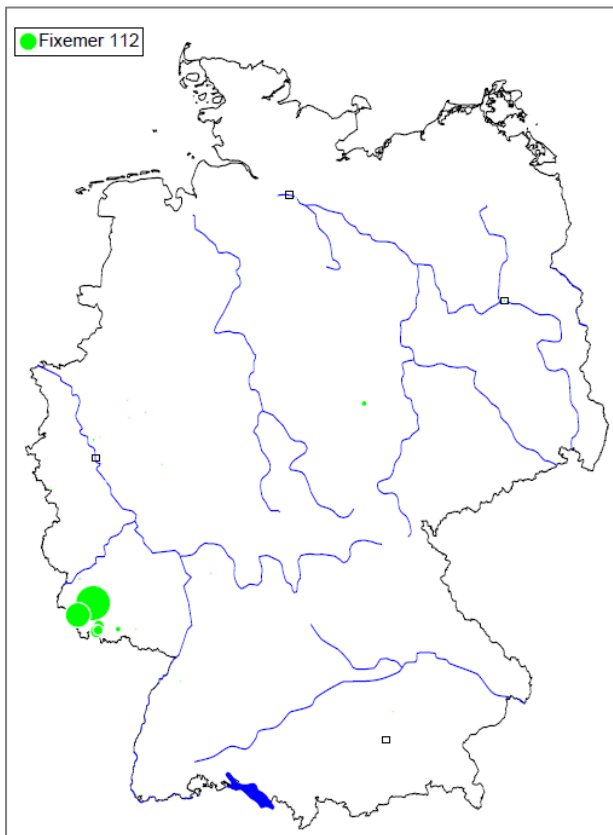
To the right of the table, under 'Kategorien', it says 'zu Bedeutung 1:' and lists categories: 'Beruf', 'Nahrungsmittelherstellung', and 'Müller'. Below this, under 'Glossareinträge', it shows '► Entrundung'.



