

Ursachen und Folgen des Bedarfs nach individuellen Softwarelösungen in den digitalen Geisteswissenschaften am Beispiel der bayerischen Dialektwörterbücher der Bayerischen Akademie der Wissenschaften

Manuel Raaf

1. Einleitung

Die oftmals speziellen und sehr auf das jeweilige Thema ausgerichteten Anforderungen geisteswissenschaftlicher Projekte an die IT verlangen meist auch nach speziellen Softwarelösungen, da bereits bestehende Programme diese Ansprüche nicht bedienen können. Der späte Wechsel bei (insbesondere Langzeit-) Projekten vom analogen zum digitalen Workflow ist ein Grund hierfür, da die gewohnten Arbeitsabläufe meist möglichst beibehalten werden sollen. Hinzu kommt die Anforderung, der breiten Öffentlichkeit online einen niedrighschwelligigen Zugang zu Forschungsdaten und -ergebnissen zu ermöglichen. Sowohl die Informationsverarbeitung (Funktionalität) als auch die Informationsdarstellung (Layout) ist daher eine nicht zu unterschätzende Herausforderung, die erfahrungsgemäß mit eigens entwickelter Software besser zu bewerkstelligen ist, als mittels bereits existenter und ggfs. gar generischer. Selbiges gilt für Datenformate/-strukturen.

Anhand der Beispiele aus den Mundartprojekten der BAdW (Bayerisches Wörterbuch, Fränkisches Wörterbuch, Digitales Informationssystem von Bayerisch-Schwaben), der letztjährig übernommenen Bayerischen Dialektdatenbank sowie dem Derivat dieser vier Datenbasen namens *Bayerns Dialekte Online* soll unter Fokussierung der Informationsdarstellung skizziert und diskutiert werden, welche Ursachen und Folgen individuelle Softwarelösungen aufweisen und welchen zusätzlichen Mehrwert sie gegenüber generischen Lösungen liefern können.

2. Ursachen & Lösungen

2.1 Allgemein

Die primären Ursachen, die eine individuelle Softwarelösung erfordern, sind i.d.R. spezifische Eigenschaften des jeweiligen Projekts bezüglich dessen Daten und/oder Arbeitsabläufen. Oft sind die Daten (Projekthinhalte) sehr granular erfasst, vielschichtig, komplex und sollen entsprechend den Anforderungen der Projektmitarbeiter:innen individuell verarbeitet sowie dargestellt werden.

Auch der niedrighschwellige Zugang für die interessierte Öffentlichkeit spielt eine zunehmend wichtigere Rolle. Für Neu- sowie Nachfolgeanträge ist zudem die Verwendung informationstechnologischer Methoden eine Voraussetzung für die erfolgreiche Antragsstellung, jedoch erwarten Evaluationsgremien bestehender Langzeitprojekte nicht zuletzt auch deshalb eine entsprechende digitale Umsetzung bisheriger Forschungsdaten und -ergebnisse (vgl. BMBF 2016; DFG 2020; Volkswagenstiftung 2021).

Liegen die Forschungsprimärdaten analog vor in Form von Fragebögen oder Zettelkästen, wie dies bei älteren variationslinguistischen Vorhaben die Regel ist aufgrund der zu Projektbeginn fehlenden

technischen Möglichkeiten, findet man für gewöhnlich keine zufriedenstellende fertige Softwarelösung, die die Struktur der Daten nach deren Digitalisierung entsprechend genau behandeln kann. Hat man bereits digitale Daten in Form von Word-Dateien, sind diese jedoch nur bedingt maschinenlesbar und bedürfen einer Konvertierung, die sich stark an die im Projekt verwendeten Formatierungen halten muss. Fehlt es diesen an Eindeutigkeit, ist eine Umwandlung ohne manuelle Korrekturarbeiten ohnehin nicht möglich. Sollen Projekte vernetzt werden, die unterschiedliche Datenqualitäten bereitstellen (z. B. redaktionell verfasste Artikel vs. Beleglisten) und/oder verschiedene Datenformate nutzen (SQL vs. XML vs. Text mit individuellen Metazeichen vs. Word vs. Print-PDF), kann dies auch nur von eigener Software bzw. zumindest eigenen Modulen für existente Software gelöst werden und bedeutet damit den Bedarf an individueller Programmierung. Fremdsoftware stellt nämlich hierfür i.d.R. nicht (die gesamte) nötige Funktionalität bereit. Bedenkt man die Spezifika des eigenen Projekts, ist dies auch nicht mehr verwunderlich: Fremde Programme können die projekteigenen Strukturen selbstredend nicht kennen oder ausreichend flexibel verarbeiten, selbst wenn es sich um Software aus thematisch ähnlichen Projekten handelt (z. B. innerhalb des Bereichs der Dialektlexikographie).

Zwar gibt es seit Beginn der Digital Humanities Bestrebungen, all diese Problemfälle im jeweiligen Anwendungsbereich oder gar darüber hinaus generisch in einer Softwarelösung abzufangen, doch mangels entsprechender künstlicher Intelligenz kann Software dies aktuell schlichtweg noch nicht leisten. Weitgehend generisch konzipierte Lösungen bleiben letztlich Inselfösungen (BBAW 2020; Textgrid; Mehler – Gleim 2016)¹ oder scheitern (LRZ 2019). Computer sind durchaus dem Menschen überlegen, wenn es um lineare Rechenaufgaben geht. So können die Inhalte tausender Bücher in wenigen Sekunden durchsucht oder auch konvertiert werden. Auch Bilder können schnell verarbeitet und z. B. zurechtgeschnitten werden. Abstrakte Vorgänge hingegen sind eine immense Herausforderung für die Informationstechnologie, weniger jedoch für den Menschen: Die semantische Disambiguierung von Homonymen ist für uns ein Leichtes im direkten Vergleich zur algorithmischen Lösung, da wir hierfür auf ein entsprechend großes Weltwissen zurückgreifen und dieses logisch adäquat verarbeiten können (vgl. Liqun Luo 2018; Rodriguez-Ramos 2018). Auch die Entscheidung darüber, welche Inhalte verschiedener Ressourcen miteinander verbunden werden müssen für eine quellenübergreifende Suche ist nach wie vor Arbeit, die vom Menschen übernommen werden muss. Software kann hier unterstützen, muss dann jedoch für den jeweiligen Anwendungsfall geeignet und damit spezifisch anstatt generisch sein.

Es bedarf somit der Erstellung eigener Programme, die innerhalb der Institution an die eigenen Belange und somit sehr projektnah ausgerichtet sind, um die genannten Probleme zu lösen. Anwendungen dieser Art werden auch als „In-House-Software“ bezeichnet, da sie speziell und i.d.R. zunächst nur für den hausinternen Gebrauch geschaffen werden. Die Möglichkeit der Verwendung bereits existenter Software sollte im Vorfeld jedoch unbedingt eruiert werden, allerdings immer im Hinblick auf die zur Verfügung stehende Zeit: Ein Ausprobieren von Fremdsoftware über Monate hinweg kann den Projektverlauf immens verzögern, wenn sich letztlich und damit (zu) spät herausstellt, dass die Software nicht im Rahmen des Vorhabens eingesetzt werden kann.

¹ Die hier beispielhaft genannten, sehr großen und zurecht bekannten Projekte bedienen mehrere Arbeitsbereiche innerhalb der digitalen Geisteswissenschaften, können doch auch innerhalb dieser nicht ohne projektspezifische Erweiterungen wirklich alle Bedürfnisse aller Projekte generisch abdecken. Somit stellen auch sie Inselfösungen dar, wenn auch recht große.

2.2 Das Bayerische Wörterbuch (BWB)

2.2.1 Ursachen

Das Bayerische Wörterbuch führte ab 1913 verschiedene Erhebungen durch, von denen bis auf die Wörterlisten, die 1958 begannen, alle abgeschlossen sind (Schnabel et al. 2020: 50ff). Die allesamt in Papierform und in unterschiedlichen Formaten vorhandenen Rückläufe dieser Erhebungen stellen den primären Quellenbestand des BWB dar. Sie wurden bis 2016 händisch ausgewertet, um als Belegmaterial in die in Microsoft Word geschriebenen Artikel einfließen zu können. Eine unterstützende digitale Komponente existierte bis dato nicht im Arbeitsablauf.

