The Daily Life of the Middle Ages, Editions of Sources and Data Processing

MANFRED THALLER

Nobody would doubt that research on the daily life of medieval societies is possible only, if medieval source material is easily accessible in published form; but, are there any requirements of this type of research, which ask for specific properties of such editions? Similarly: data processing techniques – be it for the production of cheaper books, the statistical analysis of material, which is otherwise hard to come to grips with, or the storage of vast arrays of material, which shall be made more easily accessible – have in recent years been employed in a number of historical disciplines; but, is there anything in the daily life of medieval times, which asks for specific procedures in applying such tools?

This author has already in a short communication[1], presented at an earlier round table of the Institut für Mittelalterliche Realienkunde, tried to point out that due to the specific research situation of studies dealing with the material (and not so exclusively material) aspects of everyday life, certain forms of data processing are, indeed, particularly suited to this kind of research. In this paper we will try to show: (a) that the two questions at the beginning have to be answered affirmatively, (b) that from these two answers clearly defined properties of specific research tools can be derived and (c) how they can be integrated into a coherent technical approach.

1. The Daily Life of the Middle Ages as a Special Topic of Research

If one looks at the rich collection of studies, contained in the conference volumes published over the years by the institute in Krems, one could at first glance scarcely notice anything common in the use the various contributors made of the sources accessible to them. We would like to point out that there exist three types of study, which seem to us to be particularly interesting, each of them, of course, being existent in other fields of research as well.

(a) There are contributions[2], which focus completely upon a very restricted number of sources, being dedicated to a thorough analysis of, e.g., a list of inventories, individual accounts, a small collection of pictorial sources and the like. This is in many ways the

[1] *Manfred Thaller*: Mittelalterliche Realienkunde und EDV, In: Die Erforschung von Alltag und Sachkultur des Mittelalters. Methode – Ziel – Verwirklichung, Wien, 1984 (= Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 6 = Österreichische Akademie der Wissenschaften, phil. Hist. Kl., Sitzungsberichte 433) 210-228.

[2] E.g.: *Josef Riedmann*: Adelige Sachkultur Tirols in der Zeit von 1290 bis 1330, In: Adelige Sachkultur des Spätmittelalters, Wien, 1982 (= Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 5 = Österreichische Akademie der Wissenschaften, phil. Hist. Kl., Sitzungsberichte 400) 105-131. *Rudolf Endres*: Adelige Lebensformen in Franken im Spätmittelalter, ibid. 73-104; *Gerd Zimmermann*: Ein Bamberger Klosterinventar von 1483/86 als Quelle zur Sachkultur des Spätmittelalters, In: Klösterliche Sachkultur des Spätmittelalters, Wien, 1980 (=

most certainly solid kind of research: if we describe how many items of a certain class of goods have been available at a certain castle, this is more or less what the source says – there is scarcely a chance for erroneous interpretation. The eternal question, asked of that kind of research, however, is: so what? What relevancy does it have to know, how many cloaks of a certain colour where available at a certain castle, if we do not know, if their number just expresses a private whim of its Lord or Lady – or, indeed, an anxious effort to keep up with some contemporaneous noble Jones? How representative is, whatsoever has been said?

(b) There are completely different contributions[3], which give a sweeping picture of a broad phenomenon of material provisions, of a social behaviour or of norms imposed by the society, based upon evidence selected from a wide variety of sources. The ability, to draw upon widely scattered examples, is usually seen as proof for the universality of the phenomenon described and the validity of the interpretation given. Nevertheless, each author has usually to acknowledge that, using such a vast array of diverse source material, she or he has not been able to go into a detailed appraisal of every source encountered; and indeed, even when we assume, that a few of them have been misunderstood, the validity of the overall picture will remain unchallenged. Still, it is a rare occassion, that such a broad synthesis of traces of a phenomenon does not evoke heated debates of specialists about the selected points they are familiar with.

(c) The third main type of argumentation consists, finally, of basing a hypothesis upon one central source, with which the author can reasonably argue to be so well acquainted that the basic model of interpretation is being derived from a streamlet of tradition claimed t b completel understoo – suc centra source being then augmented by proof of the occurrence of the observed phenomena in other sources, other areas and periods as well. Strengths and weaknesses of this type of study usually reflect the arguments given above, depending if the author has put more prominence to the individual source or to the secondary material accompanying it. (Obviously the precise differentiation between this type and the two preceding ones is difficult; we therefore avoid giving examples.)

Few readers will probably argue with the description just given; many however will say that what we said is not at all specific to research on everyday's life, but an inherent difficulty of historical research, so that the same typology might as well be used to describe

---

Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 3 = Österreichische Akademie der Wissenschaften, phil. Hist. Kl., Sitzungsberichte 367) 225-245.

[3] E.g. *Gernot Kocher*: Spätmittelalterliches städtisches Rechtsleben, In: Das Leben in der Stadt des Spätmittelalters, Wien, 1977 (= Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 2 = Österreichische Akademie der Wissenschaften, phil. Hist. Kl., Sitzungsberichte 325) 51-75. *Shulamith Shahar*: The History of Women in the Later Middle Ages – A General View and Problems of Research, In: Frau und spätmittelalterlicher Alltag, Wien, 1986 (= Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 9 = Österreichische Akademie der Wissenschaften, phil. Hist. Kl., Sitzungsberichte 473) 9-18. *Werner Rösener*: Zur sozialökonomischen Lage der bäuerlichen Bevölkerung im Spätmitttelalter, Wien, 1984 (= Veröffentlichungen des Instituts für mittelalterliche Realienkunde Österreichs 7 = Österreichische Akademie der Wissenschaften, phil. Hist. Kl., Sitzungsberichte 439) 9-47.

the kind of studies produced by any kind of historical research[4]. This is indeed so; still we consider the situation of the researcher of the daily life as a peculiar case: in almost all other fields of historical study we can, and rightly so, assume that knowledge tends to accumulate. That the relevance of a source discussed in isolation will ultimately become clear by other isolated sources; that the broad synthesis can become a frame of reference, to be used afterwards for checking successive findings, if they fit it or not; that an analysis of a central source can be expected to serve as a starting point for the understanding of other sources.

This will work wonderfully in (in the broadest possible sense) prosopographical[5] work: the activities of persons being restricted to a certain geographical area, we have no reason to doubt, that the immatriculation lists of a given university will be consulted by everybody who wants to trace the fate of the citizens of a community nearby. It will work great with all kind of legal systems and proceeedings; the kind of legal relationships between any two people can, after all, only take a limited number of forms. Even in a streamlined history of the economy, knowledge is constantly accumulated; a "price" being a very well defined quality, that lends itself perfectly to comparisons between studies[6].

As a result, all classical editions of source material strive for three aims: (a) to make more easily accessible the text of a document, which else could be consulted only at a specific archive, (b) to add to it all information about the entities mentioned in the sources, by hinting at other sources, where, e.g., the same persons and/or places have been mentioned as well and (c) to blaze a trail for the historian encountering the same task with another corpus later, by providing extensive registers of the entities mentioned in this source.

Now, all historians will agree that the nutrition of the populace has been a central question in daily life: is there any imaginable source, where a register of passages dealing with problems of nourishment would be provided by even the most careful editor, as part of his or her intellectual duty *except in cases where such an interest has been part of the reasons for the edition*? (As would be the obvious duty of every editor to provide a register of persons mentioned in a source, though being interested only in the legal formulae it employs). Is it, intellectually and economically, feasible, that editors make themselves acquainted with *all* aspects of material culture, so they can be relied upon, to develop the habit of indicating all exceptional references to food, architecture, clothing, furniture etc., in the same degree every trained editor of a corpus of sources will consider it necessary to indicate all exceptional legal constructions being encountered? And, particularly: is there

---

[4] Cf. *Manfred Thaller*: Praktische Probleme bei der interdisziplinären Untersuchung von Gemeinschaften 'Langer Dauer', In: *Gerhard A. Ritter* und *Rudolf Vierhaus* (Edd.): Aspekte der historischen Forschung in Frankreich und Deutschland, Göttingen, 1981 (= Veröffentlichungen des Max-Planck-Instituts für Geschichte 69) 172-189.

