

# Digital Manuscripts: Editions v. Archives

## SESSION

*ORGANIZER: Manfred Thaller*

---

*Max-Planck-Institut für Geschichte, Hermann-Föge-Weg 11, D 37073 Göttingen*

### AUTHORS:

Manfred Thaller  
Dino Buzzetti  
Stefan Aumann

KEYWORDS: digital archives, manuscript processing

AFFILIATION: Max-Planck-Institut für Geschichte, Göttingen

E-MAIL: mthalle@gwdg.de  
FAX NUMBER: +49 - 551 - 495670  
PHONE NUMBER: +49 - 551 - 495664

Manuscripts have in recent years become increasingly frequently the object of presentation on the screen. This session, for the contributions to which individual abstracts are appended, tries to look systematically at issues dealing with the problems of such digital manuscripts. It is assumed that such systems fall into one of three classes:

### “Digital Facsimiles”

A digital facsimile will typically consist of an individual source, which is scanned at a sufficiently high resolution to allow at least palaeographic work, but which consists of a few hundred or thousand pages only. The purpose of such an edition is to make one witness which resides in a specific location available.

Components of such a digital facsimile should be at the very least:

- A complete transcription, accessible via a fulltext system.
- A prosopographical catalogue of all persons mentioned, in the form of a database.
- A topographical catalogue of all locations mentioned, in the form of a database.

All these components are administered by retrieval systems, which allow a user to select those portions of the manuscript to which these elements of description or transcription pertain. This means, that the user is able to call up, for example, the

whole manuscript page(s) which contain a reference to a specific person. High end versions of digital facsimiles will as a result of such a query show the location in the manuscript, where the selected person or phrase is contained.

Beyond the basic tools given above, more advanced versions of digital facsimiles will typically include:

- Tools for producing maps for the topographical information contained in the source.
- Computer accessible knowledge representations describing, as far as applicable,
- calendar systems as used in the source,
- coinage systems as used in the source,
- a terminology database relating, for example, legal terminology used in the source.
- Graphical representations of the alphabets used by identifiable scribes.

### “Digital Editions”

Digital editions aim at presenting the same kind of corpus as digital facsimiles. Unlike these, however, they attempt to represent either all or at least a significant subset of all existing witnesses, representing in that case exactly the concept of a critical edition.

In addition to the tools provided by digital facsimiles, they will include mechanisms to:

- Represent the individual witnesses dynamically. Popularly speaking: when you look at a text, it is the text of the reconstructed original; if you press “F1”, you will see the text as occurring in witness “α”, if you press “F2” you will see the text as occurring in the witness “β” and so on.
- Link the transcription of the individual witnesses to their graphical representation. (So, if the user doubts a specific transcription of a given witness, (s)he can check the reading in the digital representation of that manuscript.)

While quite a few projects exist which have either produced early versions of digital facsimiles or are in the process of doing so, digital editions have so far not very frequently been realized, with the notable exception of the Canterbury Tales project, though they are being actively explored by a number

### “Digital Archives”

This form of presentation does not aim at a specific, relatively small “text”, but at the representation of archives as a whole. Sizes to be expected range between about 50,000 pages to some million.

To make these masses of information available, much more “shallow” descriptions will have to be

used. While, scanning operations which converted 6 – 8 million pages have successfully been performed, most existing attempts at the large scale conversion of archives have so far been hampered by the less than perfectly convincing tools to access the huge amounts of material. These difficulties are probably exaggerated by the fact that many historians and archivists have a somewhat archaic concept of databases and believe they are forced to create highly formalized schemes to administer manuscript material. The future will probably show the usefulness of a more direct translation of traditional archival tools into computer supported versions of these tools.

As the entering of descriptions / transcriptions of sources takes usually much more time than the digitization, a digital archive can, and should, successively go through various levels of accessibility. Such levels can, for example, be:

- 1) The digitized documents with nothing but their archival shelf marks.
- 2) The same documents with short abstracts in natural language.
- 3) The same documents with catalogues of prosopographical and / or topographical and / or formulaic information.

