

Eckhard Eggers, Göttingen

Datenbankbasierte Publikationen in der Onomastik

1 Vorbemerkung

Ein wichtiger Aspekt der onomastischen Forschung ist die Dokumentation von Namen in Form von umfangreichen Sammlungen, die zumeist in streng formalisierter, wörterbuchähnlicher Anordnung als gedrucktes Werk publiziert werden. Neben dem Namen selbst, der als Kopf des Eintrages fungiert, werden Informationen wie zum Beispiel die Lokalisierung bei Toponymen, die Beleggeschichte oder auch die Etymologie bereitgestellt. Weitere Informationen können hinzutreten: zum Beispiel eine Angabe über berühmte Träger des betreffenden Namens (in einem Familiennamenwörterbuch) oder Informationen zu grammatischen Merkmalen und Besonderheiten. Kennzeichnend ist, dass die Anordnung und im Idealfall auch die interne Struktur aller dieser Informationseinheiten immer einem festgelegten Schema folgt.

Als Beispiel für eine solche Struktur seien die Reihen *Hydronymia Germaniae* und *Hydronymia Europaea* angeführt. Dort liegen folgende Informationseinheiten vor: Lemma (= Name des Gewässers), Lokalisierung auf der zugrunde gelegten topographischen Karte, kurze Beschreibung des Verlaufs oder der Lage des Gewässers, Varianten des Namens, andere Namen für das gleiche Gewässer, chronologisch geordnete Belege – bestehend jeweils aus Jahresangabe, Beleg mit hinreichendem Kontext und Herkunftsangabe (Quelle, Bandzahl, Seitenzahl), Etymologie (nur *Hydronymia Europaea*), bibliographische Angaben zu Werken, in denen der Name des Gewässers besprochen wird.

Derartige Datensammlungen entsprechen bereits in ihrer Anlage einer typischen Datenbankstruktur, obwohl in der Praxis die Da-

ten vielfach von den Bearbeitern über eine Textverarbeitung eingegeben werden. Dieses Textfile¹ liefert dem Verlag die Grundlage für den Druck, wobei in der Regel der Text jedoch von einer Setzerei neu gesetzt wird, um auf diese Weise den qualitativen Ansprüchen des Verlags Rechnung zu tragen und die typographischen Standards zu erfüllen.

Im Folgenden soll am Beispiel der konkreten Arbeit an den Reihen *Hydronymia Germaniae* und *Hydronymia Europaea* gezeigt werden, wie sich die bisherige Arbeitsweise durch einen datenbankbasierten Ansatz ersetzen lässt und welche neuen Möglichkeiten sich daraus ergeben. Auf dieser Grundlage sind die in den Jahren 2001 bis 2006 erschienenen neun Bände der zwei Reihen erstellt worden, sodass sich die angewandte Technik bereits in der Praxis bewährt hat.

2 Was bedeutet „datenbankbasierte Publikation“?

Bei einem datenbankbasierten Ansatz werden die von den Autoren erfassten Informationen zu einem Namen nicht direkt in ein Textfile, sondern stattdessen in eine Datenbank² eingegeben. Mittels eines Ausgabeprogramms wird dann in einem zweiten Schritt das Druckfile erzeugt. Dies klingt zunächst kompliziert und umständlich, bietet jedoch eine Reihe von Vorteilen, die kurz vorgestellt werden sollen.

3 Vorteile einer datenbankbasierten Lösung

- (1) Die Gestaltung und das Layout³ des späteren gedruckten Werkes sind hier prinzipiell unabhängig von den Daten. Daher ist es für die Bearbeiter unnötig, sich in die Feinheiten einer komplizierten Textverarbeitung einzuarbeiten und ihre Eingaben immer genau mit denen anderer Bearbeiter abzustimmen.

