

HUMANITIES COMPUTING

»Der Neue Pauly«: EDV-gestützte Redaktion und Produktion einer Enzyklopädie

*Hubert Cancik, Matthias Kopp (Tübingen)**

1. »Der Neue Pauly«

A. Eine Enzyklopädie der Antike

a) Der Neue Pauly ist ein Lexikon der 'klassischen' antiken Kulturen des Mittelmeerraums (ca. 800 v. - 800 n. Chr.), ihrer mediterranen und altorientalischen Vorgänger und Nachbarn, ihrer vielseitigen Wirkung und wissenschaftlichen Erforschung bis in die Gegenwart. Das Werk enthält etwa 23.000 Stichwörter, verteilt auf etwa 25 Fachgebiete und Betreuer für etwa 750 Autorinnen und Autoren. (Im Vergleich dazu: Der 'große' Pauly-Wissowa (1894-1978) enthält in 83 Bänden etwa 120.000 Lemmata.) Allein im Hinblick auf diese Zahlen war die elektronische Verarbeitung aller Daten - Verwaltung, Texte, Zeichnungen, Druck - die Voraussetzung aller Planung. Nachdem der erste von 15 Bänden erschienen (1996), der zweite im Druck ist (Mai 1997), mag ein kurzer Erfahrungsbericht willkommen sein.

b) Der Neue Pauly ist kein Forschungsunternehmen, sondern ein Arbeitsmittel für alle Gebiete der Antike und Antikerezeption. Die Benutzer sind Lehrer, Altertumswissenschaftler, Theologen, Journalisten, Literaturwissenschaftler, Philosophen. Gegenüber dem bewährten »Kleinen Pauly« (herausgegeben von K. Ziegler, W. Sontheimer, H. Gärtner in 5 Bänden, 1964-75) seien - außer der Verdoppelung des Umfangs - folgende Unterschiede hervorgehoben:

1. Erweiterung der Zeitgrenze auf ca. 800 n. Chr.;
2. Verstärkte Beachtung des alten und jungen Orients;
3. Aufnahme der Byzantinistik;
4. Gleichstellung der verbalen, optischen, dinglichen Quellen;
5. Ausbau der philosophischen Begriffsgeschichte;
6. Erweiterung der Wirtschafts-, Sozial-, Alltagsgeschichte;

* Protokoll des 69. Kolloquiums über die Anwendung der EDV in den Geisteswissenschaften an der Universität Tübingen am 8. Februar 1997.

7. Aufnahme der Wirkungsgeschichte;
8. Aufnahme der Wissenschaftsgeschichte;
9. Gesamtregister;
10. Abbildungen, Schemata, Skizzen, Karten (keine Photographien).

c) Der neue Pauly bleibt ein Reallexikon. Die Stichwörter bleiben überwiegend kleinteilig, eher objektsprachlich, personen- und textnah. Die Kurzlemmata werden nicht von unübersichtlichen Dachartikeln verschluckt: einfacher Zugang zu den Grundinformationen, den Namen, Orten, Zahlen, Sachen.

Das Zentrum bleibt das 'klassische Altertum', die hellenische und römische Kultur in der Bronze- und Eisenzeit in allen ihren Ausprägungen. Aber das Feld, in dem dieses Zentrum lagert - »La Mediterranee« -, die orientalischen Voraussetzungen beider Kulturen, ihre Substrate und ihre Wirkung auf Kelten, Germanen, Slawen, Araber, auf Judentum und Christentum müssen detailliert bestimmt und neu gewichtet werden.

d) Dieser Aufgabenstellung entsprechend sind die Anzahl der Lemmata im altertumswissenschaftlichen Teil relativ groß, die Artikel relativ kurz. Die Anzahl der Mitarbeiter ist, infolge des enormen Spezialisierungsdrucks, in allen Disziplinen sehr hoch. Es gibt zahlreiche von mehreren Autoren verfaßte 'Komposit-Artikel', da sich für übergreifende Themen oder 'Dachartikel' kaum noch 'Generalisten' finden lassen. Die Einheitlichkeit der Konzeption eines Artikels - um von dem Werk im ganzen zu schweigen - ist hierdurch gefährdet.

