

Electronic bookshelf or electronic library? An examination of the medieval and early modern data bank

Touwen, Jeroen

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Touwen, J. (1993). Electronic bookshelf or electronic library? An examination of the medieval and early modern data bank. *Historical Social Research*, 18(2), 213-229. <https://doi.org/10.12759/hsr.18.1993.2.213-229>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

HUMANITIES COMPUTING

Electronic Bookshelf or Electronic Library? An Examination of the Medieval and Early Modern Data Bank

*Jeroen Touwen**

Abstract: This article attempts to illustrate the use of computer databanks in history by examining the Medieval and Early Modern Data Bank (MEMDB). The Medieval and Early Modern Data Bank, of which both an on-line version and a CD-ROM are under construction, will contain an expanding collection of historical monetary, price and wage data, concentrating on the geographical area of the Low Countries, France, England and north-west Germany in the period 800-1800. MEMDB has European offices in Leiden and Brussels. Since 1988 MEMDB offers a PC-prototype which contains 13,256 medieval currency exchange rate quotations. The article observes from a number of different angles the choices and decisions that were made by MEMDB. It relates to questions concerning the theoretical requirements and limitations of historical data banks and examines in which way MEMDB has dealt with these problems.

Introduction

In contrast with the mathematical and natural sciences, using computers in the Humanities does not always imply that the computer as a tool is fully integrated

* Address all communications to Jeroen Touwen, Department of History, Leiden University PO Box 9515, 2300 RA Leiden The Netherlands Tel.: (31)(71) 272228/272742 E-mail: LETTXC@rulmys.LeidenUniv.nl.

This article appeared earlier in *Computers and the Humanities*, 26 (1992): 4, 237-247. Reprinted with permission of Kluwer International Publishers, Netherlands.

into the discipline'. As a historian, I have noticed that many scholars in the historical discipline report on useful applications of computerized methods in their research, but all too often their activities take place at the periphery of more traditional exercises in the field. Notwithstanding the successful protagonists, history on the whole is still a discipline in which the computer is used mostly for wordprocessing only.³ During the last five years however, those who have found computing useful have started organizing themselves and their ideas. A growing number of historians struggles to master various computer skills and to apply them. These historians are not only publishing and exchanging their experiences and the results of their research, but also are reflecting upon the theoretical implications of their efforts. In order to use the computer as a historical tool, it must be incorporated in the discipline rather than allowed to remain a somewhat strange sub-discipline in the eyes of the 'real' historians.' Or, as D.I. Greenstein says, it is important to "approach the mainstream of the profession".⁴

One application of computers in history is very promising: the possibilities for rapid information retrieval now offered by digital technology have brought a new dimension to the task of searching for and analyzing historical material. Most users will admit that a computerized library catalogue is more exact and faster to handle than the old card catalogue. Likewise, historical source data can be stored and retrieved, thus making the computer useful not merely for historians involved in quantitative studies, but for the discipline at large. As J. Th. Lindblad says, "the incessant arranging and rearranging of a mass of scat-

¹ This article was written during my stay at Rutgers University, during which I had the opportunity to study and assist the Medieval and Early Modern Data Bank, after previous experience at the Netherlands Historical Data Archive. I express gratitude to the Netherlands Organization for Scientific Research (NWO), the 'Dr. Hendrik Muller's Vaderlandsch Fonds', the 'Vrijvrouw van Renswoude te's Gravenhage Fonds', and the Netherlands Historical Data Archive for their financial support. I am grateful towards the Medieval and Early Modern Data Bank and particularly professor R.M. Bell for their hospitality.

² For example, in Computers and the Humanities the contribution of historians has been relatively small in recent years.

³ In 1986 in Europe the Association for History and Computing was founded and became active. Several papers at the conferences of the Association for History and Computing have dealt with the subject of a theoretical or methodological framework. Proceedings of these conferences have been published in P. Denley and D. Hopkin, eds., *History and Computing* (Manchester, 1987) (not to confuse with the journal) and P. Denley, S. Fogelvik and C. Harvey, eds., *History and Computing II* (Manchester, 1989). (Of several of the succeeding History and Computing conferences the proceedings have been published as well.) In a broader sense, Denley reflects about the impact of the computer in the Humanities in P. Denley, 'The computer revolution and 'redefining the Humanities'', in: D.S. Miall, ed. *Humanities and the Computer, New Directions* (Oxford, 1990), 13-25.

⁴ D.I. Greenstein, 'A matter of method', in: *History and Computing* Vol 2 (1990) No. 3, 210-215, part. 215.

tered and seemingly unrelated data lies at the core of each historical inquiry. It is a type of task better to be performed by a machine than by man alone".¹ However, many scholars still experience the mandatory technical skills as a threshold that keeps them from computing. A requirement for wide application of the computer as an aid in historiography is the development of increasingly user-friendly software systems. (This requirement might be met due to increasing presence of computers in the lives of generations to come.)