# Project Presentation



## Mapping Surnames





# Project Presentation

---



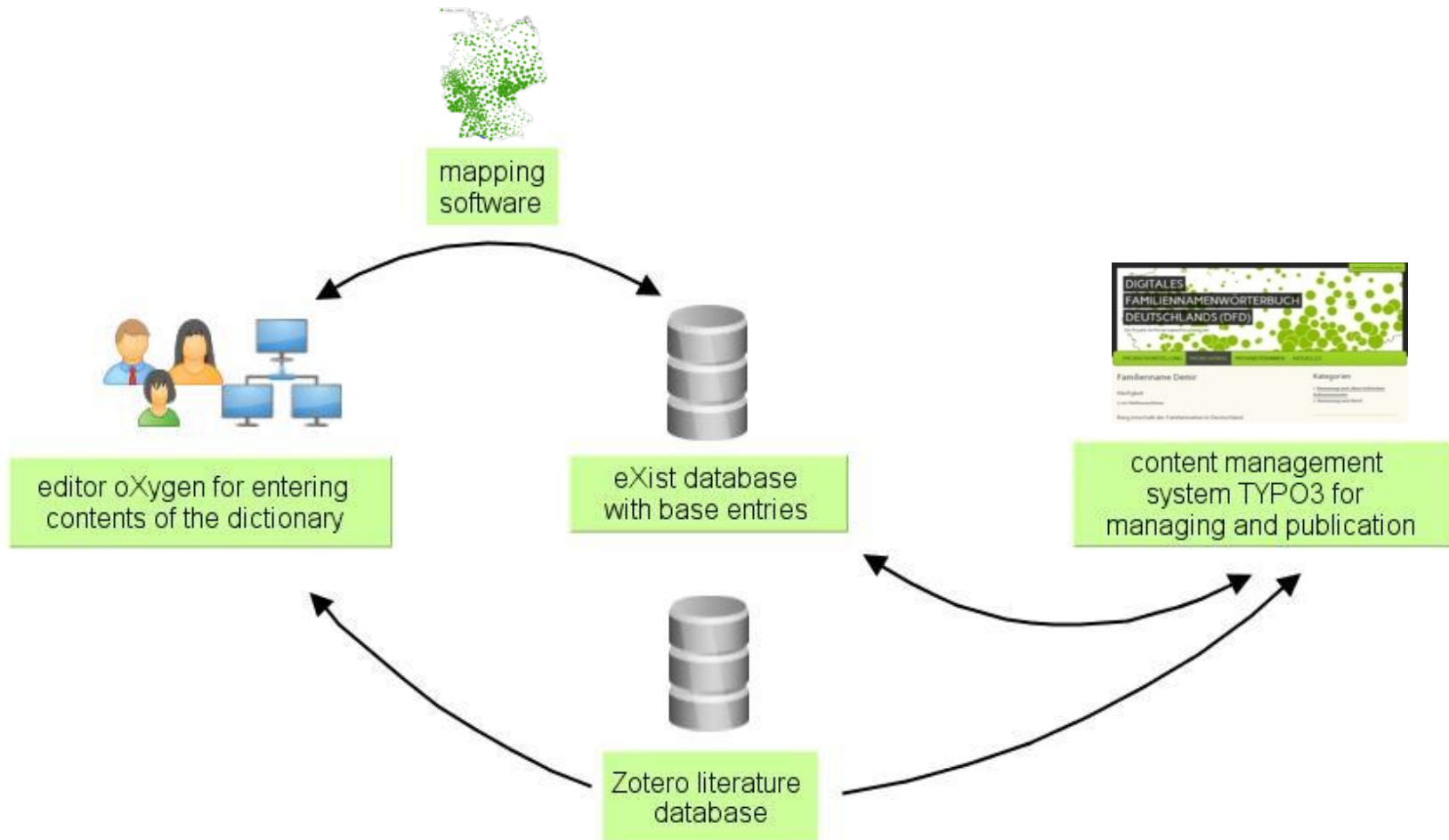
## Relevance

- shortcomings of previous surname dictionaries
- resource for other research disciplines (e.g. dialectology, migration research)





# Work Environment Onodi





# User Interfaces for Editing Contents in the CMS



The screenshot displays the TYPO3 CMS backend interface. On the left, a sidebar contains navigation menus for 'WEB' (Seite, Liste, Papierkorb, News), 'DATEI' (Dateiliste), and 'BENUTZERWERKZEUGE' (Einstellungen). The main content area shows a file tree for 'namenforschung.ne' with folders like 'Testseiten', 'Services', 'Projekte', 'IGL', and 'Projekt-Sit'. The right-hand panel is the 'Überschrift' (Title) editor, showing the title 'Surnames' and the subtitle 'Familiennamen'. Below this, the 'Typ' (Type) is set to 'Kapitelüberschrift (H2)' and the 'Layout' is 'Layout 2'. The 'Text' editor is active, showing a rich text editor with a toolbar and a text area containing the following paragraph:

Names play a significant role in defining personal identity: Not only given names, but also surnames are perceived as integral parts of individuals' personalities. The evolution of German surnames goes back to medieval times, when they developed from initially impermanent by-names. Surnames preserve linguistic material, cultural conditions and mental attitudes, the whole range of professions and also important clues about settlement and migratory movement in the late Middle Ages. This makes them a valuable source not only for historical linguistics, but also for cultural anthropology, mentality research, religious studies as well as research on migration and settlement. Accordingly, there is a strong interest from the public and the scientific community in the meanings, distribution and origins of names.