B. 'Real-Lexikon'

Im Verlag J.B. Metzler (Stuttgart) erschien in den Jahren 1839 (1837) - 1852 »Die Realencyclopädie der classischen Altertumswissenschaften in alphabetischer Reihenfolge«. Ihr Archeget ist August (von) Pauly (1796-1845). Der Typus dieses Reallexikons läßt sich weit ins 18. Jh. zurückverfolgen (vgl. Benjamin Hederich (1675-1748), Reales Schullexikon, 1717). Es ergänzt die formalen Hilfsmittel: Wörterbuch, Grammatik, Metrik.

Die Anzahl dieser Realien-Lexika und die Anforderungen an Systematik, Vollständigkeit, Lesbarkeit sind seit Hederich und Pauly ungemein gestiegen: Lexika der Literaten oder Philosophen, der Wissenschaftstheorie, der Aesthetik oder der Musik in Geschichte und Gegenwart. Allein in Tübingen wird an folgenden Projekten gearbeitet: Reallexikon der Germanistik, Handbuch religionswissenschaftlicher Grundbegriffe, Historisches Wörterbuch der Rhetorik, Lexikon der Religion, vierte Auflage der RGG, Prosopographie der stadtrömischen Priesterschaften, dazu Goethe-Wörterbuch und Luther-Konkordanz. Soeben ist der erste Band des Gesamtregisters der großen Realencyclopädie von Pauly-Wissowa-Ziegler erschienen:

Paulys Realencyclopädie der classischen Altertumswissenschaft. (Gesamtregister in zwei Teilen) - Band 1: Alphabetischer Teil. Hrsg. v. Tobias Erler, Christa Frateantonio, Matthias Kopp, Dorothea Sigel, Dorothea Steiner. Stuttgart/Weimar: Verlag J. B. Metzler, Mai 1997; 1158 Seiten, zweispaltig, mit CD-Rom.

Die Fülle dieser Arbeiten ist beachtlich: Ist das ein lokales Phänomen? Oder ein allgemeines Symptom jetziger Wissenschaft? Was für eine Sicht auf Kultur wird durch diese Lexika erzeugt?

2. Vorgaben

Mitte des Jahres 1993 waren die Vorbereitungen für das Projekt »Der Neue Pauly« soweit gediehen, daß ein detailliertes Konzept für die Produktion erforderlich wurde, daß Entscheidungen auch in Hinblick auf die EDV-Organisation getroffen werden konnten und mußten.

Damals stand bereits fest, daß die Arbeit dezentral organisiert werden würde: Die Autorenkommunikation und die fachwissenschaftliche Redaktion der Artikel sollte durch die Fachgebietsherausgeber erfolgen, die zu diesem Zeitpunkt bestimmt waren und nach Verteilung des im geplanten Lexikon verfügbaren Raumes gemeinsam die Gesamtlemmaliste erarbeiteten.

In einer Zentrale sollte das Material der insgesamt neunzehn Fachgebietsredaktionen zusammengefaßt werden, die - neben einigen Standorten an Universitäten in Westeuropa und in den USA - sämtlich an deutschen Universitäten verankert waren. Diese Zentrale sollte in Tübingen, am Philologischen Seminar eingerichtet werden, sollte unter Ausnutzung der örtlichen Facilitäten (Bibliotheken, Zentrum für Datenverarbeitung) wissenschaftliche und organisatorische Betreuung anbieten, Koordinationsleistungen erbringen. Dies betraf zunächst die Unterstützung der Fachgebietsherausgeber bei der bereits erwähnten Erarbeitung der Gesamtlemmaliste, dann die Abwicklung der Verträge mit den Autoren, die Entwicklung von Vorgaben für die Manuskriptgestaltung und schließlich die Organisation, Verwaltung und Kontrolle des resultierenden Manuskriptbestandes. Im Interesse sowohl der zeitlichen Überschaubarkeit wie der inhaltlichen Kohärenz und Vernetzbarkeit wurden die Artikel jeweils in großen Gruppen an einzelne Autoren, jeweils Fachleute auf dem entsprechenden Sektor, vergeben. Die Autoren konnten und können deshalb vom Anfangsbuchstaben »A« bis zum Anfangsbuchstaben »Z« zwischen sämtlichen Artikeln eines Themenkreises Bezüge herstellen. Darüber hinaus war und ist es idealiter für alle Mitarbeiter, Autoren wie Herausgeber, jederzeit möglich, sich über andere, nicht nur alphabetisch frühere Lemmata und deren Autor zu informieren, ggf. Absprachen zu treffen.