This “dynamic” character of collections of digitized manuscripts is actually one of the more fundamental differences, if we compare them to more traditional ways of making source material accessible.

The proposed session assumes, that *all* three kind of “products” of manuscript processing are actually being related to a group of common technologies. The intention of the session is to show the interrelationship between the problems encountered in the realization of the various types of product.

For this purpose the following course will be chosen:

- a) Starting from the experiences gained in a series of projects dealing with the administration of various types of digital source material, a contribution of Manfred Thaller describes the relationship between the conceptual problems involved in the creation of data bases holding such types of source material and the requirements for software tools to facilitate the creation of such software environments.
- b) Dino Buzzetti will then show, how on the one hand these general techniques can be used to solve a typical problem of the administration of digital editions, the processing of variants.
- c) To prove the generality of the proposed solutions, we confront this contribution with one by Stefan Aumann, who will discuss an example from the creation of a 50,000 page archival data base, to show how the general techniques

described can also be used to realize access mechanisms to large collections.

## Text as a Data Type

*Manfred Thaller*

---

*Max-Planck-Institut für Geschichte, Hermann-Föge-Weg 11, D 37073 Göttingen*

KEYWORDS: non-linear text, programming tools

AFFILIATION: Max-Planck-Institut für Geschichte, Göttingen

E-MAIL: mthalle@gwdg.de  
FAX NUMBER: +49 - 551 - 495670  
PHONE NUMBER: +49 - 551 - 495664

The administration of non linearly encoded texts on computer systems has traditionally been seen as a relatively high level problem in systems design. That is, relevant systems usually take a rather traditional approach at low level programming and add the nonlinearly functionality required for the presentation of nonlinear texts and/or linkages between digitally represented and transcribed texts by specifying appropriate functionality within the environment of the particular application system.

This creates problems, when a non-linear property transcends a specific component of an application. Assume, e.g., a text which is marked up according to two overlapping hierarchies. While these two hierarchies are represented in markup as equally important, “loading” the text into almost any target system usually means that a browser converts the text into a data object, which represents exactly one hierarchy and simply ignores the other. That is very convenient from the point of view of the target system, as, when it is being realized, the whole question of co-existing hierarchies can be ignored. The point of designing a markup scheme which allows for overlapping hierarchies, only to loose this property when the text is actually browsed into an application is not immediately understandable, however.

As a solution we propose to implement a data type “extended string”, which replaces the traditional concept of a “string” in programming application systems. This means, that any application program accepts “external information”, which is browsed into the internal extended string representation, processed in that form and re-converted into some kind of external representation before being displayed on an appropriate medium. This is far from new, of course: a good example might be a system

like X-Windows where this general logic is used to allow an applications programmer to manipulate strings by completely traditional tools, while the internal string representation takes care of all aspects of processing necessary to handle font properties in display.

We assume, however, that this logic can be carried considerably further. Let us assume, that a given text is marked up by two overlapping hierarchies, one representing the division of the text into reference units, like pages and the other some semantic division, like the names of fields of a data base system, into which a specific substring belongs. Even if the text is marked up in a way which preserves both types of division, once it is browsed and loaded into the underlying database structure, we will normally not have the possibility anymore to access the reference units. More explicitly: if such a text is browsed into a data base system which has been realized in C, the function call

```
strcmp(name1,name2)
```

will yield the same value, irrespective whether name1 and name2 are contained on the same page or not.

To change this, we propose the implementation of a data type "extended string", which has a comparison function

```
estrcmp(environment,name1,name2)
```

which by default should act just as strcmp() above. If within an application program, however, it should be preceded by a call to an environment changing function

```
estrsetsensitive(environment,  
PageSensitivity,On)
```

any following call to

```
estrcmp(name1,name2)
```

should result in different return values, reflecting whether name1 and name2 are on the same page or on different ones.