# User Interfaces for Editing Contents in the XML Editor



Gaa.xml [http://denzer@webdav.namenforschung.net/dfd/names/g/ga-gänßin/Gaa.xml] - <oxygen/> XML Editor (Ausschließlich akademische Nutzung)

File Bearbeiten Suchen Projekt Optionen Werk **a)** TEI-DFD Dokument Fenster Hilfe

neue öffnende spitze Klammer Strg+K  
 neue schließende spitze Klammer  
 neue etymologische Angabe Alt+E  
 neue Sprachangabe  
 neue metasprachliche Angabe Strg+M  
 neue Übersetzung Strg+G  
 neuer Verweis Strg+L  
 neue BedeutungsID  
 neuer Personennamen  
 neuer Familienname  
 neuer Rufname  
 neues Etymon  
 Zitieren  
 Geographische Angaben  
 Grammatische Angaben  
 Listen  
 Erneut laden  
 Tags-Anzeigemodus  
 Profiling/Bedingter Text

**b)** XPath 2.0

RN Top < > e E m 'g' lg link vgl "q"

Gaa.xml x

TEI text body entry sense sense

Vorschau der Websicht ansehen  
 Artikel-ID: e9d03651-b779-4359-a548-7127878b5e61

Artikelname: Gaa  
 Bearbeiter (Name, Vorname):  
 Überprüft durch (Name, Vorname):

Kategorien-Übersicht ansehen  
 Kategorien:  
 neue Kategorie **c)**  
 Kategorie: Bezug zu Bedeutungs-Nr.: Bedeutungs-ID:  
 neue Kategorie

Bewertung des Artikels: **d)** neue Bewertung  
 Neubewertung  
 Präzisierung der Deutung  
 Erstaufnahme  
 Neudeutung  
 Erstdeutung  
 Zusammenführung

Text Raster Autor

**e)**

Gliederung  
 Elementnamen-Filter  
 TEI "ca483c46-cb10-4d1c-9bac-8678668d"  
 teiHeader "entry"  
 fileDesc Gaa  
 profileDesc Abfrage von Komposit  
 textClass  
 keywords "categories\_DFI"  
 keywords "rating\_scale"  
 term  
 creation Abfrage von Komposit  
 revisionDesc Erstellung des Grund  
 text  
 body  
 entry "not\_specified"  
 form Gaa  
 usg "token" 291  
 usg "rank" 12807  
 etym  
 sense "primary"  
 sense "1"  
 sense "uncommon"  
 sense "uncertain"  
 sense "obsolete"  
 cit "historical\_evidence"  
 usg "norman\_distribution"

Datenquellen Explorer  
 Verbindungen  
 webdav.namenforschung.net  
 css  
 dfd **g)**  
 names  
 a  
 b  
 c  
 d  
 e  
 f  
 g  
 ga-gänßin  
 ga-gänßin  
 Ga.xml  
 Gaa.xml  
 Gaab.xml  
 Gaabs.xml  
 Gaach.xml  
 Gaack.xml  
 Gaad.xml  
 Gaadanvi.xml  
 Gaadaoui.xml



# Editing Interface and XML Code



Haupt- oder Verweisartikel?

**Familienname: Ferraro**

Formbeschreibung des Appellativs:

Häufigkeit: 208  
Rang innerhalb der Familiennamen in Deutschland: 18215

**Herkunft und Bedeutung**

Herkunftssprache:

Herkunftsland:

Herkunftsregion:

[Regionenliste ansehen](#)

```
<text>
  <body>
    <entry type="main">
      <form>
        <orth>Ferraro</orth>
        <form type="derivate"/>
        <pron/>
      </form>
      <usg type="token">208</usg>
      <usg type="rank">18215</usg>
      <etym>
        <lang xml:lang="it"/>
        <country key="IT"/>
        <region/>
      </etym>
    </entry>
  </body>
</text>
```



# Editing Interface and XML Code



Haupt- oder Verweisartikel?