Außerdem stand fest, daß der Neue Pauly, anders als seine Vorgänger, mit Hilfe von Werkzeugen der elektronischen Datenverarbeitung organisiert und

produziert werden sollte: Wenn ein Projekt dieser Größe in nützlicher Frist, mit geringem Personalaufwand durchgeführt werden soll, ist dies schon vor dem eigentlichen Satz erforderlich.

Im Laufe der Diskussion wurde schnell deutlich, daß die Daten auf jeden Fall anwendungsneutral gehalten werden sollten. Zu diesem Zeitpunkt fiel das Akronym »SGML« immer häufiger, war in einschlägigen Kreisen der Trend aufgekommen, Daten wenn möglich nicht nur in Hinblick auf ein konkretes, absehbares Ausgabemedium hin zu strukturieren und vorzubereiten, sondern sich qua anwendungsneutraler Datenhaltung auch andere Optionen, die Verwertung in anderen Medien offenzuhalten. Neben der gedruckten Ausgabe sollte für den Neuen Pauly künftig eine Publikation in anderen Ausgabemedien aus technischer Sicht mit geringem Aufwand möglich sein.

In der Tübinger Zentralredaktion mußte die Erfüllung der folgenden Aufgaben durch ein geeignetes EDV-Instrumentarium ermöglicht werden:

a) Organisation

- Führung der Gesamtlemmaliste, die einem ständigen Prozeß der Veränderung, vulgo Verfeinerung, unterworfen war, teilweise auch noch ist
- die Erarbeitung und Kommunikation von Zeitplänen
- der Abschluß von Verträgen
- die Entwicklung einer logischen Strukturdefinition (»DTD«) und, parallel dazu, die Entwicklung von Vorgaben für die Manuskript-Gestaltung für die Autoren.

b) Redaktion im engeren Sinne

- formale, stilistische, inhaltliche Bearbeitung von Manuskripten (Reihenfolge bedeutet Wertung)
- Verwaltung der Manuskript-Daten: Verweisorganisation, Vernetzung, Vorbereitung der Register
- Umfangs- und Terminkontrolle
- Akquisition von Illustrationen (Karten, Abbildungen, Schemata)
- schließlich kam auch noch die Bereitstellung von Abrechnungsdaten hinzu.

3. Realisierung

3.1 Software

Unter den genannten Aufgaben war die Entwicklung einer Strukturdefinition, die alle für die künftige Bearbeitung zu differenzierenden logischen Komponenten der Artikel des Neuen Pauly beschreiben können sollte, die umfangreichste. Mit ihr mußte die Entscheidung für ein Softwareprodukt einhergehen, mit dessen Hilfe diese Struktur anwendungsneutral dargestellt und verwaltet

werden sollte. Diese Entscheidung fiel relativ schnell für TUSTEP, da andere Möglichkeiten zur Realisierung eines SGML-basierten Redaktionssystemes - dies ist die Konsequenz der genannten Aufgabenstellung - nicht in Sicht waren. Gewichtigstes Argument war die Tatsache, daß die anwendungsneutrale Datenerhaltung, die konsequente Orientierung an den benutzerdefinierten, i. d. R. logischen Strukturen von (Text-)Daten in der TUSTEP-Entwicklung und -Anwendung traditionell gefordert und gefördert wird. Weitere Argumente waren: - TUSTEP existiert seit langer Zeit: Man muß also nicht mit Betaversionen und Patches arbeiten, auf Updates warten, hat vielmehr eine Software zur Verfügung, die über ihre Kinderkrankheiten hinaus, die lauffähig und stabil ist. - Das zugrundeliegende Datenmodell und der modulare Aufbau von TUSTEP ermöglichen eine Flexibilität, die etwa beim Dateneingang eine schnelle Bewältigung der vielfältigen Formate ermöglicht. - Entsprechenden Programmieraufwand vorausgesetzt, kann TUSTEP nicht nur an benutzereigene Datenstrukturen angepaßt werden, es bietet auch die Möglichkeit, benutzerdefinierte Abläufe weitgehend automatisiert durchzuführen, d. h. unter maßgeschneiderten Oberflächen, die keine Systemkenntnisse voraussetzen, durchzuführen. - In Tübingen ist einschlägiges knowhow vorhanden; zum einen in der Abteilung für LDDV des ZDV, deren Mitarbeitern wir für bisherige wie künftige Beratung und Unterstützung danken; zum anderen gibt es in Tübingen genügend Leute, die über die erforderliche Doppelqualifikation verfügen, d. h. über altertumswissenschaftliche Kompetenzen und TUSTEP-Kenntnisse, die ausreichen, um die Möglichkeiten des Systems zu nutzen: Da die Zeit für die Einarbeitung in eine komplexe Materie oder, anders formuliert, da qualifizierte Mitarbeiter Kostenfaktoren sind, kommt diesem Argument, der Traditionsgarantie, besondere Bedeutung zu. - Nicht zu unterschätzen ist schließlich auch, daß es bei TUSTEP in dem beim Neuen Pauly relevanten Bereich (griechische Zeichen, Diakritika, linguistische Sonderzeichen) keinerlei Zeichensatzprobleme gibt.