Taking examples from a series of ongoing projects who use experimental software based on the concept of a data type "extended string" as introduced above, the proposed paper discusses first some practical problems of its realization and the interrelationship of such an implementation with existing programming tools, taking as an example the embedding of the data type into a X-compatible widget.

It should be emphasized again, that this is just an introductory example: the number of string properties handled in that form is rather large and goes considerably beyond the scope of overlapping hierarchies. A complete description of the concept of an extended string can be found in M.Thaller, "The Processing of Manuscripts", in: M.Thaller (Ed.) *Images and Manuscripts in Historical Computing* (= Halbgraue Reihe zur Historischen Fachinformatik, vol. A 14), St Katharinen, 1992, 97-121. All the properties in question can be divided

into three groups: (a) Those which are necessary to implement nonlinearity (from which our initial example has been taken), (b) those which are necessary to connect transcribed parts of a text to bitmaps of the image it describes or the manuscript it transcribes and (c) those which deal with "graphic" properties of portions of a text.

In all three cases the questions raised relate to two different fields: on the one hand they are connected to the practical dimension of programming. This aspect is supposed to be covered by the example quoted above. On the other hand, however, the actual policies to be implemented by such a purely technical solution, reflect heavily the conceptual decisions about what a specific property of a text actually means within the context of a given discipline.

This shall be described with regard to the question of how much information is actually related to the third of the three problem areas given above, the graphic properties of a text within historical research. Speaking on the most general level, we consider a text to be "historical", when it describes a situation, where we do neither know for sure, what the situation has been "in reality", nor according to which rules it has been converted into a written report about reality. On an intuitive level this is exemplified by cases, where two people with the same graphic representation of their names are mentioned in a set of documents, which possibly could be two cases of the same "real" individual being caught acting, which, however could also be homographic symbols for two completely different biological entities. At a more sublime level, a change in the color of the ink a given person uses in an official correspondence of the 19th century could be an indication of the original supply of ink having dried up; or of a considerable rise of the author within the bureaucratic ranks. Let us just emphasize for non-historians, that the second example is all but artificial: indeed the different colors of comments to drafts for diplomatic documents are in the 19th century quite often the only identifying mark of which diplomatic agent added which opinion.

What these introductory examples should demonstrate, is, that the text – the computer interpretable representation of a written document – forms in historical research an intermediate layer between two other layers of information. On the one extreme we have abstract factual knowledge about the various entities described in a text, which allows the interpretation of it; on the other there are purely graphical characteristics of the written document, which may carry meaning, but need not do so.

That the second problem is a genuine markup problem is probably obvious: if we use a computer to prepare diplomatic drafts of the 19th century for printing, we obviously need a way to describe a

portion of the document as being “written with blue pencil”. Which, at the time of the first transcription is exactly what it says, a literal description of a graphic property, though during the process of research it may well acquire a more abstract connotation, like “author=M. Simpson”. This could of course be interpreted as such properties being eminently fitted to abstract rules for markup, because at the time of producing the markup we have not yet the faintest idea what the final representation in print, if any, of the specific graphic property is to be. The problem is however, that part of the research which is supposed to be supported – at least within an archival environment – is precisely dedicated to finding out, what the observable graphical properties mean. If a computer system shall therefore be able to support historical *research* as opposed to administering in a convenient way *results* of historical research, it has to have the capability of administering graphical properties as what they are, being able to switch to a more abstract interpretation in time, but always being able to fall back to what can actually be observed.

To bring it to a point: almost all the examples given in the discussions on standardization during the last few years dealt with how to tag a structure which is clearly understood and where the graphic representation is accidental. Historical work deals with structures in a text which we want to discover, where the graphics we see may be all the clues we ever might get.

Concludingly the paper shows how these considerations fit into the ones that resulted in the first example given, and can be turned into an organic extension of an implementation as the X-compatible “extended string widget” introduced above.

