

Datenbasen als Editionen

Autor(en): **Levermann, Wolfgang**

Objektyp: **Article**

Zeitschrift: **Geschichte und Informatik = Histoire et informatique**

Band (Jahr): **2 (1991)**

PDF erstellt am: **16.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-4264>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Datenbasen als Editionen

Wolfgang Levermann *Max-Planck-Institut für Geschichte, Göttingen*

Wenn über Datenbasen als Editionen gesprochen wird, so lassen sich -- je nach Betonung der beiden Bestandteile des Titels -- zwei Gesichtspunkte unterscheiden. Im Rahmen einer Reflexion über den Nutzen des Computers in der historischen Forschung könnte einerseits gefragt werden, ob der Computer auch Wege für eine verbesserte Repräsentation von historischem Quellenmaterial eröffnet bzw. welche besonderen Vorzüge ein Rechner im Unterschied oder als Ergänzung zu einer gedruckten Edition von Quellen bieten kann. Eine Antwort auf diese Frage würde etwa die Möglichkeit des schnellen und flexiblen Zugreifens auf das Material herausstellen oder die Weiterentwicklung datentechnischer Instrumente im Hinblick auf eine neue Repräsentation von Quellenmaterial betonen, wie sie z.B. unter dem Stichwort der «Dynamischen Edition» diskutiert wird, sie hätte sich aber andererseits auch mit besonderen Problemen wie etwa der Erhaltung schriftspezifischer und graphischer Merkmale von Texten im Rahmen einer datenbankmäßigen Verwaltung solchen Materials zu beschäftigen.¹ Die zweite Interpretation des Titels zielt dagegen auf die Veröffentlichung von Datenmaterial, das nicht mit dem eigentlichen Ziel der Edierung maschinenlesbar gemacht wurde, sondern die Basis einer wissenschaftlichen Untersuchung darstellte, die sich des Rechners als eines Werkzeuges bediente. Auch für dieses Material stellt sich aus verschiedenen Gründen die Frage nach seiner Veröffentlichung und ihr soll im folgenden Beitrag nachgegangen werden.

Historisches Quellenmaterial wird heute in immer größerem Umfang maschinenlesbar gemacht, um es z.B. mit statistischen Verfahren oder mit Methoden und Techniken der Datenbankverwaltung zu erschließen. Obgleich dies zumeist ausgesprochen zweckgerichtet geschieht und nur die wissenschaftliche Analyse der Quellen erleichtern soll, entsteht dabei doch immer auch die Frage nach einer Veröffentlichung des maschinenlesbar gemachten Materials selbst. Da dieses Material letztlich die Basis aller daraus abgeleiteten wissenschaftlichen Thesen und Resultate darstellt, sollte es schon aus Gründen der Überprüfbarkeit der auf diesem Wege erzielten

1 Vgl. in diesem Zusammenhang etwa M. Thaller: *On the Representation of Historical Texts*. Göttingen 1990 (Unveröffentlichtes Arbeitspapier des Projekts «Arbeitsplatzrechner für den Historiker»); ders.: *Secundum Manus*. Zur Datenverarbeitung mehrschichtiger Editionen. In: G. Cerwinka u.a.(Hg.): *Beiträge zur Geschichte und Ihren Grundlagen*. Festschrift Friedrich Hausmann zum 60. Geburtstag. Graz 1987, S.56ff.

Erkenntnisse der wissenschaftlichen Forschung insgesamt zur Verfügung stehen und allgemein zugänglich sein. Dies gilt umso mehr, als mit dem Medium der EDV eine im Prinzip kostengünstige Möglichkeit der Weitergabe von Daten gegeben ist.

Ein weiterer, häufig genannter Gesichtspunkt kommt hinzu: Die Aufbereitung von historischem Quellenmaterial für den Rechner ist in nahezu allen Fällen mit langwierigen und intensiven Anstrengungen verbunden; in wissenschaftlichen Projekten gehört die Phase der Dateneingabe zumeist zu den zeitaufwendigsten des ganzen Projekts. Es ist deshalb in der Vergangenheit auch immer wieder darauf hingewiesen worden, daß der für die Aufbereitung des Quellenmaterials investierte Arbeitsaufwand sinnvollerweise nicht nur dem jeweiligen Projekt, sondern auch anderen an dem Quellenbestand interessierten Wissenschaftlern zugute kommen sollte; das Material sollte ihnen deshalb für eigene Arbeiten und Recherchen zur Verfügung stehen.