3.2 Vorgaben für die Manuskriptgestaltung

Im Zuge der Autorenwerbung, spätestens mit Abschluß des Autorenvertrages, erhielten die Mitarbeiter des Neuen Pauly eine Broschüre, die neben Informationen zur Zielsetzung des Gesamtwerkes Angaben zur Gestaltung der Manuskripte enthielt.

Diese Angaben konnten nicht zu differenziert sein, da sie zum einen für Artikel unterschiedlichen Typs geeignet sein mußten - der Neue Pauly enthält neben knappen Personen- und Sachartikeln umfangreichere Artikel, die größere Phänomene beschreiben und schließlich die sogenannten Dachartikel, die, das lexikalische Prinzip der objektiven Detailinformation durchkreuzend, zusammenhängende Darstellungen unternehmen, dabei wiederum die in den kürzeren Artikeln gegebene Information voraussetzen. Zum anderen waren diese tech-

nischen Vorgaben für über 700 Autoren bestimmt, die hinsichtlich der Verwendung von Textverarbeitungssystemen und hinsichtlich der Produktion strukturierter Texte im elektronischen Medium unterschiedliche Voraussetzungen hatten.

Um die automatische Weiterverarbeitung der Artikel sicherzustellen, wurden folgende Anforderungen an die Manuskripte bzw. die in Tübingen eingehenden Daten gestellt:

- Artikel sollen jeweils in elektronischer Gestalt und in einem Ausdruck weitergegeben werden; jedes Textverarbeitungssystem, jedes Betriebssystem darf verwendet werden - angesichts der erwähnten großen Bandbreite unter den Autoren wäre hier jede Einschränkung kontraproduktiv gewesen.
- Für die Markierung der logischen Komponenten eines Artikels (Lemma, Text, Bibliographie, Zwischenüberschriften, Verweis . . .) wurden explizite Kennzeichen, sog. Qualifier, festgelegt, auf die nach Konsistenzprüfung automatisch zugegriffen werden konnte.
- Typographische Auszeichnungen (fett / kursiv / Unterstreichung) wurden nur für zwei logische Qualitäten, für Hervorhebungen im Text eines Artikels und für Autorennamen in bibliographischen Angaben, zugelassen.
- Jedem Artikel wurden die erforderlichen Verwaltungsinformationen (Name des Autors, Adresse und Bankverbindung), entsprechend gekennzeichnet, hinzugefügt.

Die Manuskripte werden zunächst vom Autor an die zuständige Fachgebietsredaktion weitergeleitet, gehen von dort nach fach wissenschaftlicher Bearbeitung und Klärung von Rückfragen an die Tübinger Redaktion.

Nachdem nun mit Manuskripten bis zum dritten von fünfzehn Bänden Erfahrungen vorliegen, kann eine außerordentlich positive Bilanz gezogen werden: Die von den Fachgebietsredaktionen bearbeiteten Manuskripte bzw. Daten genügen zu mehr als 90 % den Anforderungen.