## Digital Editions: Variant Readings and Interpretations

*Dino Buzzetti*

*Dipartimento di Filosofia, via Zamboni 38, I-40126 Bologna, Italy*

KEYWORDS: text representation, variants, text interpretation

AFFILIATION: Department of Philosophy, University of Bologna

E-MAIL: g2tbov11@vm.cineca.it  
FAX NUMBER: +39 - 51 - 258355  
PHONE NUMBER: +39 - 51 - 258357

Textual mobility and fluidity is almost a norm in

various genres of medieval literary production. Alternative readings in different manuscripts do not usually confine themselves to mere accidents of textual reproduction. Sources of this kind suggest on reflection that the traditional goal of assessing the text in the most reliable way, that is through a critical edition following the classical rules of philology, could be neither feasible nor desirable. The very idea of making clear-cut choices based upon collation seems to be neither applicable, nor altogether sound. On the contrary, an appropriate editorial policy requires that also the so-called alternatives should be edited as text. A database representation of the entire textual tradition provides an obvious solution to this problem and in notable cases a database comprising the encoded diplomatic transcriptions of all the extant manuscripts has been actually constructed. A database representation keeps closer to the varied and diversified nature of medieval textuality and contributes to its critical analysis in a way that overcomes the insurmountable limitations of the printed form of textual representation. Many medieval texts are fluid and dynamic, but as reproduced in a printed book they become fixed and immutable: the form of representation is forced upon the form of what is to be represented. A database representation offers a viable alternative in cases where the printed model, the classical model of textual representation is not altogether appropriate.

The new form of textual representation provided by a database of transcriptions can be further improved by means of digitized images integrated into it. A digital image is logical data and is not to be thought of as a mere physical reproduction of a manuscript document. A digital image can be processed; it can be linked with transcriptions; it provides elements for interpretation and analysis. Therefore, a digital image is to be conceived as a direct representation of textual content and not as a substitute for an absent document. In this respect, it is on a par with a digital transcription, which needs not replicate a physical document, another form of textual representation, but can be taken in itself as a direct form of representation of textual information. But a digital transcription is a form of representation of a new kind, and a digital image just as well. Digitized images and transcriptions are to be thought of as processable data. And a processable representation of a text is a form of textual representation very different from a non-processable one. What makes two different forms of representation a representation of the same text is their invariance with respect to their information content. What makes two representations of the same text two different forms of representation is their difference with respect to processing and analysis.

The fundamental feature of a textual representation in digital form – what makes it essentially different from any other one – is therefore its liability to processing. And there seems to be more point in this observation than just stating a platitude. For, in this respect, the very problem of textual representation in a digital form becomes the following one: how can a digital representation of textual information be effectively structured and processed in view of a specified analytical purpose? Now, one of the basic purposes of textual criticism is the analysis of variant readings, an analysis of the different forms of representation of the same text. Are they to be conceived as spurious corruptions, or are they to be conceived as genuine texts? The problem of analysing textual content cannot be separated from the problem of analysing its several representations. But they are different ones, and how are they to be connected? Is textual content by itself stable and immutable, so as not to admit a mobile form of representation, or is it on the contrary the steady and immovable form of a given form of textual representation, that forces textual content to be fixed once and for all? Medieval forms of textuality are often fluid and dynamic; in this case, the printed form of representation freezes textual mobility, and the form of representation should not be mistaken for the form of what is to be represented. On the other hand, and for the same reason, the dynamic form of a database representation – a form of representation that affords a more faithful reproduction of the varied and diversified expressions of textual fluidity – should not be mistaken for the accomplished form of its edition. A database in itself is by no means an edition, and there is a point indeed in rejecting the idea of an “archive edition”, a sort of all comprising inventory of any available piece of textual evidence. A database cannot be conceived as an edition as long as it is thought of as a sheer duplicate of its source material.