Die Forderung nach einer Veröffentlichung maschinenlesbar gemachter Quellen wird dort am nachdrücklichsten zu stellen sein, wo das verwendete Material nicht in gedruckter Form vorliegt und eine gedruckte Edition auch kaum erwartet werden kann. Hier ist eine Veröffentlichung zugleich immer auch ein Stück ursprünglicher Quellenedition, die eine Quelle einem größeren Benutzerkreis erschließt. Aber auch dort, wo gedruckte Editionen existieren, ist eine Veröffentlichung der maschinenlesbaren Version wünschenswert. Denn zum einen können im maschinenlesbaren Material Abweichungen von der gedruckten Edition vorgenommen worden sein, zum anderen erlaubt -- wie oben angedeutet -- eine maschinenlesbare Version des Quellenmaterials eine Berarbeitung dieses Materials mit all den Vorteilen, die die EDV heute für die Verwaltung großer Datenmengen zur Verfügung stellt. Selbst wenn es sich bei der maschinenlesbaren Version «nur» um einen fortlaufenden Text handelt, der einer gedruckten Ausgabe folgt, erlauben heutige Mittel der Volltextrecherche Möglichkeiten, die traditionelle Register oder Konkordanzen nicht nur ergänzen, sondern erheblich übertreffen können.

Wenngleich demnach wohl niemand den Sinn eines solchen Datenaustausches ernsthaft in Frage stellen würde, so blieb dieser Austausch jedoch in der Vergangenheit zumeist ein bloßes Postulat: Insbesondere das «Recycling» von maschinenlesbarem Quellenmaterial, seine Nutzung im Sinne einer «Sekundäranalyse», ist trotz der Bemühungen einschlägiger Daten- und Textarchive weitgehend unverwirklicht geblieben. Die Hürden, die sich der Realisierung eines umfassenden Datenaustausches in den Weg stellen, sind offensichtlich so hoch, daß die Forderung nach Datenaustausch zu einem Lippenbekenntnis verkam.

Die Gründe dafür sind sicherlich vielschichtig und können an dieser Stelle auch nur schlaglichtartig beleuchtet werden. Bis zu einem gewissen Grade sind sie sicherlich auch «historisch»: Traditionell wurde bei der Kodierung historischen Materials stark von der Quelle abstrahiert, wenn nämlich Quellenaussagen in statistische Kodezahlen umgesetzt wurden. Entsprechend erforderte ein Versuch der Wiederverwendung solchen Materials nicht nur ein intensives Kennenlernen der verwendeten Kodierung und ein gleichartig gelagertes Untersuchungsinteresse, sondern vor allem auch Vertrauen in die Korrektheit und Angemessenheit der vorgenommenen Kodierung -- ein Schritt, der vielen Interessenten zu Recht schwerfällt. Denn die Korrektheit der vorgenommenen Kodierung ist am Material selbst nicht zu überprüfen, so daß Fehler, die dabei möglicherweise entstanden sind, nicht oder nur mit erheblichem Aufwand erkannt werden können. Software, die eine «quellengetreue» Darstellung und Verarbeitung von historischem Material erlaubt und für die Zwecke einer statistischen Auswertung zusätzliche Tools bereitstellt, steht dagegen erst allmählich zur Verfügung, sollte aber solcherart Bedenken abbauen helfen.

Damit sind aber allenfalls einige der bestehenden Schwierigkeiten beseitigt, denn auch die quellennahere Form der Verwaltung historischen Quellenmaterials in Gestalt einer Datenbank ist allein noch kein Ausweg. Datenbasen mit strukturierten Daten können z.B. solch komplexe Formen der Strukturierung aufweisen, daß sie für den außenstehenden Benutzer eher abschreckend wirken. In diesen Fällen ist nicht nur intensives Studium der Strukturierungen nötig, um überhaupt erste Anfragen an die Datenbasis stellen zu können, sondern es entstehen berechtigte Befürchtungen, daß dieses Material bereits in zu starkem Maße «bearbeitet» wurde, insofern Strukturierungen ja immer auch inhaltliche Entscheidungen widerspiegeln und im Hinblick auf bestimmte Erkenntnisinteressen vorgenommen worden sind. Sollen Datenbasen daher wirklich wiederverwendbar sein, so sollten sie sowohl für gelegentliche Konsultationen geeignet sein als auch in einer «neutralen» Form zur Verfügung stehen, die sich auf wesentliche Informationsbestandteile der Quelle beschränkt und -- über den Originaltext hinaus -- nur zentrale Aspekte des Materials durch Strukturierung für Abfragen und Analysen verfügbar macht. Dies liefe letztlich auf die Entwicklung und Vereinbarung eines standardisierten Austauschformats für Quellenmaterial hinaus, das allerdings nicht leicht zu realisierende technische und vor allem inhaltliche Implikationen aufweist.²