Historians may use the computer to build databases and thus analyze their material. Increasingly, however, they will get the opportunity to use computerized historical sources that have been turned into machine-readable form by others. In many countries historical data archives have been established to give the historian access to existent machine-readable material. Such applications in historiography will have to be justified methodologically before the sceptical majority of historians, to convince them that flashy screens and fluorescent data are both useful and reliable*.

This article attempts to illustrate the use of computer databanks in history by examining one of them: the Medieval and Early Modern Data Bank (MEMDB). By examining this specific data bank we can discuss a number of problems dealing with data banks in history. Since 1988 MEMDB offers a PC-prototype which offers fast and methodologically interesting methods of source-retrieval. We will observe, from a number of different angles, the choices and decisions that were made by MEMDB. Also, we will try to ascertain whether the MEMDB-approach is likely to satisfy large groups of future users. In discussions and reports about historical computing questions have arisen concerning the theoretical requirements and limitations of historical data banks. This article will focus on six of these issues and the ways in which MEMDB has dealt with them, thereby focusing attention on some of the problems facing the historians who use the computer. The issues raised will not be highly technological or methodologically abstract; rather, we shall adopt the point of view of a traditional scholar in the discipline.

¹J.Th. Lindblad, 'Computer applications in Expansion History: A Survey', in: *Itinerario* XII (1988) no.2, 1-62, part. 4. Cf.: 'Some even believe that database technology lies at the heart of a revolution in historical research methods'.(P. Hartland and C. Harvey, 'Information engineering and historical data bases', in: Denley, Fogelvik and Harvey, eds., *History and Computing* II44-62, part. 62.) Whether historical data bases extrapolate traditional methods or revolutionize the profession, should not be the discussion. More important is the question whether we can develop a conceptual framework to deal with and incorporate these new methods.

* Cf. M. Thaller, 'The need for a theory of historical computing', in: Denley, Fogelvik and Harvey, eds., *History and Computing* //, 2-12 and more generally M. Thaller, 'Methods and techniques of historical computations', in: Denley and Hopkin, eds., *History and Computing*, 147-156. Cf. Greenstein, 'A matter of method', p. 213-214.

What is MEMDB?

The Medieval and Early Modern Data Bank was founded in 1982 by professors R.M. Bell and M.C. Howell at Rutgers University in New Jersey and is co-sponsored by the Research Libraries Group (RLG). MEMDB aims at providing scholars with a computer-based reference library holding a steadily expanding mass of information and source-data from the medieval and early modern period. Through the assistance of its European offices in Leiden and Brussels, MEMDB will keep in touch with European scholars and students. The data bank, of which both an on-line version and a CD-ROM are being prepared, will contain an expanding collection of monetary, price and wage data, including weights and measures, mainly concentrating on the geographical area of the Low Countries, France, England and north-west Germany, in the period 800-1800.

MEMDB is an example of a computer project which applies the advantages of the computer to the needs of the historian in a manner which stresses ease of access. Through a retrieval system, the historian can scan and find data, with the help of only a few straightforward commands. Source data that had to be gathered from different books and libraries can now be retrieved and compared very quickly. Since the beginning of the project in 1982, MEMDB has constantly proclaimed its existence and its ideas about historical information retrieval regularly to the scholarly community.⁷ However, until recently, use of the database for the medieval scholar has been restricted, both thematically and technically. Since 1988 MEMDB is available as a prototype on diskettes containing a large master data set with 13.256 medieval currency exchange rates.⁸ For many medievalists this is a large and valuable collection, but naturally only a few of them could really apply these data to their research. Many historians and librarians, however, have acknowledged the elegance of the user interface.⁹

⁷ A selection of recent publications: R.M. Bell, 'The Medieval and Early Modern Data Bank in Europe and North America', in: R. Metz, E. van Cauwenberghe and R. van der Voort eds., *Historical Information Systems Studies in Social and Economic History* (Leuven, 1990); R.M. Bell, and E. van Cauwenberghe, 'The Medieval and Early Modern Data Bank (MEMDB)', in: *Tijdschrift voor Geschiedenis* 103 (1990) no. 2, 260-278; M. Howell, 'The Medieval and Early Modern Data Bank', in: R. Metz, E. van Cauwenberghe and R. van der Voort, eds. *Historical Information Systems, proceedings of the 10th International Economic History Congress* (Leuven 1990); 'The Medieval and Early Modern Data Bank', in: *The Research Libraries Group News*, 12 (1987). A detailed description of the use of the PC Prototype can be found in: M. Carlin, *The Medieval and Early Modern Data Bank Handbook* (Stanford 1988). For more information: Prof. R.M. Bell, Dept. of History, Rutgers University, New Brunswick, NJ 08903.

⁸ The 13.256 currency exchange quotations from 1150-1500, covering Europe, Byzantium, the Levant and North-Africa are based on Peter Spufford's *Handbook for Medieval Exchange* (London, 1986). The system gives source information and explanatory background text for each data-item. The prototype is rather slow, but mainly the design of its user-interface is of interest here. The CD-ROM and on-line databank will have a slightly modified update of this use-interface.