3.3 Datenverwaltung

Die in der Tübinger Redaktion eingehenden Daten werden wie folgt bearbeitet: Nach der Konvertierung der Daten aus dem Eingangsformat wird das Vorhandensein der erforderlichen Qualifier bzw. Feldseparatoren angeprüft, fehlende Kennzeichen werden ggf. manuell ergänzt. Nach der Extraktion der Autordaten werden die oben erwähnten, zulässigen typographischen Kennungen in explizite Kennungen umgewandelt, alle übrigen typographischen Merkmale wie Mehrfachblanks, Tabulatoren, Einzüge usw. werden eliminiert. Nach der Ergänzung von Verwaltungsinformationen (eindeutige Nummer, Eingangsdatum, Sollumfang) werden die neueingegangenen Daten in den zentralen Datenbestand integriert und können nun redigiert werden.

Die Redaktionsarbeit in Tübingen betrifft in erster Linie die formale Auszeichnung. Die logischen Komponenten eines Artikels werden differenzierter gekennzeichnet, als dies durch Autor und Fachgebietsredaktion möglich war. Neben anderen Elementen werden etwa Gliederungen, Originalzitate, Quellenreferenzen, Verweisziele, laufende Nummern homonymer Lemmata, antike Werktitel, Ankerpunkte für Abbildungen durch paarige Auszeichnungen markiert, die Verweisziele werden zudem noch durch Angabe einer numerischen Adresse eindeutig gekennzeichnet. Diese Markierungen werden zum einen für die Druckausgabe teils unterdrückt, teils in geeignete typographische Attribute umgesetzt. Zum anderen eröffnet diese Kennzeichnung weitere Möglichkeiten zum Umgang mit den in den Artikeln enthaltenen Informationen: Die differenzierte Kennzeichnung der Gliederungen ermöglicht eine automatische Konsistenzkontrolle, Gleiches gilt für die Verweisziele und die laufenden Nummern; an die Ankerpunkte für Abbildungen werden im Laufe der Produktion Graphikdateien angeschlossen, weitere Kontrollmöglichkeiten, etwa bei den Quellenreferenzen, sind evident.

Die eigentliche Redaktion, d. h. die Bearbeitung eines Artikels aus dem zentralen Datenbestand, erfolgt in einer Redaktionsumgebung, die aus einem für die Zwecke des Neuen Pauly entwickelten System von Makros besteht. Für die redaktionelle Arbeit sind neben fachwissenschaftlichen Kenntnissen folgende Kompetenzen erforderlich:

- Kenntnis der differenzierten Auszeichnungen, Fähigkeit zur Bewegung des Cursors im TUSTEP-Editor und zur Anwendung von Buttons/Editormakros
- generelles Abstraktionsvermögen, um die anwendungsneutrale Bildschirmdarstellung, die Text und sichtbare Strukturmerkmale enthält, mit dem Druckbild vermitteln zu können.

Wichtigstes Merkmal dieser momentan am Großrechner des ZDV realisierten Redaktionsumgebung ist die Mehrplatzfähigkeit: Der gesamte Datenbestand steht jederzeit vollständig zur Verfügung. Der Zugriff auf einen Artikel sperrt diesen temporär für alle anderen Mitarbeiter. Nach Abschluß der Bearbeitung steht der Artikel sofort wieder allgemein zur Verfügung.

Im einzelnen stehen folgende Leistungen zur Verfügung: Nach Abfrage der Zugriffsberechtigung werden zunächst benutzerspezifische Profile eingestellt; damit wird unterschiedlichen Arbeitsgewohnheiten, Vorlieben für die Verwendung von Maus oder Tastenkombinationen Rechnung getragen. Anschließend wird ein zu bearbeitender Artikel ausgewählt und, nach erfolgreicher Abprüfung der Verfügbarkeit, exklusiv zur Verfügung gestellt. Im Editor, der nun aufgerufen wird, stehen zum einen Buttons bzw. Tastenkombination für die Auszeichnung einzelner Elemente zur Verfügung. Zum anderen kann von hier aus auf andere Daten (Liste aller Lemmata, Liste zulässiger Abkürzungen usw.)

zugegriffen werden. Nach Abschluß eines redaktionellen Zugriffes werden die Daten zurückgeschrieben und freigegeben. Protokolle, Sicherheitskopien und rollback-Mechanismen garantieren die Transparenz und die Reproduzierbarkeit sämtlicher Zugriffe auf die Daten.