Allein für die sehr umfangreiche Sammlung der Wörterlisten existieren rund 100.000 Exemplare mit jeweils vier Seiten, auf denen 60 Fragen enthalten sind (siehe Abbildung 1). Durch die Vielzahl an Rückläufen aus dem Untersuchungsgebiet ergeben sich über alle Sammelorte hinweg insgesamt etwas mehr als sechs Millionen einzelne Belege innerhalb dieser Sammlung. Die Anforderung an die digitale Unterstützung war daher, Belege gezielt nach Bogen, Frage und/oder Verortung aufrufen und mit wenigen Mausklicks einem Lemma zuweisen zu können.

Abbildung 1 Ausschnitt eines Fragebogens

239
Gemeinde: Abenberg Kreis: Schwabach

12. FEB. 1960

Wörterliste 6 f'32

Bitte geben Sie bei jedem Wort möglichst genau die **Bedeutung** und die **Aussprache** an (am besten in einem **Satz** oder mit einer **Zeichnung**). Achten Sie bitte auf die Unterscheidung von hellem a (wie in schwaa 'schwer') und dunklem ä (wie in Häwan 'Hafer').
Vergessen Sie bitte nie auf die Angabe Ihrer **Anschrift**, selbst dann nicht, wenn Sie uns den Bogen unbeantwortet zurück-schicken müssen.

1. Wird in Ihrer Gegend der Ausdruck **es äntelt** gebraucht, wenn der See im Sturm Wellen wirft? Wie erklären Sie sich diese Bezeichnung? Bitte Satzbeispiel. nein

2. Kennen Sie den Ausdruck **Heubader** für „Platzregen“? Wie heißt der Platzregen sonst?

3. Ist Ihnen **baisseln** (boassln), **schneebaisseln**, **schneebaissen** für „hageln“ bekannt? Welches Wort wird sonst für „hageln“ gebraucht? Genaue Aussprache? Kirsln

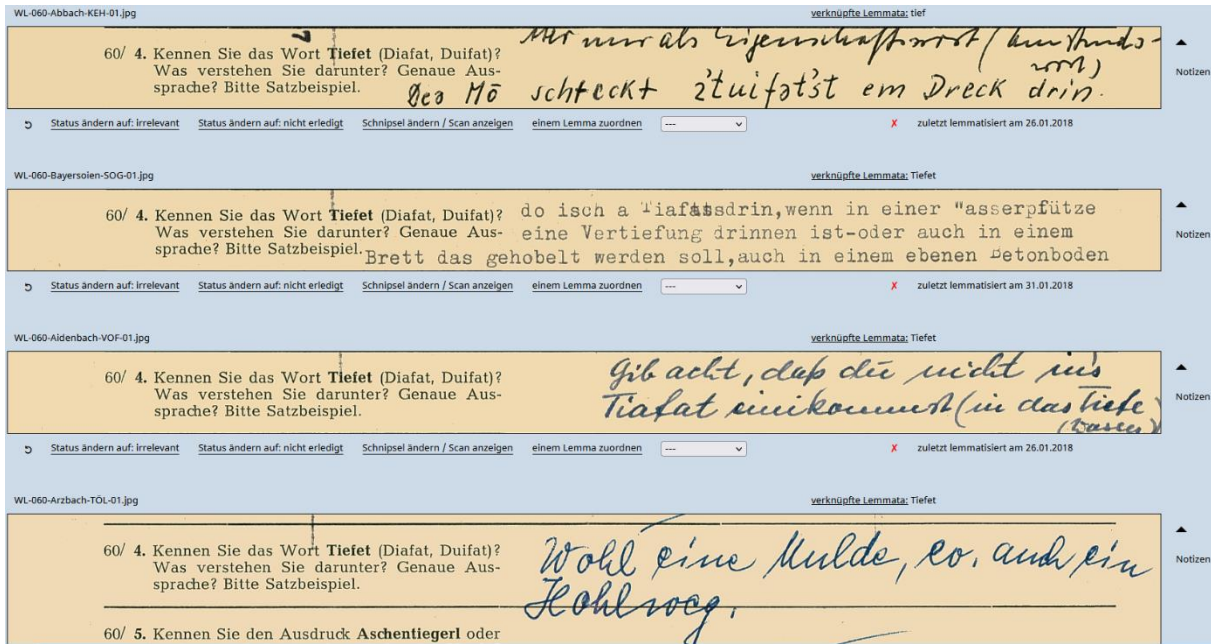
4. Kennen Sie die Bezeichnungen **Baisel** (Boassl), **Schneebaisel**, **Baizel**, **Schneebaizel** für „kleines Hagelkorn“? Wie heißen die Hagelkörner sonst? nein
Hagelkorn, Schneekorn

2.2.2 Lösungen

Im April 2016 begann die digitale Unterstützung der redaktionellen Arbeitsabläufe am BWB, indem das für das Fränkische Wörterbuch entwickelte Redaktionstool in angepasster Form für das BWB in Betrieb genommen wurde (Raaf 2016a; siehe auch 2.4.2). Wörterbuchartikel schrieb die Redaktion zunächst weiterhin mit Microsoft Word; das Tool erleichterte lediglich die Exzerpierung. Hierfür wurden die Wörterlisten in einzelne Bildschnipsel zerlegt (Raaf 2017), sodass über diese gezielt eine bestimmte Frage eines Bogens aufgerufen werden konnte. Irrelevante Belege (d.h. leer oder ohne fragenbezogenen

Inhalt) können zudem vorher sehr effizient durch Hilfskräfte markiert werden, um in der Suche ausgefiltert zu werden (siehe Abbildung 2). Bereits vergebene Lemmata lassen sich über eine Auswahlliste mit wenigen Klicks dem Beleg zuordnen, sodass das Exzerpieren mithilfe der projektspezifischen Softwarefunktionalitäten insgesamt um ein Vielfaches schneller vonstattengeht, als im zuvor analogen Betrieb (Schnabel et al. 2020: 70f). Eine lemmatisierte Belegliste lässt sich bei Bedarf in einer PDF zusammenstellen, mithilfe derer die Redaktor:innen dann den Artikel schreiben.

Abbildung 2 Redaktionstool des BWB



Im April 2019 erfolgte schließlich nach einigen Monaten der Vorbereitung der Umstieg auf einen komplett digitalen Arbeitsablauf, indem Artikel seither mittels Oxygen² in das zuvor in engem Kontakt mit der Redaktion ausgearbeitete XML geschrieben werden (vgl. Abbildung 3 und Abbildung 4). Die Entscheidung, hierfür ein eigenes XML anstatt des De-Facto-Standards TEI zu verwenden, fiel aufgrund pragmatischer Gründe: Die Interoperabilität, die für TEI einen der oft genannten Hauptgründe darstellt, geht verloren, wenn ein Derivat erstellt wird, um die projektspezifischen Strukturen abbilden zu können. Ohne die Übermittlung des zugehörigen Schemas ist damit ein solches TEI-Derivat nicht automatisch interoperabel mit anderen, ggfs. nur das originale TEI verwendenden Projekten. Das Bereitstellen des Schemas lässt hingegen den diesbezüglichen Unterschied zwischen dem Derivat und einem eigenen XML verschwinden.³ Darüber hinaus ist TEI in seiner Basisform ebenso wie das auf die Lexikographie bezogene Derivat „TEI Lex-0“ (vgl. Bański et. al 2017; vgl. DARIAH-ERIC Working Group

² Oxygen ist ein weit verbreiteter, leistungsfähiger, jedoch kostenpflichtiger XML-Editor.

³ Wer TEI gewohnt ist, wird sich in einem weitgehend am ursprünglichen Format orientierten Derivat selbstredend schneller zurechtfinden, als in einem völlig davon abweichendem. Ob und wie weit das Derivat diese Anforderung jedoch erfüllt, ist völlig frei derjenigen Person überlassen, die es erstellt. Es kann daher nicht pauschal gesagt werden, dass TEI-Ableitungen generell leichter nutzbar sind als eigene XML-Strukturen. Aus diesem Grund sind nach Ansicht des Autors beide Lösungen in Bezug auf die Interoperabilität als grundsätzlich gleichwertig anzusehen.

2017) aufgrund seines jeweils ganz eigenen Vokabulars von dem des Arbeitsalltags der Redaktion weit entfernt.