[5] Which therefore has produced a disproportionally large number of computer applications: *Karl Ferdinand Werner* (Ed.): L'histoire médiévale et les ordinateurs, München etc., 1981; *Hélène Millet* (Ed.): Prosopographie et Informatique, Paris, 1985.

[6] Indeed, a pilot project to make a directory of medieval and early mode currency exchange ratios, available on-line is currently implemented in the United States. See: The Medieval and Early Modern Data Bank, in: The Research Library Group News 12 (January 1987) 8-10.

any chance of this being done in sources, where such references would not be expected in the first place?

To sum it up: we tried to point out that historical knowledge in traditional disciplines, as political, prosopographical or legal history, has a tendency to accumulate, which is considerably stronger than just the obvious phenomenon that a later historian is able to consult what an earlier one has written. In many historical disciplines knowledge gained from one source will influence the treatment another source gets during the very process of being edited critically; if possible at all, this happens only very restrictedly with regard to the history of the daily life and the sources relevant to it.

What we would need to overcome this restriction, is essentially a technique of edition, where, together with the classical tools of critical apparatus and identifying registers, a possibility exists to address directly every portion of a source, which is interesting from the point of view of such a discipline as the study of the daily life – in its most primitive manner by having immediate access to every context, in which a certain word appears. In a utopian sense we imagine a situation, where a historian having found in a newly encountered source a term for some item of material culture not yet known to him or her, has a tool, allowing to access the same term immediately in a large number of sources.

The remainder of this paper argues that, while not feasible within the next year, the current developments in data processing make it right now at least possible to sketch, how such a tool has to look like, and how we can arrive at it. We are convinced, furthermore, that this preliminary conceptual work should take place now, that is before technical fait accomplis have been created.

This necessity, to discuss in detail the possibilities of modern information technologies now, we see particularly in two respects:

- We are at a stage now, where data processing has developed a dynamic of its own that allows it to dominate fields like type setting and publishing, quite without any concern for the feelings of the authors – *or critical editors* – whose labours are to be published. Indeed, one of the few prophecies about the development of the next years one might feel reasonably sure to make, is that within a short range of time every critical edition that becomes printed, will per definition be available at the same time in machine readable form, as every text to be printed will have been composed on some electronic device. We should explore now, if this situation creates additional possibilities for historical research, which go beyond books remaining (or becoming again) at least marginally affordable. As any argument like that is known to raise certain fears, it should be made emphatically clear that we do *not* speak of the end of the printed book: we speak about possibilities resulting out of new ways of producing printed books. We should do so all the more, because we else might discover some day, that the publishing trade has done everything that is lucrative from the point of view of a publisher, blocking other avenues in the process.
- This author seriously believes – and has discussed with regard to a number of topics in a recent series of papers[7] – that there are fundamental differences between information occurring in our present day discourse and information transmitted to us from earlier

[7] *Manfred Thaller*: Warum brauchen die Geisteswissenschaften fachspezifische datentechnische Lösungen? Das Beispiel kontextsensitiver Datenbanken in der Geschichtswissenschaft, In: *Man-*

times; if any of these suspected differences are true, they will have to be considered before applying any formal treatment of information to historical sources as such.

2. Retrieving Information Out of Historical Sources

Before we discuss, how an edition of a historical source might be made more transparent to different classes of historians, we would like to describe first what precisely it means to search for information[8] in a corpus of written material. All the tools described in the following sections have been applied to historical research[9]: when we describe them

---

fred Thaller and *Albert Müller* (Edd.): Computer in den Geisteswissenschaften, Frankfurt/New York, 1987 (= Studien zur Historischen Sozialwissenschaft 7); Ungefähre Exaktheit. Theoretische Grundlagen und praktische Möglichkeiten einer Formulierung historischer Quellen als Produkte 'unscharfer' Systeme, In: *Helga Nagl-Docekal* and *Franz Wimmer* (Edd.): Neue Ansätze in der Geschichtswissenschaft, Wien, 1984 (= Conceptus Studien 1), 77-100; Vom Beleg zum Begriff. Der Beitrag der Datenverarbeitung zur Lösung von Terminologieproblemen, In: Gerhard Dienes et al. (Edd.), Ut populus ad historiam trahatur. Graz, 1988; secundum manuscript. Zur Datenverarbeitung mehrschichtiger Editionen, In: *Reinhard Härtel et al.* (Edd.): Geschichte und ihre Quellen. Festschrift für Friedrich Hausmann zum 70. Geburtstag, Graz, 1987; Geographische Angaben in einer historischen Datenbank, In: Eratosthène-Sphragide 2 (1987).

[8] The author is well aware of the existence of the concepts of a data bank and, more specifically, of data base manipulation systems. Indeed some of his publications deal with modifications of current concepts in the information sciences necessary to make some of the more technical of these concepts applicable to historical research. This paper is directed to historians however; the author abstains intentionally from any discussion of information science topics. Indeed he is aware that he does – and intends to – blur differences between various levels of software products, like fully grown data bases, factual knowledge administered by data banks, full text retrieval systems and so on. He does so, because he believes it necessary to concentrate among historians upon the needs of the discipline, to describe – abstracting from existing software –, what kind of tools we need; when the practical consequences of such substantial decisions are discussed and implemented, he agrees fully with the necessity to discuss data basing and other areas of software on the same level of abstraction, as used in the professional development of software, and thinks he has done so. By simply popularizing, however, what is described in a computer science text book for the first years of a course in information science, one does not so much enlighten historians about data processing in his opinion, one rather obscures the fact that there are problems of data processing, which are specific for our disciplines. To the historian, who wants to inform her- or himself more thoroughly about the technical background, we recommend: On Data Basing: *Toby J. Teorey* and James *P. Fry*: Design of Database Structures, Englewood Cliffs, 1982; *David R. Howe*: Data Analysis for Data Base Design, London, 1983. On non-trivial data processing concepts: *William Kent*: Data and Reality, Amsterdam etc., 1978; *John F. Sowa*: Conceptual Structures: Information Processing in Mind and Machine, Reading, Mass. etc., 1984. On non-elementary programming techniques: *Jean-Paul Tremblay* and *Paul G. Sorenson*: Introduction to Data Structures with Applications, Auckland etc., [2]1984; *Billy G. Claybrook*: File Management Techniques, New York etc., 1983.

[9] For a first orientation: *Friedrich Hausmann et al.* (Edd.): Datennetze für die historischen Wissenschaften?, Graz, 1987; *Manfred Thaller* (Ed.): Datenbanken und Datenverwaltungssysteme

isolated, we do so, because in most, though not all, conventional computer systems[10], some or all of these approaches are mutually contradictory. If you have choosen one that is, you preclude yourself from using the others.[11]

2.1. Structural Access

Structural access we call that way of retrieving information from a source, which is probably the most familiar to all historians, who have had any exposure to computers at all. It works roughly according to the following logic:

(a) Historical source material – or information collected from a number of sources and/or from literature dealing with historical phenomena – is entered into a computer

---

als Werkzeuge historischer Forschung, St. Katharinen, 1986 (= Historisch Sozialwissenschaftliche Forschungen 20).