**Familienname: Ferraro**

Formbeschreibung des Appellativs:

Häufigkeit: 208  
Rang innerhalb der Familiennamen in Deutschland: 18215

### Herkunft und Bedeutung

Herkunftssprache:

Herkunftsland:

Herkunftsregion:

[Regionenliste ansehen](#)

```
3  /* Angabe der Herkunftssprache über Eingabefeld */
4  <entry > etym > lang{
5      color:black;
6      content:
7      oxy_label(text, "Herkunftssprache: ")
8      oxy_editor(
9      type, combo,
10     edit, "@xml:lang",
11     editable, false,
12     values, "x-DFD, x-prg-o, ar, arc, ben, bs, bg, cmn, de, da, en, et,
13     labels, "keine Angabe, altpreußisch, arabisch, aramäisch, Benga
14     line-height:2;
15 }
16
17 /* Button zum Hinzufügen einer Sprache */
18
19 <entry > etym > lang[xml|lang]:after{
20     content:
21     "A"
22     oxy_editor(
23     type, button,
24     actionID, "Herkunftssprache")
25     "A";
26 }
```



# Consistent Data Input



Weblinks

neuer Weblink

URL:

Artikel *Hufner, Hufner* in: Deutsches Rechtswörterbuch

Letzter Zugriff: (  )

März 2016

Mo	Di	Mi	Do	Fr	Sa	So
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

U

A

L

neuer Weblink

neuer Weblink

URL:

Artikel *Hufe* in: Deutsches Wörterbuch

Letzter Zugriff: (  )



```
<ref target="http://drw-www.adw.uni-heidelberg.de/drw-cgi/zeige?term=H%FCfner&index=lemmata">
  <title>Artikel <mentioned>Hufner, Hufner</mentioned> in: Deutsches
    Rechtswörterbuch</title>
  <date type="last_access" when="2016-03-29"/>
</ref>
```





# Consistent Data Input



## Weblinks

neuer Weblink

URL:

Artikel *Hufner*, *Hüfner* in: Deutsches Rechtswörterbuch

Letzter Zugriff: (2016-03-29)

März2016

Mo	Di	Mi	Do	Fr	Sa	So
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

DWB?lemma=huber

Wörterbuch

neuer Weblink

URL:

Artikel *Hufe* in: Deutsches Wörterbuch

Letzter Zugriff: (2016-03-29)

```
28 /*Weblink DatePicker in xr */
29 xr > ref[target] > date:after{
30   content:
31     "("
32     oxy_editor(
33       type, datePicker,
34       edit, "@when",
35       format, "yyyy-MM-dd")
36     ") ";
37 }
```





# Flexibility and Customizability



Deutung unsicher:

neue Bedeutung

Nr.  Motiv:

Bedeutung: BedeutungsID: s\_4da44dd1-e360-4694-82a9-f6e53e686c13 Benennung nach Herkunft zum Siedlungsnamen Ferraro, heute Ortsteil von Siderno, einer italienischen Stadt in der Provinz Reggio Calabria. (siehe Caffarelli VÖ-Jahr: , Band: , Seite: )

neue Bedeutung

```
<sense expand="uncertain">
  <sense n="2" value="Herkunft">
    <anchor xml:id="s_4da44dd1-e360-4694-82a9-f6e53e686c13"/> Benennung nach
    Herkunft <etym>zum Siedlungsnamen <district>Ferraro</district>, </etym> heute
    Ortsteil von Siderno, einer italienischen Stadt in der Provinz Reggio Calabria
    <cit>
      <bibl>(siehe <author><surname>Caffarelli</surname></author><date when="2008"/>,
      <biblScope type="part">1</biblScope><biblScope type="pp">755</biblScope>)
      </bibl>
    </cit>
  </sense>
</sense>
```





# Flexibility and Customizability



Deutung unsicher:

neue Bedeutung

Nr.  Motiv:

Bedeutung: BedeutungsID: s\_4da44dd1-e360-4694-82a9-f6e53e686c13 Benennung nach Herkunft zum Siedlungsnamen Ferraro, heute Ortsteil von Siderno, einer italienischen Stadt in der Provinz Reggio Calabria. (siehe Caffarelli VÖ-Jahr: , Band: , Seite: )

neue Bedeutung

Zitieren		
Geographische Angaben		
Grammatische Angaben		
neue Schreibvariante		
neuer Verweis orthographische Form		
neuer Verweis auf Ausspracheangabe		
Erneut laden		
Tags-Anzeigemodus		
Profiling/Bedingter Text		

neue direkt zitierte Textstelle	Strg+Q
neues Kurzzitat	Alt+V
neuer Weblink für Literaturverweise	
neues Zugriffsdatum URL	Strg+L
neuer Autor	
neuer Band	Strg+B
neue Seitenangabe	
neue Spaltenangabe	
neue Kurztitelangabe	
neuer dritter Beleg	





# New Features of Onodi

---



- Copying of UUIDs
- Managing Bibliographical References
- Publishing Workflow





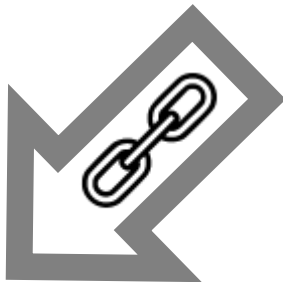
# Copying of UUIDs: The Situation



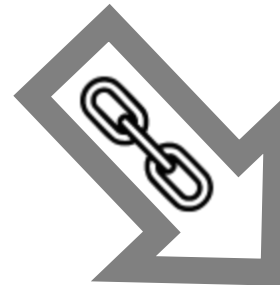
**Article**

Miller

dbbd95ff-8181-4c2d-acaa-59c46bf437a5



copy and paste  
UUID into article



**Glossary  
entry**

Entrundung  
vowel unrounded:  
Müller → Miller

fcd10016-55c7-430b-a897-33e8b16ce08b

**Thematic  
information**

English  
surnames

a2d70813-75b9-491b-a897-22e8e16ce60c



# Copying of UUIDs: The Problem



- UUID stored in `xml:id` attribute

```
xml:id="fcd10016-55c7-430b-a897-33e8b16ce08b";
```

- displayed in author view as static text

```
8 ▼ TEI:before{
9     content: "Glossar-ID: "attr(xml\:id);
10    font-weight:bold;
11    -oxy-editable:false;
12 }
```

- **problem:** in oXygen's author view not selectable





# Copying of UUIDs: A (Too) Simple Solution



- simple solution: switching to oXygen's text view
  - risk 1: users might invalidate the document by mistake
  - risk 2: selecting a UUID with the mouse is error-prone, characters in the beginning or at the end of the UUID might be omitted

```
xml:id="fcd10016-55c7-430b-a897-33e8b16ce08b">
```

```
xml:id="fcd10016-55c7-430b-a897-33e8b16ce08b">
```





# Copying of UUIDs: The Solution



- development of custom operation in Java
- class “copyAttributeValueOperation” implements oXygen’s AuthorOperation interface

```
13      //if cursor in anchor element: ID of sense element (Bedeutungs-ID)
14      ▼      if("anchor".equals(element.getLocalName())){
15              String value = element.getAttribute("xml:id").getValue();
16              copyToClipboard(value);
17
18              JOptionPane.showMessageDialog(null, "Die Bedeutungs-ID " + value +
19              " wurde in die Zwischenablage
...      kopiert.");
20      }
21      ▼      else { //if cursor somewhere else: ID of article (Artikel-ID)
22              AttrValue value =
...      authorAccess.getDocumentController().getAuthorDocumentNode().getRootElement().getAttribute
...      ("xml:id");
23              copyToClipboard(value.getValue());
24
25              JOptionPane.showMessageDialog(null, "Die Artikel-ID " +
...      value.getValue() +
26              " wurde in die Zwischenablage
...      kopiert.");
27      }
```



# Copying of UUIDs: The Solution



- copies value of `xml:id` attribute to clipboard
- “Copy ID” buttons trigger copy operation
- buttons placed with CSS next to any UUID

```
/* Copy UUID */  
TEI[xml\:id]{  
  display:block;  
  background-color:#EEE;  
  content:  
    oxy_label(text, " ")  
    oxy_editor(  
      type, button,  
      actionID, "ID kopieren");  
}
```

Entrundung.xml x

TEI

Glossar-ID: g\_0cd10016-55c7-430b-a897-33e8b16ce08b

ID kopieren

Message

Die Artikel-ID g\_0cd10016-55c7-430b-a897-33e8b16ce08b wurde in die Zwischenablage kopiert.