2 Unter technischem Aspekt wäre etwa an die Verspeicherung von Strukturierungshilfen zu denken, die nicht der konzeptuellen Gliederung des Materials dienen, sondern seine spätere Anpassung an jeweils spezifische Untersuchungsziele optimieren sollen. Schwieriger als in die Realisierung derartiger zusätzlicher Hilfsmittel gestaltet sich die Vereinbarung eines Austauschformats aber

Ein weiterer Problemkreis verknüpft sich mit der Kompatibilität der verwendeten Softwaresysteme. Soll das maschinenlesbare Material -- ähnlich einem Buch -- dem Bearbeiter auf seinem eigenen Rechner zur Verfügung stehen, so ist eine Adaptierung an die Softwareumgebung des Arbeitsplatzes erforderlich. Diese Aufgabe gestaltet sich umso schwieriger, als dabei nicht nur Unterschiede zwischen den -- typologisch gleichartigen -- Produkten verschiedener Anbieter auszugleichen sind, sondern u.U. ein Austausch von Material zwischen unterschiedlichen Typen von Softwaresystemen oder -umgebungen (Statistikpakete, Datenbanksysteme für strukturiertes Material, Volltextsysteme bzw. Systeme für prädiertes Material, Hypertext- oder Hypermediasysteme) zu realisieren ist. Diese Systeme verwalten jedoch historische Information in durchaus spezifischen Formen, die in anderen Systemen nicht immer eine unmittelbare Entsprechung haben und deshalb den Austausch zwischen den Systemen erschweren.

Die hier angeschrittenen Probleme erfordern zu ihrer Lösung erhebliche Anstrengungen im Hinblick auf eine Standardisierung sowohl der Behandlung verschiedener Quellengattungen, als auch des Umgangs mit unterschiedlichen Formen historischer Information.³ Gleichzeitig bedarf es der Entwicklung von Werkzeugen, die unterschiedliche Datenformate handhaben und eine maschinelle Umsetzung eines Datenformats in ein anderes leisten können.⁴ Dennoch bliebe selbst eine vollständige Lösung der hier

zweifellos unter inhaltlichem Aspekt. Neben der Bewahrung einer maximalen Vollständigkeit sollte die Quelle nämlich in einer Form vorliegen, in der sie nicht durch konzeptuelle Vorentscheidungen bereits weitgehend geprägt ist. Konzeptuelle Überlegungen spiegeln jeweils ein bestimmtes Erkenntnisinteresse wider, das spätere Bearbeitungen des Materials in unzulässiger Weise in ihren Möglichkeiten einschränken würde. Die Art der Repräsentation der Quelle im Normformat sollte stattdessen die Quelle offen halten für die unterschiedlichsten Bearbeitungen des Materials. Deshalb sind Strukturierungen, die den Informationsgehalt des Quellenbestandes für Abfragen und Analysen verfügbar machen, nur sehr vorsichtig zu setzen, und sollten sich auf wesentliche Aspekte beschränken, die eine erste Annäherung an das Material erlauben. Es bliebe dann immer weiterführender Arbeit überlassen, das im Normformat vorliegende Material im Hinblick auf die eigenen Erkenntnisinteressen und -ziele so zu verändern, daß sich sein Informationsgehalt der Bearbeitung maximal erschließt.

- 3 Während Standards im ersten Sinne etwa in Gestalt von Modellösungen vorstellbar sind, können sie sich im anderen Fall nur auf die Beschreibung des Umgangs mit unterschiedlichen Formen historischer Information erstrecken, nicht aber normativ für den Umgang mit historischer Information selbst sein. Vgl. zur aktuellen Diskussion insbesondere D. Greenstein(Ed.): *Modelling Historical Data*. (Halbgraue Reihe zur historischen Fachinformatik. Serie A: Historische Quellenkunden. Bd.11) St. Katharinen 1991.
- 4 Ein solches Werkzeug wurde im Rahmen des Projekts «Arbeitsplatzrechner für den Historiker» am Göttinger Max-Planck-Institut für Geschichte unter dem Namen StanFEP (Standard Format Exchange Program) entwickelt. Vgl. K. Homann: *StanFEP*. Ein Programm zur freien Konvertierung von Daten. (Halbgraue Reihe zur Historischen Fachinformatik. Serie B: Softwarebeschreibungen. Bd.6) St. Katharinen 1990; M. Gierl/T. Grotum)/T. Werner: *Der Schritt von der Quelle zur historischen Datenbank*.

angedeuteten Probleme in ihrer Tragweite begrenzt, wenn es nicht darüber hinaus auch gelingt, die Produzenten von maschinenlesbarem Quellenmaterial 1. zu einer möglichst vollständigen und quellengetreuen Datenaufnahme sowie 2. zur anschließenden Distribution ihrer Daten zu bewegen. Zwar sind diese Aspekte eng mit den vorhergenannten Problemen verknüpft, insofern die bestehenden Schwierigkeiten eines Datenaustauschs und seine bisherige geringe Verbreitung nicht gerade motivierend auf potentielle Produzenten gewirkt haben, doch tauchen unter diesem Aspekt eine Reihe weiterer hemmender Faktoren auf, die durchaus eigenständigen Charakter besitzen und zusätzlicher Anstrengungen und Maßnahmen zu ihrer Überwindung bedürfen.