A database had better be thought of as a structured logical representation of textual sources, and here can be found an answer to our problem. A database is a form of representation, and a representation of whichever sort neither is, nor can be, just a replicate of its original. The problem is indeed to put its logical features to a good use. But how, exactly, can that be done for the sake of producing an edition? The most plausible answer appears to be to organize a database as an apparatus. For that seems to be precisely what makes an edition – not just an archive – out of anything. As it has been said, representing in database form, with commentary, a textual tradition is already translating encoded textual features into structures. And that could possibly be done just for the sake of documenting one or another reconstruction of the text, which is precisely the purpose an apparatus is created for.

We can then describe the computational problem we have to face in the following way: (1) what data structures can we obtain out of digital representations, either in encoded character form or in bitmapped form, of textual materials; and (2) what kind of processing procedures do these data structures afford us? Moreover, (3) do these data structures and these processing procedures meet our needs, as far as representation and analysis of textual material is concerned? Finally, it should be firmly kept in mind that it is from this last requirement that we have to start, for it is from our research goals, and not vice versa, that we have to proceed. And it is precisely in this respect that the computational model now newly implemented into “kleio” seems to provide an answer.

The treatment of variant readings is a typical problem of overlapping hierarchies. And this problem is radically tackled within “kleio” at the basic level of system design. The computational model, data structures and processing functions, is then able to conform to the conceptual procedures imposed by the needs of text critical research. The several layers or witnesses of a text can be easily mapped into distinct sequential representations, severally implying different and mutually overlapping hierarchical representations. But we need to use a data type “extended string” in order to make all different sequential realizations of textual representation jointly compatible in a unique and consistent nonlinear representation in database form. A database representation can thus act as a consistent and unifying model of all different sequential representations of a text, a congruent structure onto which they can all be mapped simultaneously and consistently, and from which they can all be separately derived and individually displayed. By means of the “extended string” data type it is possible to reduce to a consistent unity a multiplicity of different and possibly overlapping hierarchical representations, thus meeting the needs of the editor of a text handed down by a variety of different sequential representations; or it is possible, vice versa, to derive from a single sequential representation a multiplicity of different structural representations, thus serving the purposes of a scholar approaching the text from more than a single analytical perspective.

Both the editorial and the interpretative practices seem to need a computational model of the same kind, allowing the reduction to unity, or conversely the derivation from unity, of a multiplicity of structurally different representations. But in the one case, the case of an edition, we have to start from a multiplicity of representations of the sequential structure of a particular document witnessing the text and reduce them to a unique structural representation of a non-sequential kind; whereas in the other case, the case of textual analysis and

interpretation, we have to derive from a single sequential representation of textual content, a multiplicity of structural representations of a non-sequential kind. The editor has to care about structural representations, both sequential and non-sequential, of the documents representing the text; the interpreter has to care about structural representations, both sequential and non-sequential, of the textual content represented by a document. It is structure that enables a document to represent a text, and it is structure that enables a text to be represented by a document. It is therefore the structural properties of the digital form of representation that we have to rely upon in order to apply appropriate conceptual procedures both to documents and to texts.

In this respect, the advantages of a digital form of representation over a printed one are absolutely clear. A digital representation can easily be structured both in a linear and in a non-linear form and can more aptly be employed for research purposes in text representation and analysis. A digital representation, either a transcription or an image, can improve research considerably if it is used as a structural form of representation. The reproduction of a document in all its physical properties can immediately be turned into a structural representation, because it is logical data by itself. The crucial problem is to organize digital data into data structures suitable to textual representation and in implementing processing functions suitable to textual analysis, a problem that can be solved only at the level of system design. The application of the "extended string" data type newly implemented into "kleio" to text critical problems has proved to be a substantial step towards reaching satisfactory solutions. And its application to problems of analysis and interpretation looks just as promising on the same grounds.