Die Verwendung eigener Schemata und gar sogenannter domänenspezifischer Sprachen erwies sich darüber hinaus auch in einem anderen an der BAdW angesiedelten Projekt bereits als sinnvoll (vgl. Arnold 2019; Arnold 2020; Arnold 2021): Im Mittellateinischen Wörterbuch ist es durch die Verwendung eigener Lösungen bezüglich des Formats, das in einem einfachen (und dazu noch kostenfreien) Texteditor eingegeben werden kann, möglich, zum einen sehr auf die Wünsche der Redaktion einzugehen und damit selbst hochkomplexe Anforderungen an die Struktur vergleichsweise einfach redaktionsnah abbilden zu können. Zum anderen ist es möglich, dieses Format durch eigene Konverter in jedes andere zu übertragen und somit letztlich auch den Drucksatz selbst zu produzieren, z. B. mittels LaTeX⁴, sowie eine HTML-Ausgabe für die Online-Publikation zu erzeugen. Auf Autovervollständigung, die Einbindung von kontrolliertem Vokabular, oder anderen arbeitserleichternden Eingabehilfen muss während der redaktionellen Arbeit nicht verzichtet werden, sodass dieser Ansatz eine vollständige Alternative zum XML-Editor darstellt. Für das BWB war jedoch eine domänenspezifische Sprache nicht erforderlich, sodass es bei einem redaktionseigenen XML belassen wurde (vgl. Abbildung 4).

⁴ LaTeX ist eine freie Software, die das freie und leistungsstarke Textsatzsystem TeX verwendet.

Abbildung 3 Artikel „predigen“ des BWB im Word-Format

predigen

Vb. **1** die Predigt halten, (Gottes Wort) verkündigen.— **1a** (im Gottesdienst) predigen,

°Gesamtgeb. vielf.: *pretit* „gepredigt“ Reitrain MB; *wans schee prödöngan, nacha is dö ganz Kiacha voi Lait* Hengersbg DEG; *da Här Pfarra tuat bretinga* Rgbg; *Vom Eh'stand hat der Pfarra 'predigt* {Fliegende Bl. (München) 73 (1880) 99}; *konnst du mia vielleicht erzähln, was da Pfarrer predigt hod?* {HERRLEIN Wallfahrt 15}; *er prediget an sand Matheiss tag* {ARNPECK Chron. 539,12}; *nutzt kein bredigen auch nicht* {Bilanz 1782 26}.— Ra.: *der Pfarrer predigt nicht zweimal / nur einmal* u.ä. Weigerung, etwas Gesagtes zu wiederholen,

°OB, °NB vielf., °OP mehrf., °MF, °SCH vereinz.: *°moanst du laichd, da Bfara bredigd dswoamoi fia di!* Rosenhm; *°da Pfoarra predicht blouß oamal* Wdsassen TIR; *„Der Pfarrer predigt nicht zweimal ... wenn einer nicht aufpaßt, wenn man ihm etwas sagt“* {Oberpfalz 78 (1990) 194}.

— **1b** (Gottes Wort) verkündigen: *dös werd scho' dö richtige Religion sei', dö wo der predigen kunnt* {LUTZ Zwischenfall 38}; *Categorizo ich #p³[re]dion* Tegernsee MB 11.Jh. {StSG. IV,242,11}; *do er sein hilig junger daz hilig ewangelium in der heidenscheft #u⁶ber alle die werlt bredigen hiezze* {O'altaicher Pred. 29,8-10}; *es sei dann in derselben Woche kein Feiertag, daran göttlichs Wort gepredigt* Wunsiedel 1544 {ZILS Handwerk 23}.

2 in eindringlicher Weise ermahnen, ans Herz legen, OB, NB vereinz.: *hör auf mit deim Predign!* Haag WS; *Das is die christli' Menschenlieb, woaft, die si uns predinga* {S.

SCHUBAUR, Mein Vermächtniß an Bayern, Leipzig 1831, 294}; *Da sitzt aa so oana, der allawei predigt hat: nur Ruhe – nur Ruhe!* {THOMA Werke II,256 (Lokalbahn)}.

— Auch in fester Fügung: *moralpredign* eine Moralpredigt halten G'weismannsdFÜ.

3: *°predign* „schnurren, von der Katze“ Parsbg.

Abbildung 4 Artikel „predigen“ des BWB im XML-Format

```

<bdo>
  <artikel id="bwb_predigen" wb="bwb">
    <lemma-position>
      <lemma>predigen</lemma>
      <grammatik wortart="Vb" numerus="Singular"/>
    </lemma-position>
    <bedeutung-position nr="1">
      <bedeutung>die Predigt halten, (Gottes Wort) verkündigen</bedeutung>
    </bedeutung-position nr="1a">
      <bedeutung>(im Gottesdienst) predigen</bedeutung>
    </bedeutung-position>
    <verbreitung-position>
      <verbreitung-angabe belegt="nach1958" bezirk="Gesamtgebiet" häufigkeit="vielfach"/>
    </verbreitung-position>
    <beleg-position>
      <beleg-angabe>
        <beleg-text>pretit</beleg-text>
        <beleg-kontext provenienz="zitat">gepredigt</beleg-kontext>
        <beleg-quelle>
          <beleg-region ort-landkreis="Reitrain-MB"/>
        </beleg-quelle>
      </beleg-angabe>
      <beleg-angabe>
        <beleg-text>wans schee prödöngan, nacha is dô ganz Kiacha voi Lait</beleg-text>
        <beleg-quelle>
          <beleg-region ort-landkreis="Hengersberg-DEG"/>
        </beleg-quelle>
      </beleg-angabe>
      <beleg-angabe>
        <beleg-text>da Här Pfärra tuat breitinga</beleg-text>
        <beleg-quelle>
          <beleg-region ort-landkreis="Regensburg-R"/>
        </beleg-quelle>
      </beleg-angabe>
      <beleg-angabe>
      </beleg-angabe>
      <beleg-angabe>
      </beleg-angabe>
      <beleg-angabe>
      </beleg-angabe>
    </beleg-position>
    <phraseologie-position>
      <phraseologie>
        <phraseologie-typ>Ra.</phraseologie-typ>
        <phrase>
          <beleg-typ>der Pfarrer predigt nicht zweimal / nur einmal</beleg-typ>
          <erklaerung>u.ä.</erklaerung>
          <bedeutung>Weigerung, etwas Gesagtes zu wiederholen</bedeutung>
        </phrase>
      </phraseologie>
      <verbreitung-position>
        <verbreitung-angabe häufigkeit="vielfach" belegt="nach1958" bezirk="OB"/>
        <verbreitung-angabe häufigkeit="vielfach" belegt="nach1958" bezirk="NB"/>
        <verbreitung-angabe belegt="nach1958" bezirk="OP" häufigkeit="mehrfach"/>
        <verbreitung-angabe häufigkeit="vereinzelt" belegt="nach1958" bezirk="MF"/>
        <verbreitung-angabe belegt="nach1958" bezirk="SCH" häufigkeit="vereinzelt"/>
      </verbreitung-position>
      <beleg-angabe belegt="nach1958">
        <beleg-text>moanst du laichd, da Bfara bredigd dswoamoi fia di!</beleg-text>
        <beleg-quelle>
          <beleg-region ort-landkreis="Rosenheim-RO"/>
        </beleg-quelle>
      </beleg-angabe>
    </phraseologie-position>
  </artikel>
</bdo>

```

2.3 Das Dialektologische Informationssystem von Bayerisch-Schwaben (DIBS)

2.3.1 Ursachen

Im August 2017 kam das Dialektologische Informationssystem von Bayerisch-Schwaben an die Bayerische Akademie der Wissenschaften. Es brachte eine bereits seit vielen Jahren befüllte Access-