- (2) Das Layout kann sogar zu einem ganz späten Zeitpunkt des Projektverlaufes noch komplett geändert werden, ohne dass irgendwelche manuellen Änderungen in einem Eingabetext notwendig sind. In der Anfangsphase eines Projektes lassen sich – sobald eine ausreichende Datenmenge vorliegt – leicht verschiedene Anordnungen und Textgestaltungen testen und vergleichen.
- (3) Wesentliche Vorteile ergeben sich dann, wenn mehrere Projekte – die in der Regel in ihren jeweiligen Publikationen unterschiedliche Layouts verwenden – in einer späteren Phase zusammgeführt werden sollen. Derartige Vernetzungen sind zum Beispiel denkbar, wenn alle (zunächst separat erhobenen) Daten eines geographischen Raumes in einem übergreifenden Projekt koordiniert werden sollen: Es könnten zum Beispiel alle Orts-, Flur- und Gewässernamen eines Bundeslandes in einem gemeinsamen topographischen Namenbuch zusammengefasst werden.
- (a) Da das Layout des Druckes nicht durch die Daten selbst, sondern durch ein nachgeschaltetes Ausgabeprogramm festgelegt wird, kann für das Gesamtprojekt ein einheitliches Layout gewählt werden, ohne dass die Basisdaten neu formatiert werden müssen.
- (b) Selbst für Arbeitsausdrucke aus verschiedenen Projekten kann ein einheitliches Layout hergestellt werden und es können den Bearbeitern des neuen Projektes gezielt nur die Informationen gegeben werden, die sie gerade für ihre Arbeit benötigen. Auf diese Weise lassen sich Arbeitsausdrucke sehr übersichtlich gestalten.
- (4) Auch bei der ersten Eingabe der Daten bietet eine Datenbank Vorteile gegenüber einem Textfile:
 - (a) Die Bearbeiter müssen keine alphabetische Reihenfolge mehr einhalten, sondern können quellenorientiert arbeiten. Bei der Arbeit am Sonderband zu den baltischen Ortsnamen in Ostpreußen hat sich dies als sehr nützlich erwiesen. Hier muss in

jedem Fall zusätzlich zum Lemma in seiner deutschen Form noch der russische oder polnische Name des Ortes angegeben werden. Von diesem wird auf den deutschen Namen zurückverwiesen. Bei strenger alphabetischer Anordnung müsste also die entsprechende Stelle gesucht werden, an der der russische Name einzuordnen ist, dort müsste der Verweis geschrieben werden und man müsste wieder zum deutschen Lemma zurückkehren. Jetzt kann der Verweis an beliebiger Stelle platziert werden, die Sortierung erfolgt später durch das Ausgabeprogramm.

(b) Es besteht durch die Datenbank für die Bearbeiter jederzeit die Möglichkeit, Auszüge aus dem Gesamtdatenbestand herzustellen. So lassen sich zum Beispiel alle Einträge zusammenstellen, die in einer bestimmten Quelle auftreten oder die eine bestimmte Etymologie aufweisen. Dies kann in schwierigen Fällen Entscheidungshilfen bei der Erstellung eines neuen Eintrages liefern.

4 Wie muss eine geeignete Datenbank beschaffen sein?

Das wichtigste Kriterium einer solchen Datenbank ist ihre Flexibilität. Der Umfang der Informationen kann von Eintrag zu Eintrag sehr stark schwanken. Am deutlichsten zeigt sich dies an der unterschiedlichen Anzahl an Belegen oder dem unterschiedlich groß zu wählenden Kontext der Belege. Auch die Länge der etymologischen Angabe variiert relativ stark. Zudem können Informationseinheiten auch leer sein. Man denke zum Beispiel an einen auf der topographischen Karte nachgewiesenen Fluss, zu dem es aber keine historischen Belege gibt.

Bei der Vernetzung von Projekten kommt hinzu, dass gewisse Informationen in einem Projekt auftreten, in einem anderen aber nicht vorkommen. So unterscheiden sich die Reihen *Hydronymia Germaniae* und *Hydronymia Europaea* dadurch, dass im ersten

Fall aus projekthistorischen Gründen keine etymologischen Angaben erfolgen, im zweiten Falle aber sehr wohl. Eine geeignete Datenbank muss mit solchen strukturellen Unterschieden umgehen können.