Since the PC-prototype of MEMDB was presented, new data sets have been acquired both from the home base in New Jersey and from European contributors.⁹ In the fall of 1992 MEMDB will be available on CD-ROM with an increased collection of historical source-data. In the meantime, RLG is developing the on-line version of MEMDB, which will be accessible to a wider public through RLIN and Internet.¹¹ The first CD-ROM will soon be available and will contain the original currency exchange quotations, expanded with new price and currency exchange data. These consist of about 40.000 items of Posthumus' price data from the Netherlands, 17.000 items of weekly, monthly and yearly prices of rye, wheat, barley and oats in Cologne from 1531 to 1797, and approximately 50.000 items of currency exchange rates and currency equivalents in silver- and gold-weights from the period 1400-1800.¹²

Back to facts....

One might wonder which place this source retrieval system is meant to occupy in future medieval studies, and in which way it had tried to anticipate questions and queries of its future users. We shall here examine MEMDB by discussing the following six issues. First, does the data bank strive to offer all sources that medievalists might want to consult on specific themes? Does it claim to be comprehensive on a specific subject? Second, will it be economically feasible for MEMDB to provide on-line an enormous amount of rare, specialized data, that is seldom used? Third, to what extent can the source data in the system be authentic and, therefore, reliable for an on-line user? Fourth, is random browsing possible and useful? In other words: how transparent is the data base system? Fifth, why does MEMDB present data items instead of data files to the user? If data sets are catalogued by themes and subjects, they may serve as the 'entity' that serves to access the data collection. Finally, does the data bank

⁹ Many queries that were sent to MEMDB during the last two years were concerned with the adaption of the shell of the data bank by inserting other data, often textual data. The actual items or currency quotations could for example be filled with bibliographical references and the main contents of the data base could be the blocks of text which are now the explanatory text.

¹⁰ Source data, which are vital for a good thematic coverage of the distributed Master Data Set, such as the Posthumus data on Dutch prices, have been entered into the computer from publications. Other data sets are deposited in machine-readable form, like the ones donated by Rainer Metz from Cologne. These data sets have come into being after many years of searching through archives and publications and were published before donation to MEMDB.

¹¹ RLIN is the Research Libraries Information Network. Internet is a network of computer networks that links research institutions throughout the world.

¹² The data originate from N.W. Posthumus, *Inquiry into the history of prices in Holland* (Leiden, 1964), P. Spufford, *Handbook of Medieval Exchange* (London 1986) (plus additional exchange rates by Spufford) and from collected material by Rainer Metz, Cologne.

acknowledge other data archiving institutions and supply references to these? By examining these questions in succession, we shall gain insight in the approach and ideas of MEMDB. At the same time our discussion will shed some light on the specific problems historians may encounter while using data banks.

Selecting a 'critical mass'

In general historians tend to use rare and eccentric archival sources, preferably unique for their particular study (and hopefully as unique as their particular study). Medievalists are no exception to this rule. The vast majority of sources is infrequently consulted. Will it be economical for MEMDB or any other historical data bank to provide an enormous amount of data on-line that is hardly ever used?¹¹

Probably no data bank can provide a collection of data that will entirely satisfy the needs of one user, however specific the field may be. Of course a data bank will always aim at a sufficient proportion of the users' needs, but this is especially true of historical data banks. No one expects a historian to write a book merely on the basis of data accessed through a terminal. Certain data, like the currency exchange rates in MEMDB's prototype, are useful for many different studies. Also, certain data that used to be scattered over a multitude of archives and publications, can be specifically suitable for computerized access. MEMDB calls a sufficient or satisfactory collection of these data the 'critical mass'.

If scholars, in addition to the unique documents that they study, find tabular material in MEMDB that is useful or even vital as background information for their study, one can say that the aim of the data bank has been achieved. Although in this case the data bank material may account for only a small percentage of the used sources, this is still material that otherwise could never have been found so quickly and would therefore probably never be applied to the study. From the point of view of the data bank, it is therefore desirable to establish a critical mass of data in the on-line data bank (or on CD-ROM), and store the rest elsewhere.

MEMDB thus developed a plan calling for three levels of data storage. The first level is a master data set on-line that contains the critical mass of data. All data items are stored as one large data set, regardless of origin. However, each data item is linked to a quantity of source information stating the original source and the depositor. The master data set will also be available on CD-ROM. Next, the off-line archive will hold data sets that are expected to be of lesser importance. References to these data sets will be added to the on-line data. The data sets will remain in the format in which the original research initiator deposited them. Users can order such data sets (barring copyright restrictions)

¹¹ Cf. M. Thaller, 'Databases and expert systems as complementary tools for historical research', in: *Tijdschrift voor Geschiedenis* 103 (1990) no. 2, 237; see below.

by completing an electronic order slip. On a third level, MEMDB attempts to provide an electronic reference list of the holdings of other data archives.¹⁴ This goal may appear optimistic, but as telecommunications and networks develop, and as institutions standardize their lists and formats, the efficient exchange of holdings should become a reality. The electronic catalogue of the holdings of other data archives is expected to be added to RLG's bibliographic retrieval system, known as RLIN.