Ausgangspunkt der diesbezüglichen Überlegungen ist zunächst die Tatsache, daß die Transferierung von Material auf den Rechner, d.h. die originalgetreue Aufbereitung einer Quelle in Gestalt einer Datenbasis, derzeit nur wenig oder gar keine akademische Anerkennung findet. Ausdruck dieser Situation ist etwa die Tatsache, daß bisher noch nicht einmal anerkannte Konventionen für einen bibliographischen Verweis auf solches Material existieren, durch den die Anerkennung als eigenständige Leistung und geistiges Eigentum zumindest ansatzweise dokumentiert würde. Solange aber solch elementaren Voraussetzungen im wissenschaftlichen Umgang nicht zufriedenstellend gelöst sind, werden Wissenschaftler wenig geneigt sein, die Früchte ihrer Arbeit in Gestalt eines maschinenlesbar gemachten Quellenbestandes anderen zur Verfügung zu stellen, müssen sie doch mit Recht befürchten, daß ihr Anteil an der jeweiligen Untersuchung nur unzureichend dokumentiert würde.

Wissenschaftliche Würdigung findet die Aufbereitung von Quellenmaterial für den Rechner immer erst dann, wenn auf dieser Basis eine wissenschaftliche Untersuchung durchgeführt wurde oder eine gedruckte Edition des Materials erfolgt. Da aber diese zweite Möglichkeit in vielen Fällen nicht angezeigt ist, bestimmt zumeist die intendierte wissenschaftliche Untersuchung die Art und Weise, wie die Quelle maschinenlesbar gemacht wird. Da die Aufbereitung einer Quelle bis in alle Details fast immer sehr arbeitsintensiv und mühselig ist, wird der jeweilige Bearbeiter geneigt sein, seinen Arbeitsaufwand zu optimieren und sich bei der Transferierung des Materials auf das für seine inhaltlichen Fragen und Erkenntnisinteressen Notwendige zu beschränken; alles andere wäre für ihn unter den gegebenen Umständen eine Anstrengung, die sich in keiner Weise auszahlt.

Die durchaus verständliche Kosten-Nutzen-Abwägung in der Transferierung von Quellenmaterial auf den Rechner besitzt dabei keineswegs nur quanti-

StanFEP: Ein Arbeitsbuch. (Halbgraue Reihe zur Historischen Fachinformatik. Serie A: Historische Quellenkunden. Bd.6) St. Katharinen 1990.

tative Aspekte, indem vielleicht Teile des Materials, die für die Untersuchung unbedeutend erscheinen, einfach ausgespart werden, sondern sie hat auch durchaus qualitative Dimensionen. Nicht selten überwiegen in der Aufbereitung des Materials pragmatische Überlegungen, die z.B. dazu führen, daß Kontrollen und Korrekturen nicht bei allen Teilen des Materials mit der gleichen Sorgfalt und Genauigkeit ausgeführt werden. Auch werden schwierig zu verwaltende Teile einer Quelle oft nicht mit maximaler Originaltreue auf dem Rechner abgebildet, sondern bereits bei der Eingabe verändert, um den späteren Bearbeitungsaufwand möglichst gering zu halten. Damit aber verliert das Material für weitere Benutzer erheblich an Wert, sind doch Modifikationen dieser Art am Material selbst nicht mehr erkennbar. Allenfalls eine ausführliche und präzise Dokumentation solcher «Normalisierungen» könnte hier Abhilfe bringen. Doch es ist zu befürchten, daß in aller Regel eine solche Dokumentation aus den genannten Gründen nicht erstellt wird, so daß einem weiteren Benutzer nur der (aufwendige) Weg eines Vergleichs der maschinenlesbaren Daten mit der Originalquelle bliebe.

Auf diese Weise reduziert sich nicht nur jedes potentielle Interesse an einer Sekundäranalyse solcher Daten, sondern es entsteht möglicherweise ein verhängnisvoller Zusammenhang: Geraten quantitativ und vor allem qualitativ beeinträchtigte Datenbasen entgegen der Regel dann doch einmal in den wissenschaftlichen Umlauf, so sind sie kaum dazu geeignet, das akademische Ansehen für die Tätigkeit ihrer Erstellung zu erhöhen, was wiederum das Interesse an einer Veröffentlichung solcher Materialien verringert usw.