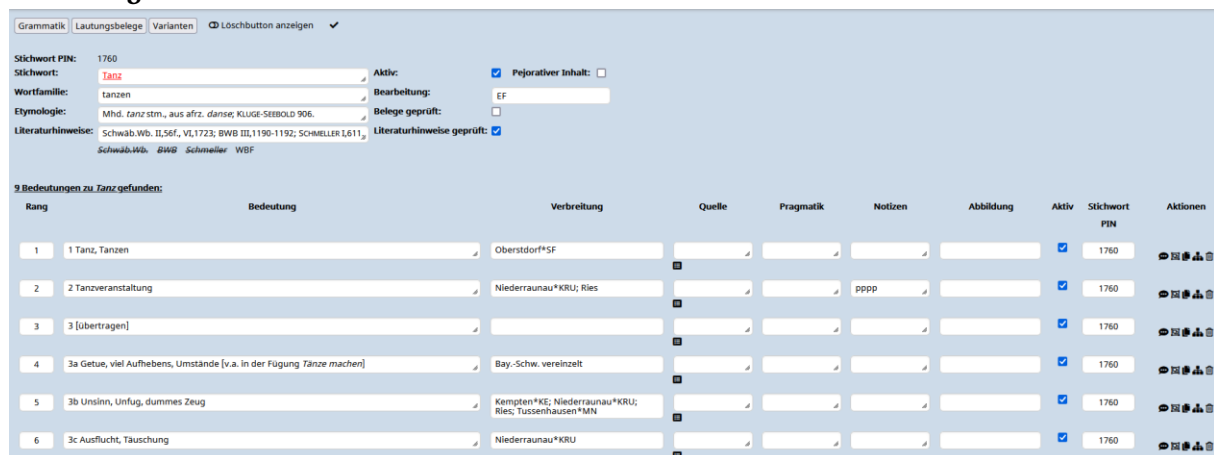
Datenbank mit, die die Grundlage für das 2013 erschienene Dialektwörterbuch von Bayerisch-Schwaben darstellte (Schwarz 2013). Digitalisate von Fragebögen waren in dieser Datenbank nicht eingebunden, sollten aber im Laufe der Zeit hinzukommen. Da ein kollaboratives Arbeiten ein einer einzelnen und ohne Kontrollmechanismen versehenen Access-Datenbank weder komfortabel noch fehlerfrei möglich war, bedurfte es einer Lösung, die dies ermöglichte und zudem die bisher nur offline vorhandenen Daten auch ins Web stellen konnte. Dabei sollte der bisherige Arbeitsablauf jedoch weitgehend beibehalten werden können – auch bez. der Eingabe und der dafür notwendigen Anzahl an Mouse-Klicks –, sodass die Lösung eine Art Klon der bisherigen Access-Formulare werden musste.

2.3.2 Lösungen

Anfang 2019 wurde deshalb ein sehr am Aufbau der bisher genutzten Access-Datenbank orientiertes, webbasiertes Redaktionstool entwickelt (Funk et. al 2020; Raaf 2020). Zunächst wurde es nur redaktionsintern verwendet, um die Artikelinhalte in gewohnter Manier einzugeben. 2020 erfolgte schließlich die Freischaltung für die interessierte Öffentlichkeit. Im weiteren Projektverlauf soll die Anwendung mit Digitalisaten, Audio- und Videoaufnahmen und Vernetzungen mit anderen Projekten aufgewertet werden. Zu diesem Zwecke werden bis Herbst 2021 zwei Typen von digitalisierten Fragebogenerhebungen in Schnipsel von Einzelbelegen aufgeteilt und lemmatisiert, sodass diese – wie auch im BWB und WBF – gezielt aufgerufen werden können.

Eine bereits bestehende und für den Anwendungsfall allumfassend passende Lösung, die Access-Datenbank mitsamt ihrer Formulare online zu stellen und im Zugriff der öffentlichen Ansicht auf Leserechte zu beschränken, existiert nach Wissen des Autors nicht. Zwar bietet Access die Möglichkeit an, Webdatenbanken zu erstellen, jedoch sind diese im Funktionsumfang eingeschränkt.

Abbildung 5 Redaktionstool des DIBS



2.4 Das Fränkische Wörterbuch (WBF)

2.4.1 Ursachen

Das Fränkische Wörterbuch entstand 1933 als „Ableger des Bayerischen Wörterbuchs“ (König/Raaf 2020: 78f). Seit 2004 werden Belege, die in den sogenannten Nachkriegsbogen zwischen 1960 und 2001 erhoben wurden, in Excellisten zusammengetragen und lemmatisiert. Bis zur Entwicklung eines Redaktionssystems, in dem die Fragebögen gezielt durchsucht und angezeigt werden können, geschah das Prüfen der Bögen in manueller Handarbeit. Um z. B. die Frage 4 des Bogens 50 bearbeiten zu

können, mussten hierfür alle Papierbögen aus den Schubern geholt, ggfs. nach Ort und Landkreis sortiert und dann in einzelnen Arbeitsschritten abgetippt, lemmatisiert und annotiert werden. Das Erfassen der Belege konnte nicht durch OCR oder ähnliche Techniken übernommen werden, da die Fehlerrate bei handschriftlich ausgefüllten Listen – insbesondere bei solchen in Sütterlin – zu hoch ist. Die Idee für die digitale Unterstützung war daher Anfang 2015 zunächst einmal nur, die Bildbereiche der Belege jeweils einzeln aufrufen zu können und die dafür erfassten Daten der Excelliste online darstellen zu können

Die verschiedenen Arbeitsschritte in der Digitalisierung der Belege (König – Raaf 2020: 84ff) gestalteten es zudem als äußerst zeitaufwändig, die durch Hilfskräfte und Freiwillige zusammengetragenen Belege aufzuteilen und zu lemmatisieren. Im Frühjahr 2020 kam daher die Anfrage, ob dies vereinfacht werden könnte: Aus z.B. dem Einzelbeleg „nimm a Stickla Brut mit“ und dessen Umschrift sollte je eine Zeile mit der jeweiligen in der Umschrift enthaltenen Grundform (hier also: „nimm“, „ein“, „Stücklein“, „Brot“, „mit“) werden. Dieser Vorgang wurde bis dato händisch durch Kopieren und Einfügen der Zeile sowie dem manuellen Eintippen des Lemmas vorgenommen – bei je nach Bogen zwischen ca. 15.000 und 22.000 Belegen eine sehr zeitaufwändige Angelegenheit. Auch die Auftrennung der Gesamtliste je Bogen zu Einzellisten je Frage und die Sortierung nach Lemma war aufgrund der Größe der Dateien oftmals schwierig, sodass auch dies übernommen werden sollte. Die Anreicherung der dann erstellten Aufteilung mit Bedeutung und Grammatik sollte darüber hinaus aus der bestehenden Datenbank übernommen werden, sofern Übereinstimmungen vorhanden sind.

2.4.2 Lösungen

Im Früher 2015 wurde mit Entwicklung eines Redaktionstools namens *LexHelfer* begonnen, um der lexikographischen Arbeit der Redaktion ein Helfer zu sein. In diesem sind die Belege alle einzeln aufrufbar, sodass digital und nicht mehr in Papierform exzerpiert werden kann (König – Raaf 2020: 88ff). Es stellt zudem nach dem Upload die Excelliste online dar, um sie granular und über die SQL-Datenbank deutlich schneller durchsuchen und sortieren zu lassen, als dies in einer Excel-Datei möglich wäre, da hier insbesondere bei vielen tausend Zeilen Excel oft träge oder gar nicht mehr reagiert. Im Laufe der Zeit ergaben sich weitere Wünsche, die in das Tool implementiert wurden, sodass inzwischen auch ganze Spalteninhalte ersetzt werden können, Statistiken zu Lemmata, Grundformen und Bedeutungen verfügbar sind, oder nach jedem Import eine Liste der neu hinzugekommenen Lemmata an die Redaktion gemailt wird.

Für die Anfrage, den Lemmatisierungsvorgang in Excel beschleunigen zu können, wurde im Mai 2020 ein VBA-Addon entwickelt, das diese Aufgaben vollautomatisch innerhalb der jeweiligen Exceldatei übernimmt (siehe Abbildung 7).⁵

Die Excellisten online darzustellen und auch zu verarbeiten, ist seit vielen Jahren mit entsprechenden Lösungen von Microsoft oder Mitbewerbern (wie z. B. OnlyOffice) möglich. Sie geben jedoch als webbasierte Implementierungen ihrer Desktop-Varianten nicht die Möglichkeiten, gezielt nur in der Grundform zu suchen und diesen Vorgang an eine weitere Suchebene zu binden, um z. B.

⁵ Mit VBA (Visual Basic for Applications), einer Programmiersprache, die von Visual Basic abgeleitet ist, können innerhalb von Office-Dokumenten Programme mitsamt grafischer Benutzeroberfläche erstellt werden, um darüber in z. B. Excel über Formulare den Inhalt des Arbeitsblattes zu bearbeiten oder Dateiinhalte (auch aus dem Web) einzulesen, etc.

nur Ergebnisse aus bestimmten Landkreisen zu erhalten. Auch die Einbindung von weiterer Funktionalität – wie z. B. der Anzeige der Belegschnipsel oder das erwähnte VBA-Addon – ist in diesen Web-Varianten nicht möglich. Aus diesen Gründen stellte deren Verwendung keine zufriedenstellende Lösung für die Wörterbuchredaktion dar.