Aus dem Gesagten folgt, dass Datenbanken, in denen Felder (Informationsbereiche) nur eine vorher schon festzulegende Länge haben dürfen, eher ungeeignet sind. Ebenso problematisch sind Datenbanken, in denen alle Einträge die gleichen Felder enthalten müssen. Beide Eigenschaften treffen auf weit verbreitete Windows-Datenbanken, wie zum Beispiel MS-Access, zu.

Im Hydronymie-Projekt wurde stattdessen die Datenbank Ask-Sam 5 (für MS-Windows) eingesetzt. In ihr ist es möglich, die Feldgrenzen nach Belieben auszudehnen. Es steht also auf jeden Fall genug Platz zur Verfügung, egal wie viele Belege am Ende vorliegen oder wie lang die Kontexte gewählt werden müssen. Auch ist es möglich, Datenbanken mit unterschiedlichen Feldern zusammenzuführen. Ob dies ratsam ist, muss allerdings im Einzelfall entschieden werden. Die Daten lassen sich aber auch in eine neue Datenbank überführen, die einheitliche Strukturen aufweist.

Die Eingabe der Daten ist einfach und erfolgt über eine vorher zu entwerfende Eingabemaske. Diese enthält bereits alle Felder und die Daten werden wie in einer ganz normalen Textverarbeitung in diese Felder eingetragen. Aufgrund der erwähnten Flexibilität verschiebt sich das Feldende während der Dateneingabe, so dass stets genügend Platz vorhanden ist. Für die Bearbeiter, die an die Eingabe mittels Textverarbeitung gewöhnt sind, bestehen daher auch so gut wie keine Einarbeitungsprobleme.

5 Wie wird das Druckfile erzeugt?

Um diese Frage zu beantworten, ist es sinnvoll, zunächst zu klären, welche Anforderungen erfüllt sein müssen: Das Druckfile muss ohne weitere Setzer-Arbeit druckfertig sein (einschließlich einer

Kopfzeile, in der das erste und letzte Lemma der Seite angegeben wird), und es soll dabei allen typographischen Standards genügen und ein hochwertiges Druckbild liefern.

Am besten werden diese Ziele nicht mit einer Textverarbeitung, sondern mit einem Satzsystem erreicht. Im Hydronymia-Projekt wurde das an amerikanischen Universitäten entworfene, mittlerweile weltweit weiterentwickelte und kostenlos erhältliche Softwarepaket LaTeX mit dem Textprozessor PDFLaTeX verwendet.

Im Gegensatz zu einer Textverarbeitung wie MS-Word handelt es sich nicht um ein WYSIWYG-System („What you see is what you get“), sondern um ein befehlsgesteuertes Satzsystem. Wenn in Word ein Wort kursiv gesetzt werden soll, so wird es markiert und der Kursivbutton wird angeklickt. Im selben Augenblick erscheint das Wort bereits auf dem Bildschirm kursiv. In LaTeX arbeitet man dagegen mit zwei Files: einem sogenannten Eingabefile und einem durch den Textprozessor PDFLaTeX daraus erzeugten, fertig formatierten Druckfile im .pdf-Format. Im Eingabefile wird explizit durch einen Befehl angegeben, wie die Formatierung aussehen soll (daher nennt man LaTeX auch eine „Auszeichnungssprache“). Im Falle einer Kursivsetzung wird das entsprechende Wort in den Klammerausdruck `\textit{...}` eingeschlossen. „Dieses Wort ist `\textit{kursiv}`“ im Eingabefile liefert also im Ausdruck das Ergebnis „Dieses Wort ist *kursiv*“. Auch die im Text vorkommenden Sonderzeichen werden (statt wie in Word über eine Zeichentabelle) in Form eines Befehls eingegeben: der Eingabe „`\v{s}`“ entspricht zum Beispiel im Ausdruck das Zeichen „š“. Diese Zweiteilung mutet auf den ersten Blick umständlich an – sie hat aber einige Vorteile, die im Folgenden kurz angesprochen werden sollen.

Am wichtigsten ist hier sicher die langfristige Datensicherheit. Auszeichnungssprachen arbeiten mit dem einfachsten ASCII-Zeichensatz und verzichten völlig auf programmspezifische interne Zeichencodierungen. Damit ist sichergestellt, dass die Daten nicht unlesbar werden, sobald eine neue Programmversion erscheint – so wie dies bei älteren Word-Versionen schon jetzt der Fall ist.