Um diesen Kreis zu durchbrechen und der Aufbereitung von historischem Quellenmaterial die Anerkennung zuteil werden zu lassen, die sie eigentlich verdient, ist eine Möglichkeit erforderlich, maschinenlesbares Material -- wie eine gedruckte wissenschaftliche Veröffentlichung -- zu publizieren und als einen eigenständigen wissenschaftlichen Beitrag zu behandeln, der zitawürdig ist und über die Autorenschaft eindeutig Rechenschaft ablegt, gleichzeitig aber auch gängigen wissenschaftlichen Ansprüchen an eine solche Edition genügt. So würde ein Anreiz entstehen, historisches Quellenmaterial auch im Zusammenhang mit «kleineren» wissenschaftlichen Projekten, wie etwa Dissertationen, so auf den Rechner zu übertragen, daß dieses Material neben dem wissenschaftlichen Beitrag, der die Quelle auswertet, als eigenständige Publikation veröffentlicht werden kann. Qualitätsverlusten der beschriebenen Art würde allein schon deshalb vorgebeugt, weil die editorische Komponente des Projekts nicht länger nur ein Abfallprodukt ist, sondern als eigenständiger und ausgewiesener Bestandteil der Arbeit fungieren kann.

Ist erst einmal ein regerer wissenschaftlicher Austausch von maschinenles-

baren Materialien in Gang gesetzt, so ist davon auszugehen, daß die sich damit ausbreitende Erfahrung, «fremdes» Material in sinnvoller Weise für eigene Zwecke verwendet zu haben, zusätzlich positive Impulse gibt. Denn wer einmal von fremdem Material profitieren konnte, wird eher bereit sein, eigenes Material auch anderen zur Verfügung zu stellen und damit maschinenlesbares Quellenmaterial für den Austausch zu produzieren.

Die «Serie C» der «Halbgrauen Reihe für Historische Fachinformatik»

Um auf dem beschriebenen Wege voranzukommen und -- neben den Bemühungen um die Entwicklung notwendiger Standards und einschlägiger Konvertierungswerkzeuge -- auch von dieser Seite her die derzeitigen Bedingungen für einen Datenaustausch zu verbessern, wurde im Rahmen des Projektverbundes «Arbeitsplatzrechner für den Historiker» eine Publikationsmöglichkeit für maschinenlesbar gemachtes historisches Quellenmaterial geschaffen. Unter dem Titel «Datenbasen als Editionen» wurde in der «Halbgrauen Reihe zur Historischen Fachinformatik» die «Serie C» eingerichtet, die in Verbindung mit dem Dansk Data Arkiv in Odense/Dänemark herausgegeben wird und sich von folgenden Prinzipien leiten läßt: 1. Die in dieser Serie veröffentlichten Datenbasen sind zwar z.T. am Max-Planck-Institut für Geschichte in Göttingen bearbeitet, nicht aber erstellt worden; sie wurden von verschiedenen Depositorys, die sie im Verlauf persönlicher Forschungsvorhaben erstellten, zur Veröffentlichung bereitgestellt. 2. Die Serie ist als «offene Reihe» konzipiert und folgt nicht den Vorgaben irgendeines, als verbindlich angesehenen Quellenkanons; sie soll insbesondere auch Material aus «kleineren» wissenschaftlichen Projekten (Dissertationen etc.) aufnehmen, das anderweitig überhaupt nicht veröffentlicht ist. 3. Das Material soll möglichst unkompliziert verfügbar gemacht werden, was nicht nur die Art und Weise seiner Repräsentation beeinflußt, sondern auch bedeutet, daß es zu finanziellen Konditionen bereitgestellt wird, die sich am Selbstkostenpreis orientieren und für alle potentiellen Interessenten damit im Bereich ihrer Möglichkeiten liegen sollten.

Derzeit sind 20 Datenbasen verfügbar, von denen drei strukturiertes Material bzw. statistische Datensätze enthalten,⁵ während die restlichen 17 Datenbasen fortlaufende Texte (Volltexte) in einer Größe von etwa 100 bis 700 Kbyte sind, die als literarische Quellen des Mittelhochdeutschen

5 Es handelt sich im einzelnen um das am Forschungsinstitut für historische Grundwissenschaften in Graz/Österreich entstandene «Urkundenbuch des Herzogthums Steiermark» (ca. 3 MB), eine am Institut für mittelalterliche Realienkunde in Krems/Österreich aufgebaute Datenbasis mittelalterlicher Bildquellen, die formalisierte Beschreibungen von 8000 Bildern in einem Umfang von ca. 3 MB enthält, und einen Datensatz zur historischen Statistik, der auf unpublizierten Aufzeichnungen zum dänischen Sundzoll aus den Jahren 1784-1795 beruht (ca. 16 - 40 MB).