Schon vor der Produktion eines gedruckten Bandes ist es möglich, den je aktuellen Zustand der redigierten Daten zu dokumentieren. Eine entsprechende Routine transformiert Daten des Tübinger Bestandes in das HTML-Format, in dem sie den Fachgebietsherausgebern zur Ansicht bzw. für Ausdrucke zugänglich gemacht werden können.

Für die Kontrolle der Konsistenz, der Numerierung und der Verweise stehen weitere Routinen zur Verfügung, ebenso für Selektion und Sortierung der Artikel nach unterschiedlichen Kriterien, für die Ermittlung des aktuellen Status, für die Terminkontrolle und die Honorierung.

4. Probleme und Perspektiven

Größtes Problem waren und sind die Termine. Da die Durchsatzraten und die Kapazitäten der Fachgebietsredaktionen teilweise zu optimistisch eingeschätzt worden waren, traten und treten in der Zentralredaktion Spitzen auf. Der Bedarf an qualifizierten Arbeitskräften und qualifizierten Arbeitsplätzen ist hier deswegen höher als erwartet. Dennoch ist die dezentrale Struktur unverzichtbar, ist der mit ihr einhergehende erhöhte Kommunikations- und Kontrollbedarf unvermeidbar. Nur so können eine effiziente Kommunikation mit den Autoren und die erforderliche fachwissenschaftliche Kompetenz und Kontrolle garantiert werden. In den Fachgebietsredaktionen wie in der Zentralredaktion hat sich mittlerweile eine Routine entwickelt, durch die anfängliche Reibungsverluste ausgeglichen werden.

Welche Perspektiven eröffnen die anwendungsneutrale Datenhaltung und das electronic publishing für ein Projekt wie den Neuen Pauly?

Die Atomisierung der Information, die in der lexikalischen Darstellung unvermeidbare Auflösung der Kontexte wurde schon früher als Gefahr wahrgenommen, der etwa durch »enzyklopädische Stichworte« begegnet werden sollte. Der Enthusiasmus, der angesichts der Aussicht auf virtuelle Enzyklopädien, auf »vielfältig durchwanderbare Informationsräume« zunächst ausgebrochen war, ist folgerichtig der Ernüchterung und der Einsicht gewichen, daß die unkontrollierte Vermehrung von nun auch elektronisch zugänglichen Informationseinheiten deren Gebrauchswert nicht erhöht. Angesichts der Vielfalt und der Zufälligkeit von Treffern nach einer Suchanfrage wird deutlich, daß die explosiv zunehmende Zahl verfügbarer Informationen nur dann bewältigt werden kann, wenn diese durch Kontexte und durch konsistente Vernetzungen ergänzt und qualifiziert werden. Die erforderliche intellektuelle Leistung wird häufig unterschätzt.

Für den Neuen Pauly lassen sich vor diesem Hintergrund ein Nah- und ein Fernziel formulieren:

- Kurzfristig muß die Einbindung weiterer Texte sowie anderer Datentypen geplant werden: Die bereits erfolgte Kennzeichnung der Quellenreferenzen ist ein erster Schritt; sie ermöglicht, im elektronischen Medium, die direkte Verknüpfung der Lexikonartikel mit Editionen. Wörterbücher, Handbücher, Bibliographien, abstracts sind weitere Textdaten, die eingebunden werden können. Graphiken, Karten und dreidimensionale Pläne, Abbildungen von Artefakten, Animationen, die etwa geographische Entwicklungen oder archäologische Fundzusammenhänge veranschaulichen, sind neben Audiodaten (z. B. griechische Tragödien) Beispiele für andere Datentypen, mit denen weitere Informationen verfügbar gemacht werden können.

- Langfristig muß die resultierende Datensammlung in Auseinandersetzung mit aktuellen Fragestellungen und mit Hilfe von Ordnungskriterien erschlossen werden, die über das Alphabet oder assoziative Verbindungen hinausgehen. »Guided tours« und andere Mittel, mit denen bewußt wieder eine Sequenz hergestellt und eine mögliche Sicht angeboten wird, können dazu eingesetzt werden.