Abbildung 6 Redaktionstool des WBF

Bitte geben Sie hier die gewünschten **Suchbegriffe** ein.
Die Eingabe mehrerer Parameter bewirkt eine UND-Verknüpfung!

Sie können die **Sortierreihenfolge** mehrerer Spalten als Ergänzung zur Suchoption definieren oder später die Ergebnisliste nach einer einzelnen Spalte durch Klicken der Spaltenüberschrift sortieren:
Die Sortieroptionen haben keine Wirkung, wenn *nur Schnipsel laden* gewählt wurde bzw. wenn *nur Bogen und Frage* gesucht werden!

Volltext: **?**
und/oder
Lemma:
Grundform:
Bedeutung:
Grammatik:
Ort:
Bogen Nr.:
Frage Nr.:

Erste Sortier-Spalte: Zweite Sortier-Spalte: Dritte Sortier-Spalte: Vierte Sortier-Spalte: Fünfte Sortier-Spalte:

aufsteigend absteigend rückläufig

aufsteigend absteigend rückläufig

aufsteigend absteigend rückläufig

aufsteigend absteigend rückläufig

aufsteigend absteigend rückläufig

Ergebnisse nach Lemma & Grundform & Bedeutung gruppieren
 alle Ergebnisse laden (ist diese Option gesetzt, werden nicht nur die ersten 1000 Ergebnisse geladen, sondern alle. Dies kann sich u. U. negativ auf die Geschwindigkeit auswirken und Ihren Browser überfordern lassen.
Wenden nur Schnipsel geladen, ist diese Option aus diesem Grund immer ohne Wirkung.)

Ihre Suche nach Grundform *schnecke* ergab 27 Treffer.

Lemma	Grundform	Bedeutung	Grammatik	Originaltext	Umschrift	Ort	Planquadrat	Kommentar	Gewähsperson	Kommentar	Bearbeiter	GP	Bogen	Frage	Bild
Schnecke	Schnecke	Anrede der Geliebten	SF NonSg	Schnecke	Schnecke	Kronach	T34,3	Junge zu Mädchen				1	116	11	
Schnecke	Schnecke	Anrede unter Verliebten	SF NonSg	Schnecke	Schnecke	Schney	U33,2					1	116	11	
Schnecke	Schnecke	Damenfrisur	SF NonSg	Schalcke	Schnecke	Oberschleichach	W30,5	als Haartracht, im Ohr		#np		1	116	9	
Schnecke	Schnecke	Freundin, hübsche Frau	SF DatSg	er geht mit seiner Schnecken spazieren	er geht mit seiner Schnecke spazieren	Heroldsberg	b33,6	Freundin		#np		1	116	9	
Schnecke	Schnecke	Gebäck	SF NomPl	Schnecken	Schnecken	Breitendiel	222,7					1	74	20	
Schnecke	Schnecke	Geblicksorte	SF NonSg	Schnecke	Schnecke	Schwabach	e33,1	als Kosenwort oder Backwerk in vielen Varianten		#np		4	116	9	
Schnecke	Schnecke	Geblicksorte, süß	S Nom	Schnecke	Schnecke	Thüngen	W26,7	"Schnecke mit Zuckerglasur"		#bild		1	2	32	

Abbildung 7 Lemmatisierungshilfe des WBF

Welches Feld soll angereichert werden? Bedeutung Grammatik beides

Aktuelle Zeile:

Die Struktur der Ausgabe ist: Bedeutung | Grammatik | Sachgruppe(n) [Link zum Lex-Helfer](#)
(Hinweis: Die Sachgruppen sind gruppiert nach der Bedeutung und ggfs. vielschichtig)

[Füllfederhalter](#)

[langer, dünner Schieferstift ohne Holzhülle](#)

Debug-Info

Hole Daten zu Grundform 'Schreiber'

1 Einträge gefunden

2.5 Die Bayerische Dialektdatenbank (BayDat)

2.5.1 Ursachen

Die bayerische Dialektdatenbank wurde 2006 veröffentlicht (Zimmermann: 2006) und basierte auf einer durch ein Webinterface abgefragten Oracle-Datenbank, die rund fünf Millionen Belege der bayerischen Sprachatlanten enthielt. Da sowohl die Oracle-Version als auch das Betriebssystem des

zugehörigen Servers nicht mehr aktuell waren, drohte die *BayDat* 2018 im digitalen Nirwana zu verschwinden. Das Rechenzentrum der Universität Würzburg hatte selbstredend keine Kapazitäten für eine Weiter- oder Neuentwicklung des Systems, da i.d.R. kein Rechenzentrum personelle Mittel für umfassende Aufgaben der Art für die gesamte Hochschule bereitstellen kann. Eine Abschaltung der *BayDat* wäre insbesondere für die Dialektwörterbücher der BAdW fatal gewesen, da sie von ihnen sehr regelmäßig konsultiert und als Quelle verwendet wird.

2.5.2 Lösungen

Dank freundlicher Genehmigung der Akademieleitung und des Lehrstuhls für Deutsche Sprachwissenschaft der Universität Würzburg sowie mit Unterstützung der Würzburger Kollegen am dortigen Rechenzentrum wurde die Oracle-Datenbank im Herbst 2018 nach MariaDB konvertiert und eine neue Weboberfläche dafür geschrieben (Raaf 2019).⁶ Hierbei wurden die Datenstrukturen angepasst, einige Relationstabellen hinzugefügt, und Fehlerkorrekturen vorgenommen. Mit der Übernahme der *BayDat* eröffnete sich zudem die Möglichkeit, die selten und ausnahmslos in Kapitälchen lemmatisierten Belege vollständig und in üblicher Orthographie zu lemmatisieren sowie mit weiteren Informationen (Bedeutung, Grammatik, Sachgruppe) anzureichern. Das WBF führte dies durch die Förderung des jeweiligen Bezirks für die fränkischen Regierungsbezirke in einem eigens dafür entwickelten Tool aus.

2.6 Projektübergreifende Suche

2.6.1 Ursachen

Zwar sind die Datenbestände von BWB, DIBS, WBF sowie der *BayDat* frei online zugänglich, jedoch nur einzeln durchsuchbar und in der Darstellung der Ergebnisse nicht einheitlich. Im Falle des BWB sind ferner Artikelinhalte nicht granular auffindbar, da sich der Index auf dem Publikationsserver der BAdW (vgl. Bayerisches Wörterbuch 2020) ebenso wie die wortbezogene Suche innerhalb des BWB-*LexHelfers* lediglich auf das Lemma stützt, nicht jedoch auf z. B. Inhalte der Bedeutungspositionen oder der Etymologie. Zudem wurden der jeweilige *LexHelfer* des BWB und des WBF anfangs nur für die interne Nutzung konzipiert, sodass das Erscheinungsbild wenig modern und insbesondere auf mobilen Endgeräten kaum komfortabel erscheint. So entstand 2018 die Idee, eine Aufhebung dieser Heterogenitäten in einem gemeinsamen System zu verwirklichen und dabei Unterschiede der Projektbestände gezielt zu vereinheitlichen. Neben der Homogenisierung waren die Anforderungen insbesondere der niedrighwellige Zugang, die Zitierfähigkeit von Suchergebnissen über persistente Links, die Visualisierung der Beleginformationen auf einer Landkarte, die Nachnutzbarkeit der Daten und die einfache spätere Erweiterung bzw. Pflege. Eine bestehende Softwarelösung für diese Belange war und ist uns nicht bekannt.

⁶ Bei Oracle (genauer: Oracle Database) handelt es sich um ein proprietäres Datenbankmanagementsystem der Firma Oracle Corporation. MariaDB ist ein quelloffenes Datenbankmanagementsystem, das aus seinem Vorgänger MySQL hervorging und inzwischen MySQL – nachdem dieses 2009 von Oracle Corporation übernommen wurde – weitgehend ersetzt hat innerhalb der OpenSource-Community.