## Digital Archives

*Stefan Aumann*

---

*Max-Planck-Institut f. Geschichte, Hermann-Foerge-Weg 11, D 37073 Goettingen*

KEYWORDS: digital archives, user interfaces, data security

AFFILIATION: Max-Planck-Institut fuer Geschichte, Goettingen

E-MAIL: saumann@gwdg.de  
FAX NUMBER: +49 - 551 - 495670  
PHONE NUMBER: +49 - 551 - 495632

Recent discussions on the possibilities to store digital manuscript material have most oftenly focused on the possibility to produce high quality representations of a rather restricted amount of digitized source material. In the archival world, on the other hand, digital systems have frequently been designed with the understanding that the digital storage of bulk material is primarily a replacement of the classical microfilming operations of archives.

Using a German project, which intends to create a pilot "edition" of a serial source of ca. 50,000 pages, this paper discusses how far archival systems can provide a starting ground for an incomparably more intensive access to bulk material than traditional techniques.

The presentation will start with a short review of the existing access methods for digital archives. It is well known that while the scanning campaign of a digitization project represents a serious organizational task, the provision of the various access tools which allow a user to access the digitized material, actually requires a considerably larger effort.

Let us recapitulate what the purpose of these access mechanisms is. The user of a digital facsimile or edition should have the possibility to select those pages he or she wants to look at by specifying characteristics of the text contained on the individual pages. The user of a digital archive should have the possibility to access by similar means all parts of the archival holdings which interest him or her in the form of high quality reproductions right at the desk in the user's room in the archive.

Traditionally this is done with the help of either full text retrieval systems or structured databases which contain descriptions of the material, which makes their preparation rather time consuming. Three forms of access can be differentiated between.

### *1) Access by Browsing*

The user encounters the manuscript(s) as a – potentially structured – collection of pages. (S)he pages through the material in the order imposed by the structure holding the documents.

This is the only traditional access tool which can be realized speedily. More popularly speaking: you go to the traditional catalogue of the archive, look up the shelf mark, enter it into the computer (or select it there from a list) and get the first page of the relevant document onto your screen.

### *2) Access by Query*

The user specifies a query in the query language of an underlying database system. This query addresses formal descriptions – which can contain partial or complete transcriptions – of the docu-

ment. As a result the user is presented with an ordered collection of qualifying pages.

Less technically: you save the excursion to the catalogue, which is itself administered by the computer as a database in which you can employ traditional database tools. The problem with such an approach, as mentioned before, is that it is usually a very complex operation to convert a traditionally very flexible and highly irregular archival catalogue into a rigidly structured database.

### *3) Access by content*

Partial or complete transcriptions are loaded into a fulltext system, presenting the complete vocabulary of some holdings as an "active list". By dynamically specifying the formulae needed, the selection is narrowed down to a manageable number of documents, which are then displayed.

Because of the heterogenous nature of traditional archival tools, such a conversion is usually easier to accomplish than the creation of a rigidly structured database. This idea to create a computer based access tool directly out of an existing one leads us one step further, to:

### *4) Access by Digitized Versions of Traditional Tools*

An existing catalogue or findbook is digitized itself. The digital version of this tool can be accessed by any of the access methods described so far. "Activating" an entry of the digitized tool initializes the display of the page(s) described by it.

Less formally: you search within a graphic reproduction of the old catalogue on the screen and click on a specific entry within it to see the first page of the file described by that entry.

This notion of using a visual object as an access tool for other visible objects leads directly to:

### *5) Access by a Graphic Overview*

The organizational scheme representing the order of the collection – for example a map of a community or territory – is presented as user interface. By activating a "house" or "location" on the map, the related documents are displayed.

More intuitively: you click on a map to start browsing through all the documents related to the village clicked on. While this is more intuitive, it can be shown however, that for actual access to information within real-size historical territories, the popular "clickable" map of toy applications may need some rethinking to reach an acceptable information density on the screen.