[10] In historical research, there exists virtually no agreement about the infrastructure necessary to apply a computer to historical research – therefore the most unreliable claim, that can be made in computer supported historical studies, is that a particular system is "extremely easy to use". Indeed, very prominent advocates of particular approaches have not formulated any command to a computer for years, have been convinced, however, by the ability of their staff of (mostly student) assistants to produce in due time the required results that, whatsoever this staff uses, it is very easy to use indeed. All our following remarks assume a different concept: that of historians chronically short of funds, who, far from being able to hire additionalstudents, employ the computer precisely to reduce the workload encountered by *themselves*. We concentrate in the following paragraphs, therefore, not upon the possibility to program computers in *procedural* steps: "read the data until you reach the first line, which contains an ampersand. Extract the following data until you encounter a comma, which is not contained between quotation marks. Store them, until you have read all the data. Now sort the data." Rather we tacitly assume that all the software, which we will explore, has the property to handle "logical objects": "Create a register of first names". (We are aware that this is only a very intuitive introduction to the meaning of "object oriented": a short explanation what the difference between the old "procedural" and the new "object oriented" style is like (though on a more elementary level) can be got from: *Patrick H. Winston*: The LISP Revolution, in: Byte 10/4 (April 1985) 209-218.)

[11] We apologize that the following sections may seem as bewildering to information scientists as, unfortunately, also to historians, who have just read a short introduction into data basing. The author is fully aware that the following classification does not run along the lines on which data structures are usually described within the classical "data models" of information science. The following approach has been chosen, however, because the author has experienced during the last decade that *these* are the difficulties, which are encountered by historians. The whole theory of data models of information science is outspokenly (and in the meantime also almost subconsciously) dominated by properties of data models, which ultimately stem from their differing behaviour, when exposed to problems of frequent on line updating. While these problems certainly can occur as well with historical data bases, an edition of a historical source, which shall be accessible by data base oriented tools, will have updates among its least central problems. We focus intentionally on problems, which are not related to update problems. Readers which, by exposure to dBase or similar systems, thought that the relational data model is the only one, might get a broader perspective by *Dionysios C. Tsichritzis* and *Frederick H. Lochovsky*: Data Models, Englewood Cliffs, NJ, 1982.

according to logical structure, which is assumed to be at least latently available in all units of material to be administered by the computer. An example for part of a charter entered in this way, might be:

```
sender$Friedericus/de Brunnense
recipient$Willelmus/de Guiningen
```

(b) To make use of any information contained in the data, you have to define its *complete structural location*. That is, you have for example to specify **name of sender**, if you want to address the name of the sender; a very useful feature, indeed, if you want easy means to distinguish between names of senders and names of recipients: a potentially calamitous situation, however, if you want to produce, say, a register of all "persons", that are both senders and recipients, in the data. In many data bases which provide for what we have called structured access, you would need for such a register at the very least one explicit command for every kind of persons you want to put into this register; in some it would not work at all.

This is no unsurmountable obstacle, if you are going to differentiate in charters between sender and recipients: but what about the private charter, where it is also mentioned that the recipient has a nephew, who, as soon as of marriageable age, shall marry the second cousin of the sender; after which event the legal proceedings, which are documented in the charters, are to be considered void? Data structured to take account of that, have usually to be made up of dozens – in extreme cases hundreds – of structurally different kinds of entities (let us for the sake of simplicity and clarity assume for this paper that an "entity" is something like an individual person). At the very best, it means that you have to write down dozens or hundreds of commands, if you want to access all of them – as, e.g., necessary to create a register. That quite a few software systems of that type are not able to combine information out of more than one kind of entity into a meaningful analysis, we just mention in passing.

We will not follow this in detail: let us just keep in mind that there are programs, which transform information into entities, between which very clear structural relationships exist; that this approach is very useful, if those structural relationships become central for the analysis; that it is definitely troublesome, if you have to ignore the structural relationships at least some times.

2.2. Semantic Access

So in the first approach, you divide your source into entities, which roughly correspond to people: and what type of people they are, becomes the central concern. The same approach, of course, can be taken with material objects: for example, a piece of furniture can be described by a structure in the sense given above, which simply divides the object further:

```
Object$Bed/Wooden
Part$Bedstead/Carved
Part$Coverlet/Silk
Part-of-Part$Embroidery/Italian
Part$Cushion/Red
```

Please notice, that we do not give any detailed instructions, how such data are formed: still most of the readers will have immediately understood that we just described a wooden

bed, which has as remarkable features a carved bedstead, a coverlet made of silk and a red cushion; the coverlet, in due turn, being distinguished further by having an embroidery of Italian origin (presumably stiched to it). The problem is that, while we obviously have to avoid the situation of the computer telling us that we have embroidered beds (due to this we made the embroidery a `part-of-part` of the bed only), it is entirely feasible that in the same group of sources we also have something like this:

```
Object$Cloak/Purple
Part$Embroidery/Italian
```

the embroidery, indeed, being a part of a cloak in a sense, in which it never could be part of a wooden bed.

According to what we said so far, this unfortunately would mean that, when asking for all embroideries appearing within our sources, we would have to memorize that an embroidery can occur as the name of an object, the name of the part of an object, the name of the part-of-part of an object, probably the name of the part-of-part-of-part of an object as well, and only heaven knows, as what else.

Why would a human being have less difficulties finding all embroideries? Because a human would differentiate between cases, where you are interested in the relationships between various levels of subdividing an object, and cases, where these subdivisions can profitably be ignored, tacitly assuming that the whole data above consist of a sequence of unrelated items, each of which is described by two fields (or descriptors, or variables, or ...) "name" and "quality".

For the purpose of this paper, we will it call *semantic* access, if a data base has the property of being able to access all entities, which have a field (or descriptor, or variable, or ...) known as "name" without regard to the structural relationships between such entities. The gain is obvious: in our preceding example, we would have not to bother any more about which kinds of persons are available, we would be able simply to say the computer: "produce a register of all entities which have a name" (or, more probably: "produce a register of all entities which have a personal name"). In theory, there is no difficulty with such an approach at all – unfortunately, however, in commercial software, there is usually a kind of trade-off between the two approaches: almost all existing systems are either good at preparing structural access to data or at giving semantic access; very few of them offer one without restricting severely the other.

2.3. Access by Content

Even if we are ready to ignore this unpleasantry, however, there is still another problem. Let us assume, we do research upon the fashion of the 15th century. We device for that purpose data, which consist of pieces of clothing, which are made up of variant structures of entities, which are interrelated somehow among each other, all having in common three fields, however, `name`, `colour` and `material`, as in

```
item$cloak/red/wool
```

Unfortunately this is not such a universal solution as we might imagine: in quite a few cases, the colour of material is an integral part of the description of its quality. A "cloak of red cloth from Flanders" might not only be of another colour than a "blue cloak of cloth from Flanders", but also be made of a distinct quality of Flandrian cloth, being

so fine that it is always associated with prestigious red. Now, dividing the information we have distinctly between `colour` and `material`, we give away some important information on the quality of the material; entering "red" alternatively into `colour` or `material`, we have to think less than straightforward in convincing the computer that we want to access all "red objects". (If we enter the colour in both, all theoretical problems are settled; we just get a headache in keeping the data consistent – particularly during corrections after initial data preparation.)

Similar problems are quite frequent in medieval data, by the way: is "Hubertus Friburger" "Hubertus", migrated to his current residence from "Friburg", or does he have a surname?

We solved structural ambiguities by concentrating on the names of the fields; now we have to ignore even these names. Indeed, there exists software, which is able to ignore all categories completely and to treat the content of the data base simply as a set of words, all of which are equally accessible; so we are interested only in the words occurring, in the *contents* of the data base; neither in its logical structure, nor in the differences between the various fields (or descriptors, or variables, or ...) describing the individual entities.

Such systems, usually known as *full text* or, more recently *free text* (retrieval) systems, have just one shortcoming: they usually insist that you ignore any structures that might be available – they will not only give you words, which consist of the same characters and might mean any of two things, they will give you words, which consist of the same characters, though we can be absolutely sure, that they have different meanings in their various contexts.