OK



# By the way:

## Are UUIDs valid values for `xml:id`?



- auto-generated oXygen UUID: first character not always letter
- problem: invalid values for `xml:id`
- solution now: prefix `s_`, `g_` or `i_`
- solution soon: CustomUUID plugin  
(<http://github.com/digicademy/CustomUUID>)

### Hauptbedeutung:

neue Bedeutung

Nr.  Motiv:

Bedeutung: BedeutungsID **s\_9248e084-fdc4-44a1-a29f-39cb8c9c09b9**  Benennung

nach Beruf zu () **mittelhochdeutsch** *smi*t 'Schmied'. Gemeingermanisch handelt es sich um eine Sammelbezeichnung für Handwerker, die mit scharfem Werkzeug harte Materialien bearbeiten, wie Metall, aber auch Holz. Später erfolgt die Bedeutungsspezifizierung zum



# Managing Bibliographical References



Screencast





# Managing Bibliographical References

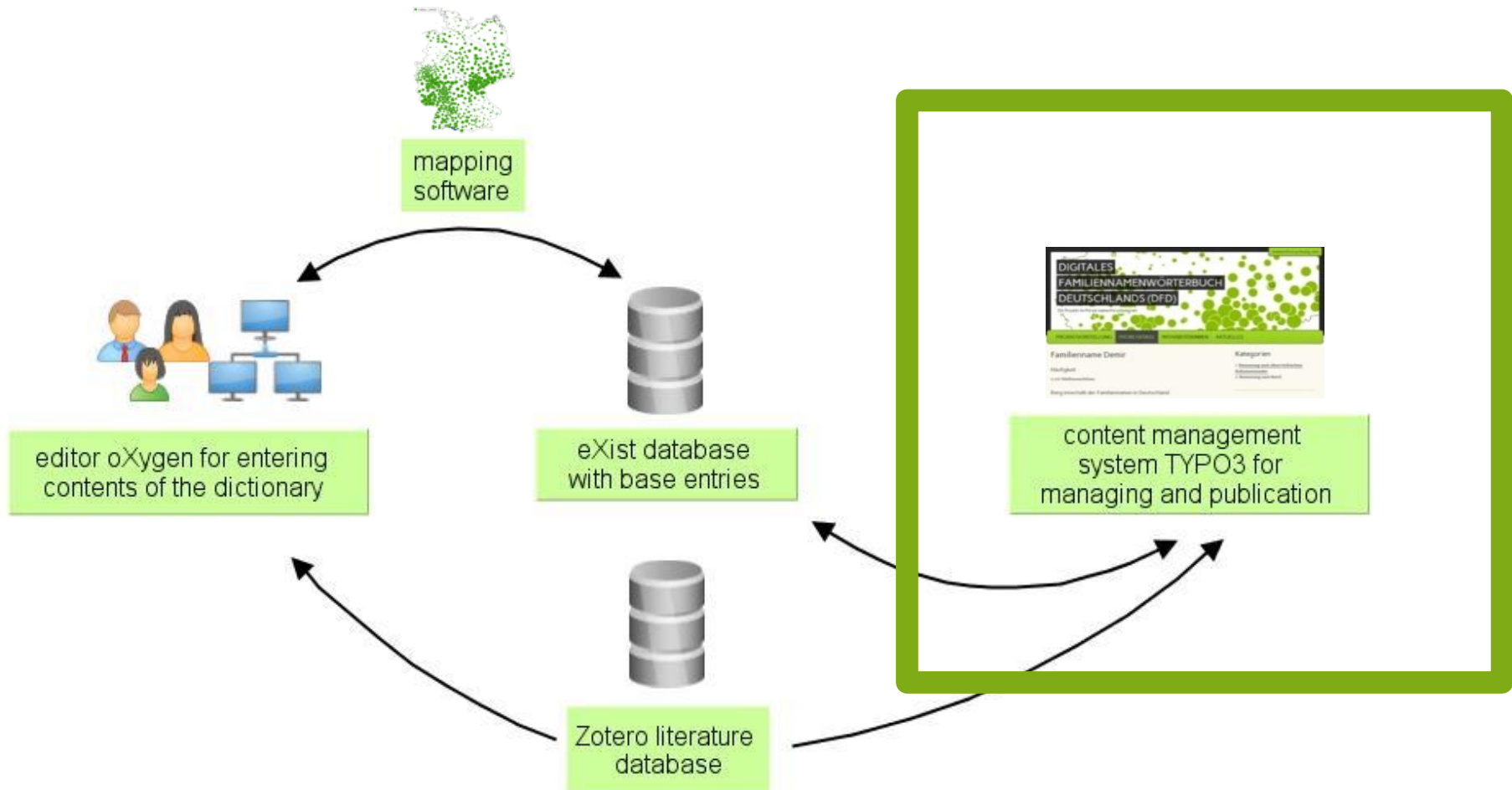


- Zotero: web-based tool to organize research sources
- eXist db scheduler calls XQuery script
- retrieves bibliographical references
- stores XML file
- eXist trigger fires on update
- creates key value list für combobox
- splits XML file per letter
- selected letter in combobox is parameter for ediarum index function