Ein anderer Vorteil ist die leichte Überführbarkeit der Daten von einem Format in ein anderes (zum Beispiel von LaTeX nach HTML oder XML, die ebenfalls Auszeichnungssprachen sind) mittels einfacher Ersetzungsprogramme. Auch dieses Merkmal dient indirekt der langfristigen Datensicherheit, zum anderen erlaubt es, die Daten auch im Internet zu präsentieren oder Hypertextsysteme aufzubauen.

Ein wesentlicher Nachteil des Satzsystems LaTeX soll aber nicht verschwiegen werden: Um so komplexe Strukturen wie einen Eintrag in einem Namenwörterbuch zu gestalten, bedarf es eines sehr aufwändigen Eingabetextes. Diesen manuell einzugeben, würde sehr schnell die Geduld des Schreibers überfordern. Hier übernimmt die Datenbank AskSam 5 eine entscheidende Funktion, indem der gesamte Eingabetext für LaTeX über das bereits erwähnte Ausgabeprogramm erzeugt wird. Solange am Layout nichts geändert werden soll, ist es nur einmal zu Projektbeginn nötig, ein derartiges Ausgabeprogramm zu entwerfen.

6 Wie verläuft die Korrekturphase in einem datenbankbasierten Projekt?

Alle Arbeiten von der Entgegennahme der Manuskripte bis zur druckfertigen Ausgabe wurden im Hydronymia-Projekt in der Göttinger Arbeitsstelle durchgeführt. Auch die notwendigen Korrekturen, die die Bearbeiter der jeweiligen Bände wie traditionell üblich, auf einem Papierausdruck der ersten Version des Druckfiles vorgenommen hatten, wurden hier eingearbeitet. Dies erfolgte allerdings nicht im Textfile, sondern in der Datenbank. Danach wurde erneut ein aktuelles Druckfile erzeugt, um an den geänderten Stellen die erfolgreiche Durchführung der Korrekturen zu überprüfen. Dieses Vorgehen war möglich, weil LaTeX ausgesprochen stabile Ergebnisse liefert: Es ändert sich tatsächlich nur die korrigierte Stelle (und die allernächste Umgebung), nicht aber die

Formatierung anderer Einträge, und sobald der gleiche Zeilenumbruch wie bei der nicht korrigierten Version erreicht ist, gibt es keine Veränderungen bis zum nächsten korrigierten Fehler.

Die Durchführung der Korrekturen in der Datenbank statt im Ausgabertext bedeutet aber, dass die Basisdaten in jeder Phase des Projektes in korrekter, dem Druck entsprechender Fassung verfügbar bleiben. Dies ist ein großer Unterschied zu der Situation, die entsteht, wenn die Erstellung der Druckfassung einer Setzerei übertragen wird. Von dort werden ebenfalls Korrekturfahnen an die Bearbeiter verschickt und die Korrekturen werden auf Papier vorgenommen – dann aber gehen die Korrekturen in die Setzerei zurück. Nur dort, und nicht innerhalb des Projektes, liegen dann die korrigierten Daten vor. Da die Setzereien aber mit sehr kostspieligen Programmen, wie früher im Falle des Hydronymia-Projektes mit QuarkXPress, arbeiten, die zudem schwer in andere Datenformate zurückzukonvertieren sind, stehen nach der Korrekturphase die Daten so gut wie nicht mehr in aktueller Form elektronisch zur Verfügung. Dies ist ein ganz bedeutender Unterschied, der leider oftmals nicht genügend Beachtung findet.

Ein sehr angenehmer Nebeneffekt der in Göttingen praktizierten Datenbanklösung ist eine erhebliche Reduzierung der Druckkosten, verbunden mit einer deutlichen Zeitersparnis.

7 Perspektiven

Die Tatsache, dass die Daten in jeder Phase in der aktuellsten korrigierten Form vorliegen, eröffnet zahlreiche Möglichkeiten der weiteren Nutzung, die hier nur angedeutet werden können.