2.4. Access by Meaning

This last mentioned problem completely aside, there remains at least one additional difficulty. When we want to get all information available about "furniture", in all the access methods we have discussed so far, we would have to specify very explicitly what we mean by furniture; in fact we would have to specify "beds" and "tables" and "chairs" and ...; and, unfortunately, we would also have to specify "beddes", "tabls" or whatsoever other variation in spelling we encounter in our sources. Now, when we do this once, we can in very many computer systems tell the machine, that it shall remember these definitions of "furniture" – a capability, which is usually implemented as a "thesaurus", often clever enough not only to understand that certain things are related to a common concept, but also that they form part of each other and more sophisticated relationships between the meaning of words[12], contained within a corpus of sources.

All of the three access methods described so far can be enhanced considerably by the use of such thesauri; even more so, when we employ a system, which combines a thesaurus with what is known as a *lemmatization* program, which, roughly, reduces any wordform encountered to a word stem before using it. This reduces significantly the effort needed to specify minor variations resulting from different grammatical usage as being equivalent in meaning for the purpose of the historian. With software, that allows structural access, we would therefore be able to ask for "objects with a name that signifies 'furniture' "; with

[12] On specialized tools for the administration of terminological information see *J. Goetschalckx (sic)* and *L. Rolling* (Edd.): Lexicography in the Electronic Age, Amsterdam etc., 1982.

software allowing semantic access, we could ask for "something with a name signifying 'furniture'"; and with content-oriented access, we finally would be able to ask for "contexts in which 'furniture' occurs".

Still, this access method, too, is not without drawbacks: quite besides thesauri being usually administered in a way, which makes changing them not all too easy, such software often is somewhat restricted with regard to the number of terms it can handle: that makes thesaurus supported systems, which handle – not in theory, but in practical operation – very restricted vocabularies, less useful for historians, who want to transcribe a source literally, and favour those, who use a pre-defined number of keywords.

And here we would like to put an axiomatic statement: a very different situation is created by any computer system, which accepts the words the user needs to employ without limitation, when compared to systems, which force the user to select phrases from a limited and restricted vocabulary. *There is* no conceptual *difference, however, between systems, where you simply have to enter code numbers and such, where you have to select out of a predefined set of keywords.* The first distinction may create the ability to handle a source as it is; the second provides an optical advantage, which may be very important psychologically, but does not bring the data closer to the original at all.

2.5. Data Revisited: Input Conventions

Before we discuss, how out of these techniques eventually a tool could be derived, which is appropriate to turn sources into computer administered editions that can be understood to consist of an almost infinite number of cross referencing systems, for every aspect of the daily life, we should go back to the question of input[13].

2.5.1 The "Structured Input" Approach

We closed the preceding section with a very strong statement in favour of source orientation: in favour of entering source material as close to the original as possible. Now:

```
Object$Bed/Wooden
Part$Bedstead/Carved
Part$Coverlet/Silk
Part-of-Part$Embroidery/Italian
Part$Cushion/Red
```

will certainly not appear as part of a medieval document, which rather would read:

"Item, a bed made of wood, with a picture carved nicely into the bedstead, all covered with a cloth of silk, in the middle of which a beautiful embroidery, done by a Florentine maiden, appears; also a red cushion on top of the bed."

Now "sticking closely to the source" can mean, as we will see soon, that indeed we keep the text in the second form. Most historians will admit, however, that the quotation as such contains quite a few terms, which are of doubtful relevance to all but purely philological research. If the "beautiful embroidery" "appears" in the "middle of which" or simply "is there", may have some relevance for intellectual history – or indeed, as mentioned, for

[13] *Manfred Thaller*: A Draft Proposal for a Standard for the Exchange of Machine Readable Sources, in: Historical Social Research / Historische Sozialforschung 40 (●ctober 1986) 3-46.

philologists using the text –, but will not usually lead to important insights regarding the daily life of the inhabitants of the bed in question.

Now, what is usually irrelevant in this sense, is also hardest to understand for a computer: that is, the relationships created between various parts of a text with the means of grammar, associative arguments, elliptical expressions, and so forth. Therefore, a well trodden path of *source conservative* or source *oriented* data processing consists simply of replacing these grammatical relationships by a structure of abstractly defined units like "object", "name of such" and so on, which are indicated by a system of limiting characters (in our example dollar signs and slashes). The vocabulary we put into these structures is completely free; so

```
Object$Bed/made of wood
Part$Bedstead/a picture carved nicely into the
Part$covered with a cloth of/Silk
Part-of-Part$embroidery/done by a Florentine maiden
Part$cushion/red
```

would be as acceptable as our first construction for almost all systems which accept this first one at all.

This decision, of restricting oneself to the semantics of a source, forgetting about the syntactical intricacies by which the language has been formed, we will refer to henceforth as the technique of structured input; its limitations are probably obvious. The reason, why such a logic is often recommended, is simply, because there are some computer systems[14] already, which are perfectly able to understand it. – As we will see that those other solutions have their shortcomings as well, however, for all sources, which have a relatively orderly structure, like account books or similar, structured input has indeed some virtues of its own.

2.5.2 The "Pre-edited Input" Approach

Still, there may be some reason, that one really wants to catch the source just as it is. A solution for this might be the following:

```
''Item, a (O (N bed N) made of (M wood M), with a picture
(O (Q carved Q) nicely into the (N bedstead N) O), all
(O (N covered with a cloth N) of (M silk M), in the middle
of which a beautiful (O (N embroidery N), (Q done by a
Florentine maiden Q) O) O), appears; also a (P (C red
C) (N cushion N) O) O) on top of the bed.''
```

Indeed software systems understanding this kind of text, which consists of objects (O ... O), which can be described by name (N ... N), material (M ... M), quality

[14] *Manfred Thaller*: κλειω . A Data Base System Specific for the Historical Disciplines, Göttingen, 1987. The current version is available from the Max-Planck-Institut für Geschichte, Hermann-Föge-Weg 11, D 3400 Göttingen; future editions will be distributed by the regular channels of the publishing trade. A German version of the manual is available as well; a French one currently being prepared. Cf. *Kevin Schurer*: Historical Research in the Age of the Computer: An Assessment of the Present Situation, Historical Social Research / Historische Sozialforschung 36 (October 1985) 43-54.

(Q ... Q) and colour (C ... C), and *may in turn contain further objects*, have been implemented[15].

This form of preparing a source for the computer, known as *pre-edited input*, is undoubtedly closer to the original than the earlier one. Also, it saves the necessity to restructure the text – you can simply transcribe it, as it occurs. The number of programs able to cope with such input, is incomparably smaller, however, than the ones that can handle the earlier one. Beyond that, this form of entering data has one major shortcoming, which may depend on the experience, one has got with using it, is viewed as considerable by many people, however: if you expand the principle to more complex constructions – and the ability to do so is the main point of starting with it – systems of *logical brackets* or *start-stop symbols* can become easily very complex, and it is extremely time consuming to find out, where in a lengthy expression, which the computer considers erroneous, a particular bracket has been forgotten.

2.5.3 Mask Oriented Input

"So why do we not simply buy an ATARI? There it is all so much easier; you just are presented with a mask and fill into it whatever you have to. That is, the computer generates on its screen a form, as would be a questionnaire on paper, where you are asked for every piece of furniture, which name, quality, material and colour it has, and which parts it consists of, repeating the same questions for every part." Well; – one answer might be: try it. We have intentionally not mentioned a very trivial fact so far: any kind of truly source oriented program assumes flexibility in a number of ways. Reasonably advanced systems of source oriented data processing all allow each field to have an almost arbitrary length; commercial (and that includes most mask oriented) programs usually allow only predetermined lengths of fields, often heavily penalizing users who fill in only part of the field. Systems of structured and pre-edited input all allow constructs of more or less unlimited complexity, as long as they are made up of the structural items or start / stop symbol sequences you defined; for mask oriented input you have to prepare the most complex structure to be possibly handled much more in detail in advance.