charakterisiert werden können.⁶ Im Unterschied zu den drei erstgenannten Datenbasen wurden die mittelhochdeutschen Texte nicht in ihrer ursprünglichen Form herausgegeben, sondern am Institut einer intensiven Bearbeitung unterworfen: Da die Texte von unterschiedlichen Bearbeitern zu unterschiedlichen Zeiten (z.T. bereits in den 60er Jahren) maschinenlesbar gemacht wurden, waren die verwendeten Kodierkonventionen -- z.B. für diakritische Zeichen, Abbrüviatur, Kursivierung, Emendation etc. -- höchst unterschiedlich. Hier galt es Vereinheitlichungen vorzunehmen, die den Benutzer der Texte davor bewahren, sich für jeden einzelnen Text in unterschiedliche Kodierkonventionen einzuarbeiten. Ausgehend von konkreten Phänomenen und Problemem, die in den verschiedenen Texten im übrigen oft verwandt und ähnlich, nur auf jeweils individuellen Wegen gelöst worden waren, wurden deshalb einheitliche Kodierkonventionen entwickelt, die sich -- nicht zufällig, sondern gewollt -- an das Druck- und Satzsystem TeX anlehnen.⁷ Die praktische Umsetzung der Texte in das einheitliche Format erfolgte anschließend mit Hilfe des erwähnten Konvertierungsprogrammes StanFEP.

Strukturierte Datenbasen werden in einer Form angeboten, die sie für das Softwaresystem clio unmittelbar vorbereitet. Auch die Volltexte sind in ein Textformat überführt, das sie mit clio bearbeitbar macht bzw. sie für die in clio enthaltenen Möglichkeiten der Volltextrecherche erschließt. Das bedeutet allerdings nicht, daß nur der mit den Datenbasen arbeiten kann, der auch clio verwendet. Während der Nutzen der strukturierten Datenbasen ohne clio zugegebenermaßen gering ist bzw. einiger Aufwand erforderlich wäre, um das Datenformat an die Bedürfnisse eines anderen Softwaresystems anzupassen, können die Volltexte schon mit jedem gängigen Texteditor im Rahmen der Möglichkeiten, die dieser Editor bietet, sinnvoll bearbeitet werden. Im übrigen werden alle fortlaufenden Texte nicht nur in einer Standard-ASCII Version, sondern auch in einer Extended-ASCII Version verschickt. Sie erlaubt eine Darstellung der Texte auf dem Bildschirm in erheblich «lesbarer» Form als die Standard-ASCII Version, die für jedes über den Standardsatz hinausgehende Zeichen eine entsprechende Kodierung verwendet.

6 s. Anhang

7 Diese Konventionen erheben in keiner Weise Anspruch auf Vollständigkeit, wenngleich sie so konzipiert wurden, daß neu auftauchende Phänomene flexibel integriert werden können. Sie wurden aus der Praxis entwickelt und dienten dem Ziel, eine Veröffentlichung des Materials möglichst zügig zu ermöglichen. Dagegen sind sie nicht dazu gedacht, die Bemühungen der «Text Encoding Initiative(TEI)» der ALLC/ACH oder verwandte Bestrebungen zur Entwicklung einheitlicher Kodierkonventionen zu unterlaufen; sobald hier konkrete Ergebnisse erreicht sind, werden sie für die «Serie C» berücksichtigt.

Zum Gesamtpaket der Lieferung gehört neben den eigentlichen Daten eine deutsch- oder englischsprachige Dokumentation, die z.B. für die strukturierten Datenbanken eine genaue Beschreibung der Struktur und der Vorgehensweise bei der Auswahl und Bearbeitung des zugrunde liegenden Materials enthält. Für die Volltexte werden insbesondere die vorgefundenen Kodekonventionen und die sie ersetzenden neuen, vereinheitlichten Konventionen dokumentiert; weiterhin werden Angaben über die Herkunft und den Zustand des zugrunde gelegten Textes bzw. der jeweiligen Textausgabe und gegebenenfalls auch über festgestellte Abweichungen des maschinenlesbaren Materials, wie es das Institut vom Depositor erhielt, von der gedruckten Textausgabe gemacht.

Schließlich enthält jede Lieferung eine Selbstverpflichtung des Benutzers über den Umgang mit dem Material und insbesondere die Weitergabe an andere. Da einschlägige bibliographische Konventionen bekanntlich noch fehlen, wird zudem ein Zitat-Vorschlag gemacht. Dieser Bestandteil der Lieferung ist somit charakteristischer Ausdruck der gegenwärtigen Situation, die nicht nur durch die bereits erwähnten bibliographischen Defizite, sondern auch durch Unsicherheiten im Hinblick auf die Rechte an maschinenlesbaren Texten gekennzeichnet ist.