Of course, the decision concerning which data belong to the critical mass requires extensive scholarly knowledge on the part of MEMDB's editorial board. The data bank developed a set of criteria concerning the reliability, usefulness and functionality of the data in order to determine which data sets should be implemented in the master data set. This is a pragmatic solution. In theory every data set should be part of the master data set, just as in theory a library should keep all its shelves open to the public. Thaller uses this issue to state what in his eyes is a shortcoming of the MEMDB concept. This policy, he argues, is based on economic considerations: to keep an enormous amount of source data available on an on-line system is not economically justified.¹⁵ MEMDB sees other reasons for restricting the master data set to a selection of data: To include all data in the master data set, apart from the cost considerations involved, would result in overwhelming the user with data that in most cases would turn out to be an irrelevant distraction.¹⁶

One other reason for discerning a master data set and an off-line archive is the editing that is done to the former. Since the retrieval system offers indexes, source-references and explanatory text for each data item, preparing data for inclusion takes time. This is also the reason why one cannot simply publish all data sets on a set of CD-ROM's. As in all data banks, the retrieval system is essential for the functionality of the information.

A fool-proof retrieval system?

Our next question can be stated as follows: does MEMDB respond to a query by providing the user with the right material from the right historical context? Can one find what one is looking for? More specifically, the issue here is whether there is either some sort of standardisation throughout the whole system, or an 'intelligent' query-program in the retrieval system that substitutes the historically appropriate equivalents to geographical names, different calendars, different spelling, etc?¹⁷

¹⁴ R.M. Bell, 'The Medieval and Early Modern Data Bank in Europe and North America'. Conference paper, International Conference on History and Computing, Amsterdam, June 1989, 10-19; Bell and van Cauwenberghe, 'The Medieval and Early Modern Data Bank' in: *Tijdschrift voor Geschiedenis*, 103 (1990) no.2, 264-266.

¹⁵ M. Thaller, 'Databases and expert systems as complementary tools for historical research', in: *Tijdschrift voor Geschiedenis* 103 (1990) no. 2, 237.

¹⁶ Bell and van Cauwenberghe, 'The Medieval and Early Modern Data Bank (MEMDB)', in: *Tijdschrift voor Geschiedenis* 103 (1990) no.2, 265.

The enormous diversity in nomenclature in archival material is a recurring problem for the builders of historical data bases. The confusion increases even more by the diversity of added computerized names and labels. There is, therefore, a great need for a standardized nomenclature, however difficult it might be to achieve. MEMDB aims at editing and standardizing the master data set, in some instances by imposing standard labels, preferably by the development of a thesaurus. In the off-line archive both the original master data set files and the other off-line data sets will be stored unedited. Donators of data sets will be asked to provide information about their activities in standardizing the data. In the on-line system this information is stored in the background information record, which is retrieved by pressing a single function key when examining a data-item.

The retrieval of data through geographical names cannot be reliable without the use of standardized names or a very adequate glossary that is consulted before a query is made, or that is able automatically to 'translate' queries. A thesaurus which would recognize all the names used to refer to one place over a period of a thousand years would be mostly difficult to compose. The MEMDB prototype does not contain such cross reference lists on the names of the places, currencies, etc, but supplies header items that contain cross-reference information or reference information for headings not actually occurring in items ('See' or 'See also' references).¹⁷ In due time, and with adequate funding, MEMDB hopes to construct a glossary. This is vital, since the larger the master data set, the more essential it will be to access items which cannot be found in another way.

It stands to reason that the use of thesauri can be criticized. The scholar will not want to rely on ready-made associations of concepts, since they imply too much interpretation. Thaller conceptualizes a solution using artificial intelligence software to evaluate and restate a user's query, in order to get comprehensive hitlists.¹⁸ Preferring to be on the trailing edge of technology, MEMDB attempts merely to supply data. But in addition it provides sufficient documentation to accompany these data, thus allowing the scholars to determine what else he can find or which indexes they should use to obtain additional relevant material.

¹⁷ Thaller deals with this question on a theoretical level, which leads him to what he calls 'Artificial Intelligence in historical data banks' in his article 'The need for a theory of historical computing', in: Denley, Fogelvik and Harvey eds., *History and Computing II*, 6. See below.

¹⁸ Cross reference lists greatly depend on scholarly knowledge and are often highly subjective. The cross-reference list available to the Spufford data was judged to be too idiosyncratic to be useful or reliable for an outsider if not extensively modified.

¹⁹ M. Thaller, 'The need for a theory of historical computing', in: Denley, Fogelvik and Harvey, eds., *History and Computing II*, 2-12. See also M. Thaller, 'Methods and techniques of historical computations', in: Denley and Hopkin, eds., *History and Computing*, 147-156, part. 149-151.

Is retrieval reliable?