2.6.2 Lösungen

Das Sprachinformationssystem *Bayerns Dialekte Online* vereint die Dialektwörterbücher BWB, DIBS, WBF und die *BayDat* in einem Portal. Es wird seit 2018 (Planung) bzw. 2019 (technische Umsetzung) im IT-Referat der Bayerischen Akademie der Wissenschaften entwickelt und befindet sich im Sommer 2021 in der öffentlichen Beta-Phase, nachdem bereits intern durch die Redaktionsmitarbeiter:innen der beteiligten Wörterbücher Tests durchgeführt und das Feedback entsprechend eingearbeitet wurde. Da es sich bei diesem System um ein technologisches wie auch fachwissenschaftliches Desiderat handelt, konnte hierfür keine bereits existente Software verwendet werden.

Abbildung 8 Startseite der BDO

BDO | Bayerns Dialekte Online

Startseite | Suche | API | über BDO | Kontakt | Impressum

BDO | Bayerns Dialekte Online

BDO gibt Auskunft über den mundartlichen Wortschatz in Bayern und präsentiert die Forschungsergebnisse der BAdW-Projekte [Bayerisches Wörterbuch \(BWB\)](#), [Fränkisches Wörterbuch \(WBF\)](#) und des [Dialektologischen Informationssystems von Bayerisch-Schwaben \(DIBS\)](#).

[BWB](#) [WBF](#) [DIBS](#)

Word Cloud aus zufällig ausgewählten Stichwörtern der Datenbank (🔄 neu laden)

Klicken Sie ein Stichwort an, um die Suche danach zu öffnen.

The word cloud contains numerous dialect terms such as: *müpfig*, *Höllgasse*, *Geschleiff*, *Anbrüchig*, *affenähnlich*, *Pflugstößel*, *glücklich*, *hölleheiß*, *Kümmellaiblein*, *Klopfbrauch*, *darnebenheben*, *Streifholzstöcklein*, *Lader*, *Lader-1*, *Kammerlein*, *herumrammeln*, *Amachen*, *estrematuren*, *Rosshändler*, *Abstößerlein*, *Unkrautsämerei*, *Kahne*, *Erni*, *Heuner*, *Butschar*, *Josefitag*, *Ladegang*, *Konsorten*, *Weterscheinen*, *Wölle*, *Agnes*, *Tennenrerig*, *hitschet*, *hittschet*, *einrechnen*, *Gekläpper*, *Dornbusch*, *kindisch*, *Rindenläuf*, *Zöbelhaar*, *Kunststück*, *Graber*, *Schlappe*, *Hasenlorbeeren*, *Stoßkarren*, *Hebkräule*, *Jägerssteig*, *Hollerilkör*, *Stallrinne*, *Schieferkasten*.

Abbildung 9 Suchfelder der BDO

Suche in: Bayerisches Wörterbuch **BWB** Fränkisches Wörterbuch **FWF**
 Dialektologisches Informationssystem von Bayerisch-Schwaben **DTBS**

Suche nach: Optionen: Groß-/Kleinschreibung beachten exakte Suche

Erweiterte Suche

Lemma: <input type="text" value="Brand"/>	Bedeutung: <input type="text"/>	Beleg: <input type="text"/>
Wortfamilie: <input type="text"/>	Grammatik: Mehrfachauswahl möglich <input type="text"/>	Sachgruppe: Mehrfachauswahl möglich <input type="text"/>
Etymologie: <input type="text"/>	Lautung: <input type="text"/>	

Untersuchungsgebiet: Landkreis: Ort:

Abbildung 10 Ausschnitt des Suchergebnisses nach „Brand“

Stichwortliste:
nur aus: **BWB** **FWF** **DTBS**

Suche in Stichwortliste...

-
- A
- B
- C
- D
- E
- F
- G
- H
- I
- J
- K
- L
- M
- N
- O
- P
- Q
- R
- S
- T
- U
- V
- W
- X
- Y
- Z

- brämsen
- Bramser
- bramsig
- brämsig
- brämsig
- Bran
- Branche
- brand
- Brand**
- Brandblater
- Brandbrief
- Brandburger
- Brände
- Brande
- brandeinen
- Brändel
- Brandelein
- Brändelein
- brandelen
- bränd(e)lingen
- brändeln
- brandeln
- brandeln
- branden

Details anzeigen Belege auf Karte zeigen

Brand

Wörterbuch: **BWB** Link zur PDF: Band 3, Spalte 46–52

Wortart: Substantiv
Genus: Maskulinum

Bedeutungen

1 Feuersbrunst, das Brennen

1a Feuersbrunst, Schadenfeuer

Verbreitung **1**

Belege **1**

Brend Kochel TÖL

haind jahrt sô da grâußô Brand Mittich GRI

Dâu hâu't's an Brând BRAUN Gr.Wb. 60

den grozzen schaden, den diu stat genomen hat von dem fewr und von prant 1365 Stadtr.Mchn (Dirr) 478,7f.

wurden sie gleichfals vom Fewr angesteckt, und waren nun alle drey ein lautterer [heller] Brandt MOSER-RATH Predigtmärlein 181

Phraseologie **1**

übertr.: Streit

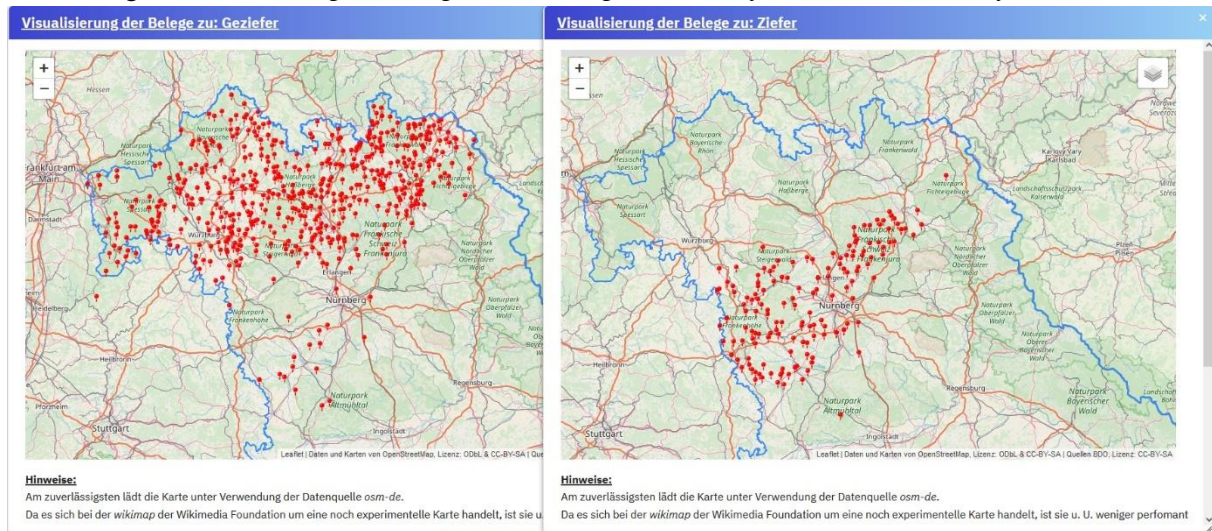
Belege **1**

Abreibung, Zurechtweisung durch Prügel od. Worte

Belege **1**

†Unruhestifter

Abbildung 11 Visualisierung zum Vergleich der Belege von „Geziefer“ mit denen zu „Ziefer“



3. Folgen

Durch eigene Anwendungsentwicklung ergeben sich neben bedeutsamen Vorteilen ggfs. auch Nachteile, die nachfolgend anhand der Erfahrungen in den illustrierten Projekten wiedergegeben werden. Sie haben jedoch keineswegs Anspruch auf Vollständigkeit, da abhängig vom jeweiligen Projekt weitere auftreten könnten.

3.1 Vorteile

3.1.1 Software nach Maß

Der wohl gewichtigste Mehrwert einer individuellen Softwarelösung ist die Tatsache, dass dadurch ein maßgeschneidertes Produkt entsteht, das für optimale Arbeitsabläufe sorgen kann, und damit den Wissenschaftler:innen eine deutliche Zeitersparnis beschert. Im Falle des BWB z. B. beschleunigte sich das Exzerpieren laut Aussagen der Redaktor:innen um ein Vielfaches durch die Verwendung des Redaktionstools *LexHelfer*.

Auch Darstellung und Vernetzung sind so sehr projektnah und damit individuell realisierbar gemäß den Wünschen der Redaktion.