However, we introduced the question at the beginning of this chapter as colloquially, as we did, to avoid even the slightest hint at intellectual propriety. Indeed, asking the question as we did, in the context we introduced it, is one of the easiest ways to convince everybody with more than a superficial knowledge of data processing that the historian asking it does certainly not know anything about computers at all.

One of the most important concepts in data processing, which is often obscured by introductions into the field gained via the user's manual of a software system or a micro computer, is the difference between the *background system* and the *user shell*. Indeed, we blurred this difference ourselves so far: important, from a data processing point of view, are the classes of access to be possible within a system (and on a level too technical to be discussed here, the logical properties a data structure has to have to make the chosen ways of access possible). The parts of a programming system responsible for the administration of the system, and those, which decide, which of the characters the user

[15] See *Kevin Schurer* as in the last note; *Timothy J. King*: The Use of Computers for Storing Records in Historical Research, in: Historical Methods 14 (1981) 59-64.

has entered shall be stored where, can be seen completely independent of each other. At least conceptually, every program system *parses* through the input data to divide them into meaningful segments of information, which have to be delimited by some convention, and, as soon as a suitable number of such segments has been encountered, they are stored in a data structure that has few similarities to the way the characters are being typed into the machine.

So, for the programs responsible for selecting out of the source those items of information we need, it is perfectly unimportant, if we entered the data via a file of structured data, as a pre-edited text or by using the mask-oriented "shell" of a particularly user-friendly program – as long as that mask generator is flexible enough to handle the data structure[16].

"Parsing", as the process of deciding, which part of the input data shall be considered to mean what, is called, indeed can go to very great complexities: the term is most frequently employed in computational linguistics, where the "conventions limiting the different entities and signifying relationships between them" are the grammatical rules of natural language; even on such a level of complexity we still have to decide, however, what are the modes of access, which we want to employ. (And computational linguistics is very far from a point, where we could expect programs, which "simply understand" arbitrary text to store it safely into a logical structure, even in contemporaneous, leave alone historical, language.)

2.5.4 Books vs. Data Bases

Why, then, this lengthy discussion? We mentioned already, that one of the reasons to present this paper now, is, that many historical sources are typeset by computers, that is, are available in machine readable form: and what we said before is necessary to understand, how a source prepared to become a book, could by specialized software be made accessible along the lines we mentioned for research into the daily life of the Middle Ages, as well.

To explain this, we have to point out, that the conventions of pre-edited input, which we described before, are, indeed, rather similar to certain conventions being used for the preparation of computer supported typesetting. For this purpose we compare two examples.

---

[16] We hold the frequent misunderstanding that a beautifully looking screen implies great power of the data structures being administered, for sufficiently dangerous to feel responsible for a warning in clear terms. Of course the relationships are more complicated: in cases, where software systems, which handle structured input, simply expect texts entered with a text editor (or word processor) there is nothing that prevents a historian to use an arbitrary mask generator for smoothing the input process, simply storing intermediately as characters indicators for structural relationships, which are to complex for the mask oriented shell to understand. Still, we would like to repeat our warnings against programs which are all beauty and no substance. It is very typical that in *Roland und Marcus Gröber*: Die Anwendung des Computers bei der Auswertung von Kirchenbüchern, in: Computergenealogie 2/5 (1986) 135-139, the use of a mouse and even a light pen is seen as more or less mandatory, but the authors – otherwise very computerwise indeed – have obviously no knowledge at all about even such trivial techniques of *historical* data processing as the handling of names with varying orthography.

```
Huius testes sunt: {\italics Friedericus de {\bold Cawenberghe}},
{\italics Cuonradus de {\bold Gravestetten}} et {\italics Markwardus de
{\bold Seyfring}}.
```

would in the typesetting system[17], that has also been used to produce the present article, result in

Huius testes sunt: *Friedericus de* **Cawenberghe**, *Cuonradus de* **Gravestetten** et *Markwardus de* **Seyfring**.

Now, we do not want to argue that the idea to print names of persons in italics and names of places bold, is a necessarily good one: what shall be illustrated, however, is that a convention like the one above, which just uses different fonts to emphasize systematically portions of an edited source, can be immediately translated into the input convention of a programming system, which is able to parse pre-edited text, as in:

```
Huius testes sunt: (person Friedericus de (place Cawenberghe)),
(person Cuonradus de (place Gravestetten)) et (person Markwardus de
(place Seyfring)).
```

That is: the kind of information, that has to be provided to produce a book, which uses any kind of semantically meaningful rule for switching between different fonts, can at the same time also be used for the various modes of access we discussed before.

2.6. Access Revisited: Retrieval Languages

We introduced first different properties of various ways to access the information within historical sources; then we described different ways to prepare source material for that purpose and explained that the two areas are completely independent of each other. Now, access does not only consist of the effort needed to prepare data for it, but also of the knowledge one has to have to specify, what information to access. Once again, the specific tools provided by a programming system, to specify what it has to do, are independent of the kind of data structures it can handle. Indeed, there are programming systems, which provide more than one way to access information.

We will illustrate this point with the help of a programming system[18], which in turn is discussed further below. For this illustrative comparison, we assume that we want to select out of a data base all sentences appearing in a set of charters, which contain a word starting with `rubr` ("rubrum" et sim.), to find out all descriptions of red objects – this would be necessary, if we want to evaluate the importance of an object having a specific colour, an interest for which the editor of the chartulary has almost certainly provided no tool at all.

2.6.1 Menu Driven Systems

The first way of asking the computer to do so, consists of telling it vaguely that we are interested in examining the text of the chartulary, being then supplied by the program

[17] TEX. Besides the beautiful description given by *Donald E. Knuth*: The TEXbook, Reading, Mass. etc., 1984, which despite of all its beauty is next to useless for the beginnner, there exist now, fortunately, a few very good introductions like *Norbert Schwarz*: Einführung in TEX, Bonn etc., 1987.

[18] $\kappa\lambda\epsilon\iota\omega$, as in note 14.

with tables, which describe what is possible at a certain moment; as we pick from these tables what we want the machine to perform, they are usually known as *menus*. To do so, we need three pieces of information: (a) Where the text of the chartulary is stored (we will assume that somebody stored it as `chartulary`), (b) under which name the reference list of all words in that chartulary is stored – as opposed to an independent register of, e.g, persons – (we will assume, that somebody stored that list as `words`) and (c) how to activate the program.

As computers are slightly different, this third step will vary; so for the time being we just assume, we know how to do that. As soon as we activated the program, we will read on the computer screen:

```
Please select one of the following activities:

b     Selection of a new data BASE
r     Selection of a new REPERTORY
w     Selection of a new/additional WORD form
c     Redefinition of the CONTEXT for search and/or display
d     DISPLAY of the current list of quotations
s     SAVE list of quotations in a quotation library
m     MOVE the cursor within the quotation list
@     Leave Kleio
*     Leave this menu

Please enter the selected character and press ENTER/RETURN.
```

Understandably enough, the computer has to know, which source we are interested in; as he considers a source to be a data base, we type the letter `b` on our keyboard and touch the key that is labeled ENTER on some machines, RETURN on others and TRANSMIT others still. (To be a little bit more concise, we will speak about the ENTER/RETURN key from now on.) The machine reacts with

```
Please enter the nsme of a new data base:
```

to which we respond by entering `chartulary` and pressing ENTER/RETURN once more. Now we have decided about the source, and have to tell, which of the potentially numerous specialized registers we want to consult. Still having the initial menu in front of us, we enter `r`, hit ENTER/RETURN and are greeted by:

```
Please enter the name of a new repertory:
```

which we supply as `words`, hitting ENTER/RETURN again. Still in front of the initial menu, we decide, that we are finally getting to the colour red now and enter `w`, followed by ENTER/RETURN. This results in the computer pleading:

```
Please specify a word form or a quotation list.
```

to which we react by typing `rubr` and ENTER/RETURN. Now the menu, that has accompanied us so far, changes. Our screen looks like:

```
Please select one of the following activities:

c     Look for a COMPLETE word form
```

```
b     Look for words BEGINNING with this text
e     Look for words ENDING with this text
l     LOAD a previously stored quotation list
@     Leave Kleio
*     Leave this menu

Please enter the selected character and press ENTER/RETURN.
```