# Work Environment Onodi





# Publishing Workflow: TYPO3 CMS



- dictionary website is driven by content management system TYPO3
- TYPO3 extension “dfd”
  - data model
  - scheduler task for retrieval of documents which are ready for publication
  - storage in TYPO3 database
  - provides plugins for search, list and detail view of dictionary contents





# Publishing Workflow: Articles “Ready for Publication”



Miller.xml x

**Publikationsschritte:**

- ☒ Freigabe zur Reda
- ☒ Überprüfung
- ☐ Überarbeitung
- ☒ Freigabe zur Publ

index.xml x

```
1 <?xml version="1.0" encodin
2 <index>
3   <uuid name="Aydin">e7e1
4   <uuid name="Brinkschmid
5   <uuid name="Brinkschnei
6   <uuid name="Brinkschnie
7   <uuid name="Brinkschröd
8   <uuid name="Celik">dee7
9   <uuid name="Demir">ddaf
10  <uuid name="Ferrai">e19
11  <uuid name="Ferrari">be
12  <uuid name="Ferraro">b8
13  <uuid name="Ferrerri">e2
14  <uuid name="Fexmer">b2f
15  <uuid name="Fixemer">a3
16  <uuid name="Fixmer">fe4
17  <uuid name="Futterschne
18  <uuid name="Gsottschnei
19  <uuid name="Hexelschnei
20  <uuid name="Hillesheim">e1372680-ed7d-440e-89b6-5600d76e9a1d</uu
21  <uuid name="Hillesheimer">cb43a4c5-ce63-40da-9817-f8df650b8751</
```