3.1.1.1 Technologie-Mix

Durch eigene Software können Technologien entsprechend kombiniert werden, um ein Maximum an Nutzen und Flexibilität zu erreichen. Ist man z. B. auf die Verwendung von NodeJS oder PHP allein eingeschränkt, lässt sich sicherlich in den meisten Anwendungsfällen dennoch alles damit realisieren. Jedoch ist das Verwenden anderer Technologien für manche Funktionalitäten ggfs. leistungsfähiger oder schlichtweg einfacher zu implementieren. Im Falle von *Bayerns Dialekte Online* wird ein bunter Mix an Technologien verwendet, um die einzelnen Bereiche bestmöglich umzusetzen, der hier nur exemplarisch genannt werden soll, ohne auf dessen Inhalte näher einzugehen: Bash, JavaScript, PHP, Python, TeX, XML, XQuery, XSL und SQL (aufgrund des Exports aus den Quelldatenbanken).

3.1.2 Wiederverwendbarkeit

Bedenkt man in der Konzeption der Software bereits, dass sie wiederverwendbar und erweiterbar sein soll, lässt sich hier ein weiterer Vorteil realisieren, der später große Flexibilität ermöglichen kann und die ursprüngliche In-House-Software auch in ähnlichen Projekten außerhalb der eigenen Institution nutzbar werden lässt. Im Falle des *LexHelfers*, der in seiner ersten Variante für das WBF geschrieben wurde, zeigte sich ferner, dass u. U. auch dann eine Wiederverwendbarkeit gegeben sein kann, wenn diese nicht Teil der Konzeption ist: Die Schnipselgenerierung sowie -darstellung (Raaf 2017) kann auch für das BWB und alsbald auch für das DIBS verwendet werden.

Eine 100%ige Wiederverwendbarkeit in anderen Anwendungsfällen – noch dazu ohne jeglichen Anpassungsaufwand – bleibt jedoch eine Utopie, da dies bedeutete, eine generische Software mit leistungsfähiger künstlicher Intelligenz geschaffen zu haben.

3.1.2.1 FAIR

Ein bezüglich der Wiederverwendbarkeit besonders wichtiger Aspekt sind die FAIR-Prinzipien (vgl. GoFair 2021): Die Daten sollen auffindbar (*findable*), zugänglich (*accessible*), interoperabel (*interoperable*) und wiederverwendbar (*reusable*) sein, um eine möglichst große Nutzbarkeit und Wiederverwendbarkeit auf allen Ebenen sicherzustellen. In eigener Entwicklung kann die Einhaltung dieser Prinzipien von Anfang bedacht und kontrolliert werden, wohingegen sie in Fremdsoftware erfahrungsgemäß oft an zumindest einem dieser Punkte scheitert. Fairerweise muss jedoch erwähnt werden, dass sich dies in den letzten Jahren gebessert hat, zumindest innerhalb der digitalen Geisteswissenschaften: Moderne Web-Anwendungen, die XML-Exporte zumindest von Suchergebnissen anbieten und mit anderen Anwendungen bzw. Repositorien verknüpft sind, erfüllen zunehmend die FAIR-Prinzipien.

3.1.2.2 Offene Lizenz

Die Frage nach der Lizenz ist zwar nicht zwingend von der verwendeten Software abhängig – sie hängt in erster Linie an etwaigen Urheberrechten der verarbeiteten Daten –, jedoch u. U. von dem von der Software verwendeten Datenformat: Proprietäre Software stellt die verarbeiteten Daten womöglich nicht in einem für die Nachnutzung geeigneten Format bereit. Somit bräuchte hier eine offene Lizenz nichts, da nichts unter dieser freigegeben werden kann. Doch auch quelloffene Software ist kein Garant dafür, z. B. einen XML-Download anzubieten oder eine Online-Repräsentation als Permalink unter z. B. CC-BY bereitzustellen. Ist die Anforderung danach nicht in der Software implementiert, steht sie selbstredend nicht zur Verfügung. In eigenen Programmen hingegen kann dies sehr einfach gelöst werden und der Datenbestand mit der gewünschten, möglichst offenen Lizenz versehen sowie dadurch letztlich zur Weiterverwendung angeboten werden.

3.2 Nachteile

3.2.1 Hoher Aufwand

Unabhängig davon, ob die Software als Desktopanwendung auf einem lokalen Arbeitsrechner verwendet werden soll oder es sich um eine Webanwendung handelt, fällt insbesondere zu Beginn ein hoher Entwicklungsaufwand an. Die Verwendung von Frameworks und Open-source-Bibliotheken erleichtert und beschleunigt manches sicherlich, jedoch bleibt die Notwendigkeit bestehen, Zeit und

Arbeitskraft zu investieren. Fertige Software benötigt auf der anderen Seite jedoch meist auch einen hohen Einarbeitungsaufwand, da der Zweck einer solchen Software doch sehr speziell ist – schließlich handelt es sich um ein Nischenprodukt innerhalb der (digitalen) Geisteswissenschaften. Müssen zusätzlich eigene Module entwickelt werden, um Funktionalitäten nachzurüsten, vergrößert sich dieser Aufwand entsprechend, da eine fundierte Einarbeitung in den Aufbau der Fremdsoftware vonnöten ist. Eine zusätzliche Gefahr hinsichtlich des Aufwands stellen eigene Erweiterungen zudem aus folgendem Grund dar: Aktualisierungen der Hauptsoftware können die Kompatibilität der Module zu dieser brechen, sodass hier u. U. nach größeren Updates des Hauptprogramms Anpassungen oder vollständig neue Entwicklungen der Module nötig wären.

3.2.2 Softwarepflege

Die Pflege der Software muss ebenfalls bedacht und sichergestellt werden, damit die Entwicklung nicht nach wenigen Jahren Einsatz aufgrund von fehlenden Updates in ihrer Nutzung eingestellt werden muss (siehe hierzu als Beispiel 2.5). Dies stellt in Zusammenhang mit der allseits bekannten, leidlichen Befristungsproblematik einen doppelt zu bewertenden Nachteil dar, da die Pflege von Fremdsoftware nach Weggang des Entwicklers bzw. der Entwicklerin für gewöhnlich nochmals aufwändiger ist. Unter Einhaltung von Entwicklungsstandards ist jedoch zumindest sichergestellt, dass Freelancer über einen Werkvertrag für kleinere Anpassungen im Rahmen von Updates o.ä. gefunden werden können.

3.2.3 Problematische Daten oder Abläufe

Gibt es potentiell problematische Daten im Projekt, stellen auch diese einen nicht zu unterschätzenden Nachteil dar. Es kann sich dabei um Daten handeln, die erst aufwändig aufbereitet oder konvertiert werden müssen, um in die Software eingespeist werden zu können. Es können jedoch auch solche sein, die vielfach fehlerhaft sind: Ortsdaten z. B. sind erfahrungsgemäß nahezu immer problematisch, da sich die Schreibung des Ortsnamens unterscheidet von der offiziellen Schreibweise, Tippfehler enthalten sind, der Altlandkreis verwendet wurde, oder der Ort gar nicht mehr existiert. Die Lösung dieser beiden Problemarten ist jedoch nicht nur für individuelle Software nötig, sondern viele auch bei vorhandenen Programmen an. Sie stellt somit keinen exklusiven Nachteil für Individualsoftware dar, sollte jedoch unbedingt auch bei dieser bedacht werden.

Neben den Daten selbst ist es auch möglich, dass die Arbeitsweise mit dem geschaffenen Tool nicht optimal abläuft und letztlich eine andere produktiver gewesen wäre. Je nach Projektstand ist ein Wechsel ggfs. (zumindest vorerst) nicht mehr möglich. Doch auch dies ist nicht exklusiv für Individualentwicklungen gültig, sondern kann sich auch nach einigen Monaten der Verwendung von Fremdsoftware einstellen.

3.2.4 Abstimmungsschwierigkeiten

Mit selbstgeschriebener Software eröffnet man Anwender: innen die Möglichkeit, sehr viele Wünsche zu äußern zum Layout und Funktionsumfang. Hier mögen sich eventuell auch nicht alle immer einig sein, sodass es zu Abstimmungsschwierigkeiten kommen kann. Bestehende Software eröffnet hier deutlich weniger oder sogar keinen Diskussionsbedarf, da es schlichtweg keine Spielräume für Wünsche gibt. Im Falle der bayerischen Dialektwörterbücher konnten die Abstimmungsschwierigkeiten durch kurze Wege und die durchweg gute Kommunikation bisher sehr gut gelöst werden. Lediglich im

LexHelfer des DIBS gibt es vereinzelt Probleme beim Export nach XML, sowie in *Bayerns Dialekte Online* vereinzelt in der Transformation der Datensätze nach HTML. Zwar konnten sie durch Workarounds abgefangen werden, jedoch verkompliziert das den Export bzw. die Transformation an manchen Stellen doch sehr und müsste daher in naher Zukunft durch eine bessere Klärung des Arbeitsablaufes gelöst werden.