Es sind insbesondere diese rechtlichen Probleme, die die Verhandlungen über die Veröffentlichung von Materialien in der «Serie C» bisweilen recht schwierig gestalten. Dies gilt umso mehr, als auch die Verlage sich zu regen beginnen und bestrebt sind, ihre Rechte an solchen Texten zu bewahren, da die Aufnahme maschinenlesbaren Materials in das Verlagsangebot ein lukratives Geschäft werden könnte. Auf diese Weise wird nicht nur die Veröffentlichung von Material, das bereits verfügbar wäre, blockiert, ohne daß es in absehbarer Zeit im Angebotskatalog eines Verlages auftauchen würde; vielmehr ist auch damit zu rechnen, daß die Preise eines kommerziellen Anbieters erheblich über den an den Selbstkosten orientierten Preisen der «Serie C» liegen werden.

Es bleibt zu hoffen, daß hier sinnvolle Lösungen gefunden werden, die dem wissenschaftlichen Interesse an einem breiten Austausch von maschinenlesbarem Material gerecht werden und ihn nicht an rechtlichen und kommerziellen Hindernissen scheitern lassen. Was die erhoffte Signalwirkung der Reihe anbetrifft, so geben die ersten Erfahrungen im übrigen durchaus Anlaß zu Optimismus, da dem Herausgeber in der Zwischenzeit bereits vier neue Angebote von Wissenschaftlern vorliegen, die an einer Veröffentlichung der von ihnen erstellten Datenbasen interessiert sind. Über die Herausgabe weiterer größerer Textcorpora wird derzeit mit guten Aussichten verhandelt.

Anhang: Halbgraue Reihe für Historische Fachinformatik. Serie C:

Band 1: Die Vorauer Bücher Mosis. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Deutsche Gedichte des XI. und XII. Jahrhunderts, hrsg. von Joseph Diemer, Wien 1849, S. 3 -- 85. St. Katharinen 1990 (Datenumfang: ca. 98 / 97 KB.)

Band 2: Das St. Trudperter Hohe Lied. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Das St. Trudperter Hohe Lied. Textkritische Ausgabe: Text, Wörterverzeichnis und Anmerkungen von Hermann Menhardt, Halle 1934 (Rheinische Beiträge und Hilfsbücher zur germanischen Philologie und Volkskunde, Bd. 22). St. Katharinen 1990 (Datenumfang: ca. 279 / 264 KB.)

Band 3: Notker Labeo, Psalmen nach der Wiener Handschrift. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Notkers Psalmen nach der Wiener Handschrift, hrsg. von Richard Heinzel und Wilhelm Scherer, Straßburg und London 1876. St. Katharinen 1990 (Datenumfang: ca. 606 KB.)

Band 4: Deutsche Predigten des XII. und XIII. Jahrhunderts. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Deutsche Predigten des XII. und XIII. Jahrhunderts, hrsg. von Karl Roth, Quedlinburg und Leipzig 1839 (Bibliothek der gesamten deutschen National-Literatur von der ältesten bis auf die neuere Zeit, Bd. 11, Teil 1). St. Katharinen 1990 (Datenumfang: ca. 86 KB.)

Band 5: Altdeutsche Predigten II: Die Oberaltacher Sammlung. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Altdeutsche Predigten, hrsg. von Anton E. Schönbach, Bd. 2, Graz 1888. St. Katharinen 1990 (Datenumfang: ca. 441 / 439 KB.)

- Band 6: Altdeutsche Predigten III: Die deutschen Predigten des Priesters Konrad. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Altdeutsche Predigten, hrsg. von Anton E. Schönbach, Bd. 3, Graz 1891. St. Katharinen 1990 (Datenumfang: ca. 691 / 686 KB.)
- Band 7: Altdeutsche Predigten und Gebete. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Altdeutsche Predigten und Gebete, hrsg. von Wilhelm Wackernagel, Basel 1876. St. Katharinen 1990 (Datenumfang: ca. 599 / 595 KB.)
- Band 8: Speculum ecclesiae: Eine frühmittelhochdeutsche Predigtsammlung. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Speculum ecclesiae. Eine frühmittelhochdeutsche Predigtsammlung (Cgm. 39), hrsg. von Gert Mellbourn, Lund und Kopenhagen 1944 (Lunder Germanistische Forschungen, Bd. 12). St. Katharinen 1990 (Datenumfang: ca. 351 / 344 KB.)
- Band 9: Eilhart von Oberge, Tristrant. Fragmente des alten Gedichtes und die Bearbeitung. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Eilhart von Oberge, hrsg. von Franz Lichtenstein, Straßburg und London 1877 (Quellen und Forschungen zur Sprach- und Culturgeschichte der germanischen Völker, Bd. 19). St. Katharinen 1990 (Datenumfang: ca. 366 / 337 KB.)
- Band 10: König Rother nach der Heidelberger Handschrift. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: König Rother, hrsg. von Theodor Frings und Joachim Kuhnt, Bonn und Leipzig 1922 (Rheinische Beiträge und Hilfsbücher zur germanischen Philologie und Volkskunde, Bd. 3). St. Katharinen 1990 (Datenumfang: ca. 174 KB.)
- Band 11: Konrad der Pfaffe, Das Rolandslied nach der Heidelberger Handschrift. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Das Rolandslied des Pfaffen Konrad, hrsg. von Carl Wesle, Bonn 1928 (Rheinische Beiträge und Hilfsbücher zur germanischen Philologie und Volkskunde, Bd. 15). St. Katharinen 1990 (Datenumfang: ca. 309 / 308 KB.)