Die Option einer Vernetzung mit anderen Projekten ist bereits oben skizziert worden. Daher sollen abschließend zwei andere Formen der Weiternutzung angesprochen werden: die – auch bereits in der Praxis durchgeführte – Erstellung von Indexbänden und die Einbringung von zusätzlichen, nicht gedruckten Informationen in

die Datenbank, die für eine spätere philologische Auswertung genutzt werden können.

Das Ausgabeprogramm der Göttinger Datenbank wurde von Beginn an so ausgelegt, dass zu jedem Eintrag bereits die Befehle für eine automatische Indexerzeugung generiert wurden. Diese bleiben für die Leser des gedruckten Werkes unsichtbar, sind aber für den Fall, dass ein Index erstellt werden soll, jederzeit verfügbar. Dies wurde bereits für den 2006 erschienenen Index zu den Sonderbänden der Reihe *Hydronymia Europaea*, der einen alphabetischen und einen rückläufigen Teil umfasst, in der Praxis genutzt. Ebenfalls datenbankbasiert ist der ein Jahr zuvor erschienene Gesamtindex zur Reihe *Hydronymia Germaniae*. Beiden Indizes ist eine CD-ROM mit einer Hypertextversion beigelegt.

Ebenfalls der philologischen Auswertung der Daten dient die Möglichkeit, dass in einer Datenbank in zusätzlichen Feldern, die nicht ausgedruckt werden, ergänzende Informationen abgelegt werden können. So ist es zum Beispiel möglich, ein neues Feld hinzuzufügen, in welchem – am besten nach einem hierarchischen numerischen Schlüssel – die Bildungsweise eines Namens codiert gespeichert werden kann. Durch die Sortierung der Daten nach diesem Schlüssel erhält man für Arbeitszwecke den betreffenden Band nach Bildungsweisen geordnet. Natürlich ist es sinnvoll, für diesen Zweck auch die Klassifikation sichtbar zu machen, aber möglicherweise ist es unnötig, alle Belege oder die Lokalisierung zu haben. Diese können dann durch eine kleine Modifikation des Ausgabeprogramms ausgeblendet werden, sodass sich ein übersichtlicherer Arbeitsausdruck ergibt.

Die hier genannten Möglichkeiten sind nur ein kleiner Ausschnitt aus dem Spektrum an zusätzlichen Optionen, die der Einsatz einer Datenbank eröffnet.

Anmerkungen

- 1 Der Ausdruck „File“ wird hier weitgehend synonym zu „Datei“ verwendet.
- 2 Im Gegensatz zu einem Textfile, das einen fortlaufenden Text enthält, ist eine Datenbank hierarchisch strukturiert. Sie besteht aus sogenannten „Datensätzen“ oder „Einträgen“, die ihrerseits wiederum eine Struktur – die sogenannten „Felder“ – aufweisen. Angewandt auf ein Gewässernamenbuch bildet die Gesamtheit aller aufgenommenen Informationen zu dem FIN X-Bach einen Datensatz. Die Felder dieses Datensatzes sind die verschiedenen aufgenommenen Informationseinheiten: Name des Flusses, Lokalisierung, Belege, Etymologie usw.
- 3 „Layout“ bezeichnet hier das endgültige Aussehen einer Seite im gedruckten Werk.

Summary

The documentation of names in large printed collections (in form of highly structured articles including the name's history, its etymology and variation etc.) has always been an important task of onomastic research. This paper presents a new, more effective way of publishing in onomastics, consisting of a combination of a database (AskSam 5) and a typesetting system (LaTeX). The input of the typesetting system is generated by an output program of the database. Some advantages of this solution are discussed: the independence of the data from the layout in the printed book, which allows to change the layout whenever needed (e.g. when combining different projects), and a higher data security by using basic ASCII code in the database combined with a markup language for professional typesetting. All corrections are made in the database. This keeps the database in accordance with the correct, finally printed text and opens new perspectives in using the data for additional projects, e.g. for indexes or other forms of publication (e.g. popular or short versions). All steps of this new approach are presented: the required features of the database, how to produce the printfile, and how to make the corrections. A detailed discussion of the advantages and the resulting savings of money and time aims to encourage the readers to follow our solution.