Our third issue deals with the extent to which the source data are incorporated in their authentic form or appearance. Does the user examine original unmodified source data, or has either the archiving institution or the depositor of the data set modified, changed, interpreted and selected the data? In other words, are the data close to the sources and can the user find out about this relationship? As Metz put it in a description of the software package TUSTEP, "as a prerequisite for source-oriented processing it is necessary to make the sources completely machine-readable without changing or modifying the original information in advance by coding or selecting."²⁰ However, modification will always occur, and to different degrees. For example, in the Archivo General de Indias in Sevilla (Spain) a large part of the archival sources are scanned and made accessible in a large 'picture data bank' on optical disc.²¹ Thus the documents can be retrieved virtually in their original form. One can search on key-words that are linked to each image, but no computer analysis or calculation can be conducted on the contents of the source-material. (Besides, this admirable project requires an immense storage capacity!)

As an almost diametrically opposed kind of institution (more tool than archive), MEMDB aims to present a rather eclectic mass of machine-readable data, instead of a large part of an existent archive. These data are largely entered into the computer by others, and thus a considerable degree of 're-shaping' of the data-structures is inevitable. However, MEMDB argues, if the origin of the data is well documented, there is no harm. Thus the first requirement that the developers of MEMDB had in mind was that the system had to "retain all the documentation conveyed with the data set, not just the tabular results, and had to make this documentation readily available."²²

Focusing on tabular material in the retrievable units, MEMDB has chosen a data item characterized by 8 fixed and 2 optional fields. The fields, however, are repeatable, thus allowing large data items which are all accessed via indexes of the same sort.²³ Blocks of explanatory text are linked to each item, starting with commentary relevant to the specific quotation but - as in the case of the Spufford currency exchange data - sometimes even including a full

²⁰ Metz, R., 'TUSTEP: a software package for source oriented data processing in history', in: Denley and Hopkin, eds., *History and Computing*, 241-250, part. 241.

²¹ The computerisation project of the Archivo General de Indias recently received the European 'Erasmus prize' as an innovative project. More on this project in: P. Gonzalez, 'The computerisation project of the Archivo General de Indias', in: *Data, Computers and the Past, Proceedings of the conference 'Archiving and Disseminating Historical Machine-readable Data'*, Leiden, april 27-28 1990, Cahier 5 of the Low Countries Association of History and Computing (Vereniging voor Geschiedenis en Informatica), (Hilversum 1992), 52-67.

²² Bell and van Cauwenberghe, 'The Medieval and Early Modern Data Bank (MEMDB)', in: *Tijdschrift voor Geschiedenis* 103 (1990) no. 2, 262.

²³ Referring to the MEMDB on-line system currently being developed.

introduction. In addition, records containing source information and heading records are linked to the data items. In this way all the essential information from deposited data sets can be supplied. The background text is, of course, a subjective addition. MEMDB does not let the user download this background text. The data items themselves may be downloaded by the user for more elaborate analysis than is feasible during the on-line session.²⁴

Of course in many cases selections are made by the data set initiator at the data entry stage, because of thematic or chronological restrictions, or as a result of unusable or perished sources. Since MEMDB concentrates on tabular material, the coding of the items should not result in irretrievable loss of information.²⁵ In theory, all relevant information considering record modification or decisions on standardisation can be stored in the source-documentation records.²⁶

The future MEMDB master data set will provide another but often neglected advantage of on-line data banks. Since the mass of data forms a dynamic collection, which is constantly being updated, the retrieved information is sure to be reliable. Should a user find inconsistencies in the data, a footnote can be added to the relevant master set items (in this matter the European branches, close to the original material, will undoubtedly prove valuable). Likewise the inclusion of any new material of the same sort can be indicated in footnote records attached to certain data items. Such methods ensure a more reliable and more frequently updated collection of sources, surpassing any printed publication. Updating is one of the main advantages of on-line data banks. Now that CD-ROMs are now so inexpensive, it would even be possible to update CD-ROMs, and supply new discs with both the original and supplementary materials.

²⁴ MEMDB-Online supplies the user with a PC-program that enables reviewing previous searches and composing new search-commands before establishing a connection. Once the program has checked the syntax of the queries and stored them, the dialup connection can be made and the searches executed. After the result sets are downloaded, the user can log off and examine the result sets at leisure. This cycle can be repeated, thus maximizing functionality and minimizing connection time and cost.

²⁵ The importance of careful documentation of these irregularities in a data set, either by the research initiator or the archiving institution, was earlier pointed out in: L.J. Touwen, 'Experiences in Data Archiving. Developing the History of European Expansion Databank', in: P.K. Doom, R. van Horik and L.J. Touwen, eds. *Nederlands Historisch Data Archief I: Eindverslag van een pilot-project* (Amsterdam 1990) 15-23.

²⁶ MEMDB considers the on-line data as published material analogous to published material in a library: the data can be used and reproduced provided source-references to their origin are stated. Thus a data set donated to MEMDB becomes equivalent to a scholarly publication.

Systematic searching versus random browsing

The fourth question deals with the possibility of 'wandering aimlessly' through large collections of material, perhaps gaining new insights, versus more specific searching. MEMDB aims not to 'overwhelm' the user with data that might otherwise act as a distraction. But do scholars who want to use electronic data in their research have to know beforehand what they are looking for? Through examples we will simulate the experiences of a first-time user who wishes to browse through the system.