Band 12: Lamprecht der Pfaffe, Das Alexanderlied. Vorauer und Straßburger Version. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Lamprechts Alexander, hrsg. von Karl Kinzel, Halle 1884 (Germanistische Handbibliothek, Bd. 6). St. Katharinen 1990 (Datenumfang: ca. 291 / 272 KB.)

Band 13: Die Weingartner-Stuttgarter Liederhandschrift. Ein maschinenlesbarer Text, erstellt von George Fenwick Jones, Hans-Dieter Mück, Heike Mück, Ulrich Müller und Franz Viktor Spechtler, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Die Weingartner Liederhandschrift [Begleitband zur Faksimile-Ausgabe], Stuttgart 1969: S. 1 -- 310: Transkription, bearbeitet von Otfried Ehrismann. St. Katharinen 1991 (Datenumfang: ca. 483 / 442 KB.)

Band 14: Die Berliner Neidhart-Handschrift c (mgf 779). Ein maschinenlesbarer Text, erstellt von Ingrid Bennewitz-Behr, Diane Donaldson, George Fenwick Jones und Ulrich Müller, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Die Berliner Neidhart-Handschrift c (mgf 779). Transkription der Texte und Melodien von Ingrid Bennewitz-Behr unter Mitwirkung von Ulrich Müller, Göppingen 1981 (Göppinger Arbeiten zur Germanistik, Nr. 356: Neidhart-Materialien, Bd. 1). St. Katharinen 1991 (Datenumfang: ca. 548 KB.)

Band 15: Meister Otte, Eraclius. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Eraclius. Deutsches Gedicht des dreizehnten Jahrhunderts, hrsg. von Harald Graef, Straßburg und London 1883 (Quellen und Forschungen zur Sprach- und Culturgeschichte der germanischen Völker, Bd. 50). St. Katharinen 1990 (Datenumfang: ca. 203 / 192 KB.)

Band 16: Rudolf von Ems, Barlaam und Josaphat. Ein maschinenlesbarer Text, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Barlaam und Josaphat von Rudolf von Ems, hrsg. von Franz Pfeiffer, Leipzig 1843 (Dichtungen des deutschen Mittelalters, Bd. 3). St. Katharinen 1990 (Datenumfang: ca. 616 / 577 KB.)

Band 17: Wolfram von Eschenbach, Lieder, Parzival und Titurel. Maschinenlesbare Texte, erstellt von Roy A. Wisbey, bearbeitet von Astrid Reinecke auf der Grundlage der Edition: Wolfram von Eschenbach, hrsg. von Karl Lachmann, 5. Aufl., Berlin 1891. St. Katharinen 1990 (Datenumfang: ca. 11, 891 und 45 / 11, 826, 42 KB.)

Band 18: Das Urkundenbuch des Herzogtums Steiermark. Maschinenlesbare Texte, erstellt von Ingo H. Kropac, bearbeitet auf der Grundlage der Edition: Urkundenbuch des Herzogthums Steiermark, bearb. von Joseph von Zahn, Bd.e 1 -- 3, Graz 1875, 1879 und 1903. St. Katharinen 1990 (Datenumfang: ca. 1183, 1025, und 613 KB.)

Band 19: The Sound Toll (mit englischsprachiger bzw. dänischsprachiger Dokumentation). Ein Datensatz zur historischen Statistik, erstellt von Hans Chr. Johansen. Mit einer Einführung von Hans Jørgen Marker. St. Katharinen 1991 (Datenumfang: ca. 16 MB.) Die genannten 16 MB beziehen sich auf ein Format, bei dem eine Strukturierung nach Art hierarchischer Datensätze vorgenommen wird; bei einer Umsetzung auf «flat cases» ist mit etwa 40 MB zu rechnen.

Band 20: REAL: Bilder mittelalterlichen Alltags. Herausgegeben und kommentiert v. Gerhard Jaritz. (erscheint voraussichtlich 1991) (Datenumfang: ca. 847, 916, 851 und 371 KB)