### *6) Access by Fragment*

Significant sections of the manuscript – for example illuminated initials or miniatures – are administered as a primary database. By activating such a fragment the part of the complete manuscript

from which it is taken is displayed.

Few experiences with this kind of approach exist yet; it remains to be seen whether such a tool which has been used experimentally within the realm of digital facsimiles can successfully be extended to large scale digital archives.

After having shown examples of the basic access mechanism, we go on to demonstrate, that the actual software functionality required to implement these techniques is very closely related to the functionality which has been implied by Dino Buzzetti's discussion of variant readings.

By this we assume to have demonstrated, that the various possibilities to use digitized manuscript material are closely related to each other: which supports the thesis, that the appropriate response of archival institutions to the new technologies should primarily be in the creation of an institutional framework, which is sufficiently flexible to allow one and the same institution to act as a logistical host for a few groups of manuscripts with very intensive editorial information assigned to them, while acting at the same time as supplier of very shallowly described mass documents.

This may seem doubtful for one reason: documents, into which extremely intensive editorial preparation has been invested create different problems of copyright and protection against illegitimate distribution than mass documents with few, if any, explanatory information attached to each individual page.

We close our considerations on digital archives therefore with a discussion of the protection mechanisms employed within the organizational and software environment from which the examples of this paper are drawn. Data security in the case of archives arises broadly from three reasons.

a) The institutions from which the source material originates have been awakened recently to the problems of copyright with regard to digitized source material.

Museums are afraid that they will be robbed of large revenues if cheaper pictorial reproductions of their holdings, and particularly reproductions, which can easily be copied, get around. This is not quite so obvious in the archival case, but certainly represents a reason for much concern for an author of a digital facsimile or edition.

b) While nobody in a small city archive really believes that they will loose huge sums because their early 16th century account books can easily be copied, there is a widespread fear in the archival world that the systematic digitization of source material will threaten the position of the archives in two ways. On the one hand there is a widespread feeling that these technologies will let the archives lose control over their material. There is probably no technical answer to that: it is part of the impli-

cations these technologies have for the organisation of the research process. A more immediate fear, particularly in smaller archival institutions, is related to the fact, however, that many archives get funded among other reasons because the local authorities get convinced of the importance of an institution which has so and so many users a year. This effect, it is feared, will get lost when large portions of the archival holdings are accessible from the outside. c) A third problem arises with sensitive material, as, for example, in the case of an attempt to convert the holdings of the archive at the former concentration camp in Auschwitz into digital form. While the manipulation of high quality images is not quite as easy as that of low quality reproductions on which it is usually demonstrated, in the case quoted the danger of falsifications produced by some right-wing lunatics to prove the non-existence of the holocaust is quite real.

Within the various projects implicitly discussed here, we have not yet found any definite solutions to these problems. However in general the following procedures will probably be implemented. To protect the rights of the institution generating the material, it will be distributed in an internal format, which can only be accessed with a specific copy of the program issued with it; which should solve the problems described under a) and b).

In that area we assume that any protection scheme can only protect as long as no serious criminal attempt is made to break it. (If you want to produce a non-copyrighted version of a fairly traditional publication, you can do so just as well.) In the last case, however, where historical integrity is in question, and the potential offenders have a clear criminal potential, this is deemed insufficient. In principle it will always be possible to display visual material on a computer and dump a copy of the screen into a file, where it can then be processed further. While it requires quite some effort to recreate out of such dumps the original quality, it could in principle be done. The distribution of the material is not the problem in a case like Auschwitz: the more people see the authentic sources about the holocaust, the better. It has to be possible, however, to prove easily that a specific visually reproduced document has not been tampered with. For such purposes digital reproductions of images or manuscripts can contain embedded "watermarks" or "seals" which are as difficult to break as the identification codes for credit cards and similar devices.

The presentation concludes by an attempt to show briefly, how these mechanisms for the protection of manuscript security fit into the overall logic of manuscript processing, which is supposed to be the covering theme of this session.