As we want to know about rubrum as well as about rubra, rubro and all the other forms, declension and the fantasy of a not necessarily perfect Latin scribe has provided, we enter b, followed by ENTER/RETURN. This action results in:

```
Please select one of the following activities:

w     Redefinition of the list of quotations by a new WORD form
a     AND: restriction by an additional word form
o     OR: expansion by an alternate word form
n     NOT: restriction by a negated word form
@     Leave Kleio
*     Leave this menu

Please enter the selected character and press ENTER/RETURN.
```

This is less immediately comprehensible as the previous menus; it hints at further possibilities (like selecting sentences, which contain any of twelve different words, but not any out of five other ones). For the time being, we just select w, as indeed we just selected a new word form, and press ENTER/RETURN, which brings us back to the original menu, that is to

```
Please select one of the following activities:

b     Selection of a new data BASE
r     Selection of a new REPERTORY
w     Selection of a new/additional WORD form
c     Redefinition of the CONTEXT for search and/or display
d     DISPLAY of the current list of quotations
s     SAVE list of quotations in a quotation library
m     MOVE the cursor within the quotation list
@     Leave Kleio
*     Leave this menu

Please enter the selected character and press ENTER/RETURN.
```

where we now select d (plus ENTER/RETURN) to look at the quotations dealing with red, which we just selected. As a result we are asked:

```
Please select one of the following activities:

s     Display of the list of quotations on the SCREEN
p     Display of the list of quotations on the PRINTER
m     MOVE the cursor within the quotation list
```

```
w      Selection of a new/additional WORD form
c      Redefinition of the CONTEXT for search and/or display
@      Leave Kleio
*      Leave this menu

Please enter the selected character and press ENTER/RETURN.
```

Now, before we start wasting paper, it might be useful, to see if the quotations are really worth to be printed, so we type **s** (ENTER/RETURN) and are asked:

```
How many quotations shall be processed?  (Zero = last menu)
```

We decide that we just want to look at the first three selected quotations, enter 3 (ENTER/RETURN) therefore and have them in front of us.

Most readers will probably feel sure by now that, by experimenting a bit further, they can convince the program to print the quotations dealing with red things, if these are interesting enough to merit the paper.

We went through this process in such detail to illustrate two points: on the one hand, what we have given as a description, would for most purposes be sufficient already to work with the computer. On the other hand, it is probably plain by now that, if so many different steps are encountered in looking for and displaying words, there is ultimately a limit to what we can do with such a logic: as, whosoever wrote the program, had to be able to foresee all possible junctions of the search process, where the user might want to select one of two or more alternative courses. So menu oriented access is extremely easy to use; but inherently restricted in fexibility.

2.6.2 Command Languages

The other extreme would be, simply to device a command language, which allows users to ask for everything, that can be expressed in the rules of the syntax of this language. In the command language of the system we are discussing here, which is available as an alternative to the menu-oriented shell we discussed before, the same kind of quotations, which we just selected, would be printed, if we give the commands:

```
quaero nomen=charters; pars=:text="rubr"
scribe
```

Now, such an approach, as the simple comparison of the length of this section with that of the previous one shows, is much more concise: and at the same time much more powerful to express complex wishes for the computer. To master all the tricks of a command language, takes, however, usually much more than the thirty minutes most people need, to get the impression that they control a menu driven program.

3. Requirements of a "Historical Workstation"

So far we have described what options are available in various fields of computer systems, where trade-offs between the various advantages and drawbacks of different approaches have to be made. On the back ground of these descriptions – and still having in mind our vague concept that, what research about the daily life of the Middle Ages would need, is a kind of "super accessible edition" –, we will first try to define a best of possible worlds and finally see, how realistic it is to expect entering it in the not too distant future.

Computers are applied throughout historical research; many specialist fields pose their own specific problems – be it source criticism, research upon legal terminology and others. In the introduction, we tried to describe that in our opinion the most specific problem of research into the daily life is the need to use very many different sources with very many shifting foci of attention; this translates into the need to employ all the access methods we described above concurrently: a need that is emphasized, therefore, in the following sections. Still, the problems, that are central here, *can* be solved isolated: to produce a meaningful result, within the context of other historical disciplines, they *should* rather be solved in an attempt to address oneself to the specific needs of the historical disciplines at large. In an attempt, that is, to study the possibility of creating an optimal tool for all fields of historical research.

3.1. A Popular Definition

Popularly speaking, a specifically historical tool for computer supported research can best be described by the way it should be used by the historian.

We assume that, as a by-product of the continuing office automation throughout the university, it will be possible for every researcher who wants to do so, to have at his desk a small computer, used at least as a replacement for the typewriter placed there so far. Such a small, personally controlled computer is in many academic disciplines increasingly often fitted out with specialized software tools that are geared towards the particular needs of a given discipline: in some engineering disciplines such "specialized needs" concentrate upon the possibilities to create three-dimensional drawings on the screen; in some philological disciplines the provision of lexica in machine readable form is considered as particularly useful. Such a personally controlled computer, together with the software components dedicated to the needs of a specific discipline, is usually known as a *workstation*[19].

What would for the historical disciplines be the specific requirements of a workstation? What do we suppose a historian to do with it?

- A historical workstation will obviously need specialized text processing programs, which provide for multiple apparatus, characters not frequent in present-day studies – like ŏ and similar symbols. So it *should* become a sophisticated replacement of the typewriter.
- The workstation, however, should also allow to access source material in a more thorough way than possible so far. This we can understand roughly as follows: the historian interested in furniture, who just heard, that a colleague has completed an edition of the wills of a given city, should be able to receive, together with the stately bound volume of the printed edition, a magnetic representation of the same source, which can be made accessible to the workstation: enabling the historian, in turn, to access systematically all phrases in these wills dealing with furniture.
- When, doing so, our potential user has selected a number of source references as important for his or her own research, the workstation shall enable him to transfer all the portions of the source relevant to the specific topic into a collection of "excerpts"

[19] A good introduction into what a "workstation" is in the publishing trade, easily comprehensible for a historian, is given by *John W. Seybold*: The Desktop-Publishing Phenomenon, in: Byte 12/5 (May 1987) 149-154.

– i.e., the "own" data base of the researcher –, which is continuously updated and changed, as new material is collected from various sources.

- At least during the second stage of this process – but more probably already at the first –, the user of the workstation will probably encounter some portions of the material, which it is not possible to understand; or which are suspected to merit, in the context of the specific research, more attention than the original editor of the source has paid to it, or has been able to pay. An example for the first situation would arise, if our researcher wants to compare prices in the newly encountered source with prices in material familiar to him or her: as relevant research may have progressed since the time of the edition, the currencies and units of measurements used should be automatically connected to the most recently available treatment of the meaning of such terminology. The second case can occur, if the researcher suspects, that additional insight into the source may, e.g., be got by comparing systematically the names of persons in the source with prosopographical information, the original editor did not consider relevant; or did indeed not have access to, as it just emerged as recent result of the own efforts of the user of the workstation.
- Finally, the workstation should enable the historian to draw "services", which are not so central, but useful: preparing statistical descriptions or analyses of suitable phenomena in the source; drawing, with the help of the machine, series of maps, showing the development of the spatial distribution of some distinct component of material culture in a series of documents spread across a larger span of time and space.