MEMDB invites the user on the opening screen to 'scan' a list of 'indexes'. Thus, the command 'SCAN LONDON' will result in a list of all indexes that contain the string 'London', indicating the number of data items. Likewise, SCAN PLACE LONDON will result in an index list in which index PLACE is 'London'. Pressing one function key leads directly to the data items themselves. From the data items, pressing one key leads to the explanatory text pertaining to these data items, while another key leads to their sources.

One can argue that browsing through a mass of data offers clues, which lead to further research. In many data banks, however, browsing could lead to confusion, caused by an excess of information. We shall therefore consider how a user with a particular query would browse through the MEMDB retrieval system. Let us assume that a student knows that in 1400 A.D. a shipment of cloth from Bruges to Venice was priced at three thousand Flemish pounds.²⁷ When he wants to discover the actual size of the shipment, he can find necessary information in MEMDB. MEMDB provides a set of specific terms that can be scanned and examined: the student can 'scan' the names of Bruges and Venice, or all places that start with 'Br' (to make sure he finds all spelling variations), and can also select all currency exchange data of Flemish pounds between for example 1395 and 1405, in order to convert Flemish pounds to other currencies of the time (see figure 1 and 2). This can all be done in a matter of minutes. When he has found the desired data, background information explains any intricacies relating to the Flemish pound around the year 1400.²⁸

Let us suppose that another student is describing 'regents' in 15th Century Leiden and wants relevant material about the period. She is, however, unsure of which direction to take, since her research is less advanced than that of the student of Flemish pounds. In this case a directory of available data sets, or a thesaurus containing keywords describing data sets, could be more appropriate than direct access to individual data items. In theory, however, this user can create a data set from the MEMDB master data set by scanning all data from the

²⁷ Cf. Bell and van Cauwenberghe, 'The Medieval and Early Modern Data Bank (MEMDB)', in: *Tijdschrift voor Geschiedenis* 103 (1990) no.2, 260.

²⁸ He will find out that the moneys of account in the principalities of the Low Countries were particularly complex, since 'men of all nations came to Bruges and the men of Bruges stayed at home' so that many currencies could be found. He will then read which currencies were used in Bruges at that time, and their silverweight equivalent.

| Search: FIN S3 DAT 1397-1405 | | | | Result: 9 Items |
|--|---------------|-------------------|---------------------|-----------------|
| Action: F3 DIS 8 LONG | | | | |
| # | DATE PLACE | A | B | CAT |
| 1. | 1396» Malines | Flemish ponden» 1 | Livres, s» 256li. | Accounting»> |
| 2. | 1396» Malines | Flemish ponden» 1 | Schelling» 106s.8d. | Accounting»> |
| 3. | 1398 Lubeck | Flemish ponden» 1 | Mark, sch» 6mk. | Commercial»> |
| 4. | 1398» Malines | Flemish ponden» 1 | Livres, s» 272li. | Accounting»> |
| 5. | 1398» Malines | Flemish ponden» 1 | Schelling» 113s.4d. | Accounting»> |
| 6. | 1399» Malines | Flemish ponden» 1 | Schelling» 120s. | Accounting»> |
| 7. | 1399» Malines | Flemish ponden» 1 | Livres, s» 2881i. | Accounting»> |
| 8. | 1399» Malines | Flemish ponden» 1 | Schelling» 20s. | Accounting»> |
| 9. | 1403» Lubeck | Flemish ponden» 1 | Mark, sch» 5mk.6s. | Commercial»> |
| 1369/12/25-1398/04/07 Malines | | | | Item 1 |
| Flemish ponden groot 1 | | | | |
| Livres, sous, and deniers payment of Brabant 256 li. | | | | |
| Accounting | | | | |
| [] 256 livres | | | | |
| <Text exists. Press F4 to display.> | | | | <end of item> |
| Esc: finished with this item. | | | | |
| NEXT ACTION: | | | | |
| F1 Actions F2 Indexes F3 Long F4 Text F5 Sources | | | | |

Figure 1 Screen of the MEMDB-prototype which shows data items on Flemish Pounds between 1397-1405. Note that item 1 is displayed in long form.

| Search: FIND S3 DATE 1397-1405 | | | | Result: 9 Items |
|--|---------------|-------------------|---------------------|--------------------|
| Action: F5 DIS 1 SOURCES | | | | |
| # | DATE PLACE | A | B | CAT |
| 1. | 1396» Malines | Flemish ponden» 1 | Livres, s» 2561i. | Accounting»> |
| 2. | 1396» Malines | Flemish ponden» 1 | Schelling» 106s.8d. | Accounting»> |
| 3. | 1398 Lubeck | Flemish ponden» 1 | Mark, sch» 6mk. | Commercial»> |
| 4. | 1398» Malines | Flemish ponden» 1 | Livres, s» 2721i. | Accounting»> |
| 5. | 1398» Malines | Flemish ponden» 1 | Schelling» 113s.4d. | Accounting»> |
| 6. | 1399» Malines | Flemish ponden» 1 | Schelling» 120s. | Accounting»> |
| 7. | 1399» Malines | Flemish ponden» 1 | Livres, s» 2881i. | Accounting»> |
| 8. | 1399» Malines | Flemish ponden» 1 | Schelling» 20s. | Accounting»> |
| 9. | 1403» Lubeck | Flemish ponden» 1 | Mark, sch» 5mk.6s. | Commercial»> |
| Munro, J. H. A. Material from the Stadsarchief van Mechelen, Stadsrekeningen 1368-1435, privately communicated by Professor John Munro of the University of Toronto. | | | | Sources for Item 1 |
| CITED IN: | | | | |
| Spufford, P., with Wilkinson, W. and Tolley, S. | | | | CONTINUES |
| Esc: finished with sources. | | | | |
| NEXT ACTION: | | | | |
| F1 Actions F2 Indexes F3 Long F4 Text F5 Sources | | | | |