Datenquellen Explorer

Verbindungen

webdav.namenforschung.net

css

dfd

names

resources

schema

status\_of\_publication

delete

edit

publish

publish

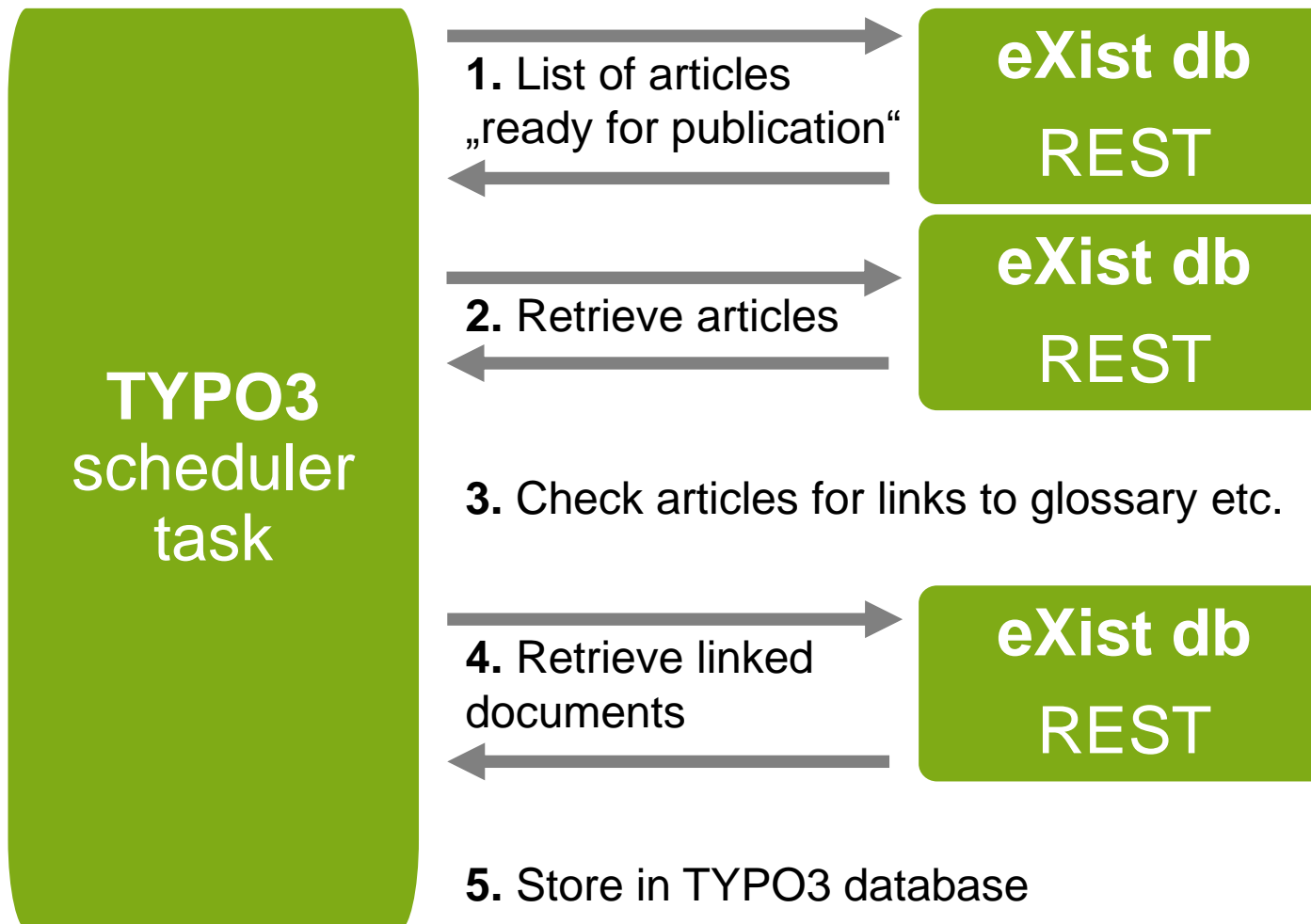
index.xml

status\_of\_publication

08.03.2016 | Digital Dictionary of Surnames in Germany (DFD) | Horn, Hambuch, Reker | 28



# Publishing Workflow: Scheduled Retrieval Task





# Conclusion



- customized, user-friendly interface
- users produce richly annotated documents – without advanced XML skills
- copying of UUIDs is easy
- bibliographical references centrally stored, easily inserted
- automated publishing process – two mouse clicks are enough

**Questions?**