The beginning and the end of this list is indeed in principle available right now: typesetting and/or word processing systems exist, as well as a large array of software for statistical or geographical/topographical display. While the integration of these various components into a workstation, that can be used by an average historian, is a problem of quite some difficulty, it is ultimately a purely technical problem; and, though some of the problems mentioned have a specific historical dimension, we can more or less rely upon them being solved in principle by the commercial developers of computer related tools – as the solutions of these problems can be sold to financially much more potent parts of the public than the historical research community is.

Remain the problems of the accessible editions and of access to specialized knowledge: together with the question, what level of knowledge may be demanded from a historian to make such an approach useful to him or her. Indeed, this may be the central question deciding about the feasibility of it all. There have been examples of large specialized reference works[20], turned into conventional data bases in relatively recent years, which afterwards have almost remained unused, because the results to be gained from accessing them did not seem worth the effort to learn how to do so. We, therefore, suppose that all services of a workstation have in principle to be accessible by what we described before as a menu driven user shell. We assume, that it is a useful addition to the tools available

[20] Almost pathetic: *Ben Ross Schneider, jr.*: The London Stage Project: its Status and Future, In: *Joe Raben* and *Gregory Marks* (Edd.): Data Bases in the Humanities and the Social Sciences, Amsterdam etc., 1980, 31-34.

to a researcher, if, encountering the name of a person about which he or she holds no previous knowledge, it is possible to consult within five or ten minutes a prosopographical data base, where this person may be mentioned with some marginal probability, getting a result – or rather information, if it is sensible to look further for information dealing with such a person – within another five to ten minutes; *if and only if the knowledge how to do so can be used for similar consultations and can be acquired initially within* an *hour* or *less*. The greatest problem with existing historical data bases, in our opinion, is that they do not favour their *casual* consultation. There are many historians, who would occasionally like to consult a data base, holding some specialized historical information, as, e.g., a data base of all pictorial sources of medieval Austria: there are extremely few, however, who are ready to spend a day or two in learning how to do so, as they know that, while such a consultation may be extremely important for a question occurring at a specific point of research, such a question is improbable to turn up in short intervals.

3.2. Libraries of Sources ...

So, the first condition for a useful workstation, that endeavors to make accessible source information, is purely quantitative: there exists probably no single historical source, which is so central that even the most optimal access to it would be a sufficient reason for historians to make themselves familiar with a new working environment just for the sake of this source.

We envisage, therefore, a workstation, which contains a general purpose data base system, which is able to access a large number of data bases: these data bases being available on individual magnetic media – to make things more concrete, we can envisage the now available floppy disks, though the technology involved would ultimately have to be somewhat different[21].

Now, this data base system has to be completely hybrid with regard to the various access mechanisms, which we described in a preceding section, and has to support all access types equally well. That is a necessity, which has extremely important consequences for the person designing the system (which shall not be discussed any further here, however, because we think that the reasons for this decision – as well as why it is a very central one – should be clear by now).

Even the best data base system is useless, if there are no data: so we have to make provisions, making it probable that historians are willing to contribute machine readable sources for the workstation – or indeed, discover that after having edited a source in a basically traditional way, the source has become during that process available as a machine readable data base, and can now be used with the workstation as well.

We propose the following behaviour of the data base system to be used by the workstation: as background system we propose to employ one, which administers all data as a structured network of semantic terms (providing structural and semantic access parallely,

[21] The so called "laser disks", more correctly: "optical digital storage media", which right now are becoming standard hardware. On considerations regarding historical research see: Immo Gathmann: Datenbanken und Informationsverwaltungssysteme. Probleme ihrer Implementierung auf nur einmal beschreibbaren Speichermedien, in: Historical Social Research 41 (January 1987) 76-87.

therefore). Access by content shall be possible; every individual historian has the tools and the responsibility, however, to create those selections of fields and of terminology, which shall be accessible by content – as *there exist historical disciplines, like the study of the daily life and the material culture, which have to access their sources with so rapidly varying emphasis that it is not sensible to expect any historian to provide all possibly needed access paths through the sources in question.* By similar reasons, the user of the workstation shall have the possibility to refine structural and semantic means of access; while it is not possible to decide beforehand, which parts of the vocabulary of a source may be central to a given historian, we can assume that there exist generally definable entities – like persons, material objects, pieces of land –, which are of some importance to virtually every historian, and can be described by categories – like name, location, age –, which are of equally general interest. This is not the case for any access method related to meaning, as the specific relationships between categories of explanatory frameworks and the vocabulary of the source are the very aim of the studies to be undertaken; particularly in a discipline like the study of the daily life. For these purposes, a workstation shall therefore provide tools for the administration of thesauri, and general tools related to the philological background of the sources, as, e.g, tools for lemmatization. The sources themselves shall be made available in a non-lemmatized form and be always accessible without recourse to any thesaurus, even if some machine readable sources thesauri are available optionally.

We assume that a system of structured input can be most easily made flexible enough to be sufficiently powerful for providing all kinds of access modes simultaneously. As the development of the last ten years seems to show, however, that the easier availability of systems supporting structured input directs an increasing number of historians also to the analysis of sources, which in fact would be served better by systems using conventions of pre-editing; and, furthermore, as an increasing number of historical sources is bound to be available prepared for electronic typesetting, we propose that the workstation shall contain a parser that can convert pre-edited input, supplied by a user, into structured data to be processed later by the input modules of the data base acomponent of the workstation. This modularity should result in much more flexibility than an attempt to integrate both kinds of parsing mechanism into the data base components directly.

3.3. ... and of Background Knowledge

While we did assume so far that a workstation shall be able to access as many individual machine readable sources as possible, but will be used to access each of them only relatively rarely – or at least by very few historians more often than occassionally –, there exists "historical knowledge", which is generally enough to be applicable to many sources, as information on chronology, currency systems (and indeed prosopography, if more narrowly defined prosopographical information is extracted from the context of the entire source). Such knowledge can – and indeed has been – administered by dedicated data bases. We assume that a historical work station shall contain information of that kind enveloped in specific programs (roughly: so called "expert systems"[22]), which allow the

[22] Cf. *Richard Ennals*: Artificial Intelligence. Applications to Logical Reasoning and Historical Research, Chichester etc., 1986.

historian to consult them, instead of printed information dealing with the same subject. Why would it be potentially more sensible to consult an "expert system" on chronological problems instead of simply a book? One might quote increased comfort, greater flexibility and so on: for all practical purposes, however, we do not think that all these arguments together will convince any historian, who does not simply *like* computers to begin with. The real importance for such integrated expert systems, in our opinion, is to be found in two special aspects: on the one hand, the kind of knowledge we are talking about now, is precisely that one, where the "addidivity" of historical research, about which we talked in the beginning, is most important. If one historian discovers that a certain counting unit means something different from what so far has been assumed, this discovery has a more practical effect upon all other historians than even a very detailed study about a long chain of events in some community, as it can immediately be applied in any context, where this counting unit occurs. Printed compendia, however, are very cumbersome, when it comes to take notice of such singular discoveries. This, indeed, *is* one of the areas, where an electronic tool, by virtue of being updated more easily, is simply more accurate than a traditional one.

The other field of application, and this is the most important one, lies in the immediate application of such "background" knowledge to a given historical source: e.g., by telling with one command, that prices within the source shall be interpreted according to a given table of identified currency units – or, in a very advanced version, by automatically comparing every person appearing in a source with a large prosopographical data base, to ease the identifying tasks of the editor of the source.