Figure 2 Screen of the MEMDB-prototype showing sources of data item 1.

period and the places she studies. In an on-line session, she would download all relevant information for further analysis, thus saving connection charges. With the CD-ROM, she can, at her own convenience, narrow or expand her searches until she finds a useful set of prices, exchange rates, or other relevant economic data. One can easily compare this with searching through an electronic library catalogue, but scanning the header-lists, in which the indexes are browsed, offers a better overview. In some library catalogues this is also possible.

In most instances, the header-lists of MEMDB will indicate the presence of potentially suitable material. Once again, the user will eventually have to reckon with the off-line holdings, and this can be done via the electronic references to off-line data sets.

A third example may illustrate searching the data bank with a more defined goal.²⁹ A scholar has done some research and found that Christopher Columbus' voyage to America cost 1.5 million maravedis. He wants to know what the sum equates to in today's currency. MEMDB will rapidly provide the answer. The scholar can browse the index to find items on maravedis (see figure 3). He then looks up the items on the currency in the late fifteenth century, and examines the full display, containing the exchange rate with the florin of Aragón. With single keystrokes he learns about the sources and background of the florin of Aragón. Using a new display of all exchange rates around 1492, he also obtains the exchange rate with the Flemish groten. In this currency he can look up the cost of basic commodities like eggs, butter, bread and sugar around 1500, and calculate the cost of these in maravedis. He can also discover the daily wage of a carpenter or mason which offers another way of interpreting an amount of money.

Through the exchange rates and wage and price data, the user can convert the 1.5 million maravedis to a certain amount of commodities or days of work. By having access to contemporary prices, the user can approximate the real cost of Columbus' voyage. With additional information from MEMDB, comparisons with the cost of crusades or military campaigns may also be made, although it should be kept in mind that such conversions build on certain assumptions (converting to contemporary prices using wages gives quite a different picture than using the price of daily bread). Nevertheless, by having a wide range of original source quotations at one's disposal, both random browsing and systematic searching are quite possible. At the moment, the contents of MEMDB are too restricted for such exercises. After the addition of series of price data, scholars will find that price comparisons such as the above are performed much more quickly and easily than is possible without the computer.

²⁹ Bell, *The Medieval and Early Modern Data Bank ...*, in: Metz, van Cauwenberghe and van der Voort eds., *Historical Information Systems*, 74-76.

| SCA MARA | | 1374 Headings | |
|------------------|-------|---|-------|
| Action: SCA MARA | | Headings 637-645 | |
| # | ITEMS | HEADING | INDEX |
| 637. | 176 | Maravedis | Term |
| 638. | 76 | Maravedis and dineros | Term |
| 639. | 12 | Maravedis and dineros of Castile | Term |
| 640. | 1 | Maravedis de moneda nueva | Term |
| 641. | 1 | Maravedis de moneda vieja | Term |
| 642. | 1 | Maravedis de moneda vieja (i.e., pre-inflation of late» | Term |
| 643. | 1 | Maravedis de novenes | Term |
| 644. | 1 | Marburg | Place |
| 645. | 0 | Marburg -- see Schillinge and pfennige of Marburg; Sch» | Term |

NEXT ACTION:
F1 Actions F2 Indexes F3 Long F6 Items

Figure 3 Screen of the MEMDB-prototype when scanning for Maravedis.

Searching data or searching data sets

Many related institutions establish an off-line electronic data archive instead of presenting source-data on-line.³⁰ When an archiving institution offers an on-line data bank, it has to decide whether it is going to offer the holdings as a catalogued collection of data sets or as a (directly accessible) mass of data material. Social science data archives usually archive their holdings in the first manner (with or without on-line access). It is clear that the MEMDB master data set follows the second approach. Many CD-ROMs also present a thematic collection of source data accessible via a retrieval system.³¹

CHRONOS, developed by the Netherlands Historical Data Archive (NHDA), is an example of an on-line historical data archive system.³² The CHRONOS

³⁰ For example the Danish Data Archive (DDA) in Odense, Denmark, and the ESRC Data Archive, in Colchester, England, which both have special branches that archive historical data sets.

³¹ For example 'U.S.History on CD-ROM' (Bureau Development Inc.), CD-ROM with historical 'Dutch Printers Devices', classified in Iconclass (Royal Library, The Hague, Netherlands), the Patrologica Latina Database on multiple CD-ROMs, the 'Perseus' CD-ROM with sources and studies on Ancient Greece, the Spanish 'Admyte' CD-ROM (MicroNet), etc.