4. Fazit

Die genannten Beispiele aus der germanistischen Variationslinguistik zeigen auf, wann und warum individuelle Softwareentwicklungen gegenüber bereits bestehender Software zu bevorzugen sind und worin hierbei die Vor- und Nachteile liegen können. Es wird ebenfalls dargestellt, dass fertige Programme die Nachteile in aller Regel nicht abfangen könnten, sodass sie nicht im logischen Umkehrschluss gegen die Entwicklung von In-House-Software sprechen.

Vor- und Nachteile müssen hierbei immer projektspezifisch abgewogen werden.

Trotz des Plädoyers für Individualentwicklungen hat dieser Artikel keineswegs das Ziel, bereits bestehende Softwarelösungen innerhalb der digitalen Geisteswissenschaften grundsätzlich als ungeeignet darzustellen. Deckt Fremdsoftware die eigenen Bedürfnisse vollends ab, spricht nichts dafür, das Rad neu zu erfinden. Erfahrungsgemäß ist diese Abdeckung allerdings selten gegeben, sodass der sicherere Weg der der Eigenentwicklung ist.

Bibliographie

- Arnold, Eckhart 2019– : *Beschreibung der MLW-Notation*. URL: <http://purl.badw.de/zy9Hmx#4-die-notation>, Stand 22.06.2021
- Arnold, Eckhart 2020: *Introduction to DHParse*r. URL: <https://gitlab.lrz.de/badw-it/DHParse/r/-/blob/master/Introduction.md>, Stand 22.06.2021
- Arnold, Eckhart 2021: Ein neues Back-Office für das Mittellateinische Wörterbuch (MLW). In: Bayerische Akademie der Wissenschaften (Hg.): *Jahrbuch 2020*. München: Bayerische Akademie der Wissenschaften. 63-65. URL: <http://purl.badw.de/2xfvSh#63>, Stand 22.06.2021
- Bański, Piotr, Bowers, Jack and Erjavec Tomaž 2017: TEI-Lex0 Guidelines fort he Encoding of Dictionary Information on Written and Spoken Forms. In: Kosem, Iztok, Tiberius, Carole, Miloš Jakubiček, Kallas, Jelena, Krek, Simon and Baisa Vít (Hg.): *Electronic lexicography in the 21st century. Proceedings of the eLex 2017 conference*. Brno: Lexical Computing CZ s.r.o.
- Bayerisches Wörterbuch 2020: *Index des Bayerischen Wörterbuchs*. München: Bayerische Akademie der Wissenschaften. URL: <https://publikationen.badw.de/de/bwb>, Programmierung und Gestaltung von Stefan Müller und Daniel Schwarz, Stand 10.06.2021
- BBAW (Hg.) 2020: ediarum. URL: <https://www.ediarum.org>, Stand 29.06.2021
- BMBF 2016: Freier Zugang schafft mehr Wissen. Pressemitteilung. URL: <https://www.bmbf.de/de/freier-zugang-schafft-mehr-wissen-3340.html>, Stand 10.06.2021
- DARIAH-ERIC Working Group 2017: TEI Lex-0. URL: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>, Stand 10.06.2021
- DFG 2020: Digitaler Wandel in den Wissenschaften. Impulspapier. URL: <https://doi.org/10.5281/zenodo.4191345>, Stand 10.06.2021

- Funk, Edith, Raaf, Manuel, Schwarz, Brigitte and Welsch, Ursula 2020: „Dialektologisches Informationssystem von Bayerisch-Schwaben (DIBS)“. In: Lenz, Andrea and Stöckle, Philipp (Hrsg.): *Germanistische Dialektlexikographie im 21. Jahrhundert*. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik, Beiheft), 105-142
- GoFair 2021: *Fair Principles*. URL: <http://www.go-fair.org/fair-principles/>, Stand 29.06.2021
- König, Almut and Raaf, Manuel 2020: Das Fränkische Wörterbuch (WBF). In: Lenz, Andrea and Stöckle, Philipp (Hrsg.): *Germanistische Dialektlexikographie im 21. Jahrhundert*. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik, Beiheft), 77–104.
- Liquan Luo 2018: *Why Is the Human Brain So Efficient?*
URL: <https://web.archive.org/web/20210629080632/https://nautil.us/issue/59/connections/why-is-the-human-brain-so-efficient>, Stand 10.06.2021
- LRZ (Leibniz-Rechenzentrum) 2019: *GeRDI*. Projektbeschreibung.
URL: <https://web.archive.org/web/20210601123054/https://www.lrz.de/forschung/projekte/forschung-daten/GeRDI/>, Stand 10.06.2021
- Mehler, Alexander and Gleim, Rüdiger 2016: *eHumanities* Desktop. Projektwebseite.
URL: <https://www.texttechnologylab.org/applications/ehumanities-desktop>, Stand 10.06.2021
- Raaf, Manuel 2016a: *LexHelfer BWB*. Online-Datenbank der Forschungsprimärdaten des Bayerischen Wörterbuchs. URL: <https://lexhelfer.bwb.badw.de>, Stand: 10.06.2021
- Raaf, Manuel 2016b: *LexHelfer WBF*. Online-Datenbank der Forschungsprimärdaten des Fränkischen Wörterbuchs. URL: <https://lexhelfer.wbf.badw.de>, Stand: 10.06.2021
- Raaf, Manuel 2017: Precise Annotation of Questionnaires for Dialect Research: The Bavarian Dictionary and its Digitization. In: Kosem, Iztok, Tiberius, Carole, Miloš Jakubíček, Kallas, Jelena, Krek, Simon and Baisa Vít (Hg.): *Electronic lexicography in the 21st century. Proceedings of the eLex 2017 conference*. Brno: Lexical Computing CZ s.r.o.
- Raaf, Manuel 2019: *BayDat v2*. Die bayerische Dialektdatenbank. URL: <https://baydat.badw.de>, Stand: 10.06.2021
- Raaf, Manuel 2020: *LexHelfer DIBS*. Online-Datenbank der Forschungsprimärdaten des Digitalen Informationssystems zu Bayerisch-Schwaben. URL: <https://lexhelfer.dibs.badw.de>, Stand: 10.06.2021
- Raaf, Manuel 2021: *BDO – Bayerns Dialekte Online*. URL: <https://bdo.badw.de>, Stand: 10.06.2021
- Rodriguez-Ramos, Jaimie 2018: *Brains vs. Computers*.
URL: https://web.archive.org/web/20210629080518if_/https://becominghuman.ai/brains-vs-computers-f769548010f1, Stand 29.06.2021
- Schnabel, Michael, Raaf, Manuel and Schwarz, Daniel 2020: Bayerisches Wörterbuch (BWB). In: Lenz, Andrea and Stöckle, Philipp (Hrsg.): *Germanistische Dialektlexikographie im 21. Jahrhundert*. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik, Beiheft), 47–76.
- Schwarz, Brigitte 2013: *Dialektwörterbuch von Bayerisch-Schwaben*. Vom Allgäu bis zum Ries. Augsburg: Wißner
- TextGrid Konsortium. 2006–2014: TextGrid: Virtuelle Forschungsumgebung für die Geisteswissenschaften. Göttingen: TextGrid Konsortium. URL: <https://textgrid.de>, Stand 10.06.2021
- Volkswagenstiftung 2021: *aufbruch*. Die neue Förderstrategie der Volkswagenstiftung. URL: [https://www.volkswagenstiftung.de/sites/default/files/downloads/Aufbruch%20Broschuere online 2.pdf](https://www.volkswagenstiftung.de/sites/default/files/downloads/Aufbruch%20Broschuere%20online%202.pdf), 10.06.2021

Zimmermann, Ralf 2006: BAYDAT- Die bayerische Dialektdatenbank. In Schwitalla, Johannes, Stahl, Peter, Wegstein, Werner, and Wolf, Norbert R.: *Würzburger elektronische sprachwissenschaftliche Arbeiten*; Bd. 1. Würzburg: Universitätsbibliothek Würzburg