3.4. The User Interface

Now, in some of the preceding paragraphs, we emphasized that tools like the one we describe here, are useful only, if the historical user can learn how to control them in a minimal amount of time; and apply, what has been learnt once, to very many different sources. In the preceding sections we explained in detail, what we consider an easy, menu oriented way to approach a source: we tried to explain also, however, that menu oriented systems, easy to use as they are, have certain inbuilt restrictions regarding flexibility. This, therefore, would precisely be the approach *not* to take, if the resulting workstation shall be handling as many different classes of sources as possible. Indeed, out of our very first arguments it follows, that while a "genealogical work station" or a "workstation for an economic historian" might be conceptualized very much as an extremely specialized tool, the distinguishing property of a "workstation for research on the daily life" would have to be, that it is as unspecialized as possible – as what we gain here is not (as in genealogy) the possibility to identify persons, or (as in economic history) the one to concentrate specifically on the development of prices, but the possibility to apply as many different tools to as many different types of sources as casually as possible. – Which in due turn makes such a workstation, of course, much more useful to the historical disciplines in general than a more narrowly defined one could ever become.

This particular circle can, in our opinion, be squared by distinguishing clearly between two situations: on the one band, we have the historian concentrating upon a certain source, e.g., an account book; on the other the historian, who, using many sources at a time, wants to use that very same account book, as part of his general base of sources, casually. The

"casual user" wants to be bothered as little as possible by specificities of the workstation; the "accounts book specialist" wants to use as much of the workstation's power as possible. In our opinion, the solution consists in building the workstation basically upon a powerful command language, and put a menu driven system on top of it, which translates the menu items chosen into the command language. The casual user will never transcend below this uppermost of the system; the specialist engaged in getting the maximum out of a source, will have the possibility to do just that, by "looking behind" the menus. – And, if a mechanism is provided to influence the menus presented to the next "casual user", who is accessing the same source, the specialist, preparing a particular source for his own purposes, will, as a side effect, indeed prepare the material for such a more casual user as well; that is, undertake the classical editorial tasks.

4. Possibilities for Realization

So much on the concepts; it would be pointless to describe them in such detail, if they would not be related to some on-going work. Indeed, that is the case:

- A *data base managment system* that includes all the principles discussed here, is currently being developed, under the name $\kappa\lambda\epsilon\omega$, at the Max-Planck-Institut für Geschichte, building upon the experiences gained there with an earlier system known as CLIO since 1978.
- The description of the *workstation* presented, forms part of a project, currently being formulated at the same institute, negotiations with a number of other institutions about possible forms of cooperation being at this stage the main focus.
- The software needed to *convert pre-edited input into structured one*, has been planned pretty much in detail and is supposed to be developed within the framework of the subgroup on standardization of the Association for History and Computing[23].
- Prototypes of Public Domain Data Bases, which can be accessed by the data base management system $\kappa\lambda\epsilon\omega$, dealing both with sources proper and with the background knowledge necessary to handle them, are currently being prepared by a number of research institutions[24]. They are supposed to become available during the spring of 1988

[23] Further information on the Association and its standardization subgroup, which is preparing a comprehensive publication on all problems regarding the exchange of machine readable historical data, is available from the author.

[24] Institut für Mittelalterliche Geschichte, Universität Freiburg/Breisgau (PD Dr. Dieter Geuenich); Forschungsinstitut für Historische Grundwissenschaften, Universität Graz (Dr. Ingo Kropač); Historical Institute, Hebrew University Jerusalem (Dr. Michael Toch); Institut für mittelalterliche Realienkunde, Krems/Donau (Dr. Gerhard Jaritz); Stadtarchiv Regensburg (Dr. Heinrich Wanderwitz / Bettina Callies); project-group "Migration und horizontale Mobilität im Mittelalter" (Dr. Peter Teibenbacher, Graz)

to test, how the distribution and the *casual* use of machine readable source material can best be organized.

Address all communications to:
Manfred Thaller
Max-Planck-Institut für Geschichte
Hermann Föge Weg 11
D-3400 Göttingen

MEDIUM AEVUM QUOTIDIANUM

newsletter 10

Krems 1987

Herausgeber: Medium Aevum Quotidianum. Gesellschaft zur Erforschung der materiellen Kultur des Mittelalters. Körnermarkt 13, A–3500 Krems, Österreich. – Für den Inhalt verantwortlich: Univ.Prof. Dr. Harry Kühnel. – Druck: HTU–Wirtschaftsbetrieb Ges.m.b.H., Karlsgasse 16, 1040 Wien.

INHALTSVERZEICHNIS / CONTENTS

Vorwort

Das vorliegende Heft 10 von Medium Aevum Quotidianum–Newsletter widmet sich einem Aspekt, der in der internationalen historischen Diskussion seit geraumer Zeit einen entscheidenden Schwerpunkt darstellt: der Anwendung computerunterstützter Methoden.

Es braucht nicht gesondert betont zu werden, daß in der Alltagsgeschichte, und nicht allein in jener des Mittelalters, diesbezüglich ein bedeutender Nachholbedarf zu konstatieren ist. Doch auch in unserem Fachbereich mehren sich Einsicht und Zeichen, daß EDV mehr bieten kann als Textverarbeitungsprogramme zur schnellen und billigen Produktion von Publikationen. Ein entscheidender Vorstoß in diese Richtung ergibt sich beileibe nicht nur für den deutschsprachigen Raum auf Grund der Initiative des Max-Planck-Instituts für Geschichte in Göttingen und dessen EDV-Spezialisten Manfred Thaller. Wir sind deshalb besonders froh, gerade ihn für einen sehr grundsätzlichen Beitrag in Medium Aevum Quotidianum–Newsletter gewonnen zu haben. Auf Grund der internationalen Relevanz des Aufsatzes haben wir uns entschlossen, diesen in englischer Sprache wiederzugeben.

Der zweite umfassendere Beitrag des Heftes, verfaßt von Ingo Kropač und Peter Becker (beide Graz), widmet sich einem Anwendungsbeispiel, dessen inhaltlicher Schwerpunkt zwar auf dem Gebiet der mittelalterlichen Prosopographie zu suchen ist, welches jedoch nicht nur wegen der angewandten Methoden auch für die Alltagsgeschichte des Mittelalters von Wichtigkeit erscheint.

Beide Beiträge sollen den Auftakt darstellen für eine in unserer Reihe fortzusetzende Diskussion zur Anwendung computerunterstützter Methoden.

Der bereits seit längerem angekündigte Band mit den Ergebnissen der von Medium Aevum Quotidianum im Oktober 1985 mitveranstalteten Tagung "Migration und horizontale Mobilität vom Mittelalter bis zum Ende des Ancien Régime" befindet sich im Druck (Campus-Verlag, Frankfurt am Main – New York) und wird noch heuer (für unsere Mitglieder als MAQ–Newsletter 11/12) zur Auslieferung gelangen.

Ein besonderer Hinweis muß auf den bereits im vorigen Heft erwähnten Kongreß gegeben werden, den Medium Aevum Quotidianum in Zusammenarbeit mit dem Institut für mittelalterliche Realienkunde der Österreichischen Akademie der Wissenschaften vom 27. bis 30. September 1988 in Krems veranstalten wird. Er befaßt sich mit dem Thema "Mensch und Objekt im Spätmittelalter und in der frühen Neuzeit. Leben – Alltag – Kultur" und läßt im Rahmen der Diskussion von international anerkanntesten Fachleuten vor allem im theoretischen und methodischen Bereich neue Ansätze und Wege erwarten. Voreinladungen an unsere Mitglieder wurden bereits vor einiger Zeit versandt.

Wie gewohnt wird sich einer der im Jahre 1988 erscheinenden Newsletter den Zusammenfassungen der am genannten Kongreß gebotenen Referate widmen. Darüber hinaus wird, wie bereits zuvor angekündigt, Teil II der Auswahl-Bibliographie zur materiellen

Kultur des Mittelalters erscheinen. Im Planungsstadium befinden sich einige Hefte, die sich mit der Erforschung von Alltag und materieller Kultur des Mittelalters in einzelnen Ländern auseinandersetzen werden und einen Schwerpunkt insbesondere auf Forschungs- und Literaturberichte legen sollen. Das erste Heft dieser Reihe wird vermutlich Dänemark gewidmet sein.

Gerhard Jaritz, Schriftleiter