³² The Netherlands Historical Data Archive has recently been officially recognized as the national center for archiving, cataloging and disseminating historical datasets in the Netherlands. A prototype of its archiving system 'Chronos', which is built in the

prototype is a data set-oriented catalogue, that also allows for examination of a section of the holdings stored on-line. The user looks for data sets in the catalogue, searching via keywords, regions, periods or names. She is given information about the size and sources of the data sets found and about their availability on-line. The user is then offered the opportunity to examine, analyze or even download the data from a list of data sets (provided there are no copyright restrictions stated by the donor and the data set is kept in the on-line archive).

The attraction of this type of system is that the user is virtually led by the hand through the system: menus are self-explanatory, and suggestions are made about possible actions, such as the execution of simple statistic procedures to evaluate a range of data. Menus can skilfully be skipped by the more experienced user, but are of great help to anyone who is less familiar with computerized searching. CHRONOS works mostly like a library search catalogue, but goes further in that data sets can be modified and analyzed, albeit through limited, preprogrammed procedures. MEMDB creates a slightly higher threshold for a user who wants to calculate an average or a frequency chart: this requires downloading the set to the PC and applying some multipurpose software program to the data.

As mentioned earlier, with MEMDB the user is not aware that the system is made up of different data sets. All data are merged in the master data set, and searches can only be made on the detailed level of data items. Using rather simple commands (there are only eight possible commands in the MEMDB system) the user sends his queries to the data bank and whatever tabular output he wishes. Although this requires the user to devise smart queries, the user is not bothered by levels of catalogue or data collections, by different data files or by different formats of data - and is not confused by long marches through menus.

Since the MEMDB master data set will have a carefully selected collection of information in a well-defined field, it will serve as a useful tool for scholars interested in specific periods and regions. In this sense MEMDB resembles a reference handbook. One can look up answers to specific questions encountered during analysis or research. By way of contrast, one could say that CHRONOS resembles a bookshelf, where we read the titles on the back of the books, before opening the books themselves. Scholars will look at this electronic bookcase to check if there is any data set of interest. This will satisfy many users, but the users that have specific questions ready cannot query all data sets at once. Whereas the MEMDB seeks to provide on-line access to its critical mass of data (as a reference tool), the NHDA as a data archive aspires to

SAS-software package, is accessible through the Dutch computer network Surfnet and in the near future internationally through Internet. For more information see P.K. Doom, ed., *Netherlands Historical Data Archive II: Chronos Historical Data Archive System, Handbook of a prototype* (Amsterdam 1990).

catalogue existent data sets that have been compiled by researchers and are available for secondary analysis.³³

Navigating through the information jungle

Our final issue is whether the data bank supplies complete reference to holdings of other archiving institutions. We mentioned earlier that MEMDB plans to do this, but the subject deserves some extra attention. Most historical source material does not confine itself to national borders and becomes increasingly interesting for the researcher when amassed from different regions. Thus a historical data bank should strive to offer a comprehensive collection of source data. However, it is often difficult for the user to determine just how many data banks need to be consulted (there is a cost-factor involved). It is, therefore, increasingly important to cross-reference from one database to another. Network users are already familiar with listservers that distribute holdings and catalogues of different institutions. One example here is the inventory of electronic texts of the Georgetown Center for Text and Technology, which is available on-line and provides an updated list of projects involving electronic texts from around the world.³⁴

MEMDB's plan to provide an electronic reference list to other historical data banks is an important supplement to the data bank's own holdings, since all researchers want to survey all possible material in their field. In the current situation, a historian will be fortunate to have on-line access to one historical data bank in his own country. A well-equipped university library might offer links to a few other databank systems, either on CD-ROM or through dial-up mode. However, this will change, and we shall certainly see extensive, user-friendly networks that can exchange and coordinate the information available at various institutions. In this respect historians (or humanists for that matter) can comfortably adopt a wait-and-see attitude, since developments in history follow pioneering initiatives in the natural sciences and in business.³⁵ While the people

³³ In this regard MEMDB is more like the text-corpora data bases (as there are, for example, the Thesaurus Linguae Graecae (TLG), the American and French Research on the Treasury of the French Language (ARTFL), the Dartmouth Dante Database, and the Fontes Anglo-Saxonici, etc).

³⁴ Neuman, M. et al. *The Georgetown University catalogue of projects in Electronic text*, Georgetown Center for Text and Technology, Georgetown University (Washington, 1991)

³⁵ To get an impression of the vast expansion of both the number and the size of information services, one can look at the DIALOG Data Base catalog, listing thousands of online accessible data bases. The number of on-line accessible data banks in the world ranges in the thousands. Large commercial data banks provide all kinds of information, from listings of companies to the Dow Jones index. A new category of institutions has appeared that does not provide data bases but offers to execute one's searches as an intermediary, a 'guide through the jungle' (for example 'Savage Information Services' and 'Information On Demand').