...atermarked Papers

...Bustarret

...No. 2 (May, 2002), pp. 143–169

Figure 3. Betaradiography vs processed image. Profil #55: Crowned griffon, manuscript of Bossuet (RothBoss VC, #283). (Photo: Bibliothèque Nationale de France, digitized by Laboratory of Optics, University of Besançon.)

Finally, the image resulting from the extraction of the outline is superimposed on the restored image (i.e. rendered homogeneous and free of weave pattern), so as to intensify the watermark outline.

The processing that is carried out is aimed at restoring an image as close as possible to the original. Unlike manual restoration, there is no risk of sub-interpretation or omissions while copying; and although reading a betaradiograph may present some difficulties in detail, this technique remains more accurate than manual ones (De La Chapelle and Le Prat, 1996). The processed image is faithful to the original imprint, and with enhanced legibility.

2.3. COLORS

The image obtained by betaradiography is in black and white. Color can easily be used to display images in various hues to bring out details better and help the user pick out variations. However, the undiscerning use of color can have the opposite effect, and a profusion of colors can be more confusing than helpful. The choice thus has to be made while bearing in mind the need to provide an interface that is easy to read, and offering a range of possibilities with real advantages compared with gray levels.

For the sake of verisimilitude, we have selected as a reference an image in which the dominant color is cream, which is close to the natural color of commonly used paper. The standard screen display of the *Profil* database thus includes both the descriptive watermark data and two boxes in which the reference images are placed in icon form. This coloring also has the advantage of enabling ready identification among all the other information (cf. Figure 4).

Papiers filigranés   N° : 304

1. FILIGRANE   Type Triple   Classe B/Y

Marque portrait de Napoléon , de profil, en médaille dans un triple cercle portant la devise "NAPOLEON EMPEREUR ET ROI" et un motif décoratif de 9 cercles imbriqués ornant le cadre

Détail M maj. rom. ombrées

ContreMarque aigle impériale empiétant la foudre // signe supplémentaire: "I* Johannot"

Détail CM cursives anglaises penchées

	Initiales	Position M	Position CM
Papetier	JOHANNOT, I*.		
Provenance		**Date Fabrication**	
Réf. Filigrane			

Rem. Filigrane M: diamètre du triple cercle : 9,6/9,9.

2. PAPIER   Dimensions : Fragment   Feuille 236 x 380 mm

Type	Vélin	**Qualité**	**Epaisseur**
Technique	Main		**Tranchefil**
Couleur	Ecru		**Ecart LC**
Emplac. Filigrane	Centré	**M** Droite	**CM** Gauche
Rem. Emplacement			

Réf. Papier

Rem. Papier signe supplémentaire en bordure inférieure au milieu de la feuille

3. DOCUMENT REPRODUIT

Cote BN	NAF 13969	
Folio	4, 5	**Pliage** in-2°
Nom Auteur	STENDHAL	
Prénom Auteur	Henri Beyle dit	
Titre Oeuvre	*Correspondance*	
Identification Texte	Lettre à Pauline Périer	
Support		
Nature Document	Lettre autographe	**Statut génétique**
Date de rédaction	14.08.[1815]	**Indice Date** Partiellement autographe

Figure 4. Standard *Profil* screen display.

Among the other possible colors, we have selected two as giving good image readability. We have moreover chosen colors that restore two complementary images: positive (dark watermark on a light background) and negative (light watermark on a dark background). The former is in gradations of green, the latter in gradations of brown. We limited the number of extra colors to two to reduce the amount of data, of course, but also and above all for ergonomic reasons; the goal as we stated above is to help researchers in identifying the paper support and discerning as well as possible the paper and watermark characteristics. Various trials on combinations of colors have shown that increasing the number of colors brings no appreciable advantage, hence the limit of two colors in addition to that of the reference image.

To summarize this section, we have seen how the iconographic documents used in the Profil database are created using betaradiography (original document in black and white) to obtain an image after processing. In what follows we shall present the functions used.

3. Exploring the Image Database

3.1. NAVIGATION

The basic principle of navigation is of the hypertext type: clicking on a sensitive zone on the screen triggers an action such as change of viewpoint, zoom, and running other functions in the application. As we saw above, the standard screen has two boxes in which reference images are placed. This means that the user sees a miniaturized image in the form of an icon (cf. Figure 4). Two icons are provided on the screen because most modern papers have so-called 'double' watermarks, and this type of watermark is digitized in two distinct images (mark and countermark).[4] Clicking on one of these allows zooming up to other scales.[5]

In general, navigation is between an *information zone* made up of one set of files on watermarked papers, paper makers, lexicon and occurrences, and an *image zone* composed of references images and colored images. The navigation interface clearly shows that various images can be retrieved through icons which are also colored (cf. Figures 5 and 6).

Thus Profil navigation enables images to be explored through different views of various sizes and/or colors. Whatever type of display is being used, it is at all times possible to quit the exploration of the image zone and return to the standard screen.

Apart from the navigation feature, there are other functions provided which, as we shall see in the next section, are image-related; these include the possibility of searching for the set of images with a given motif, and displaying arrays of examples illustrating the main variants of the terms featured in the iconographic lexicon.
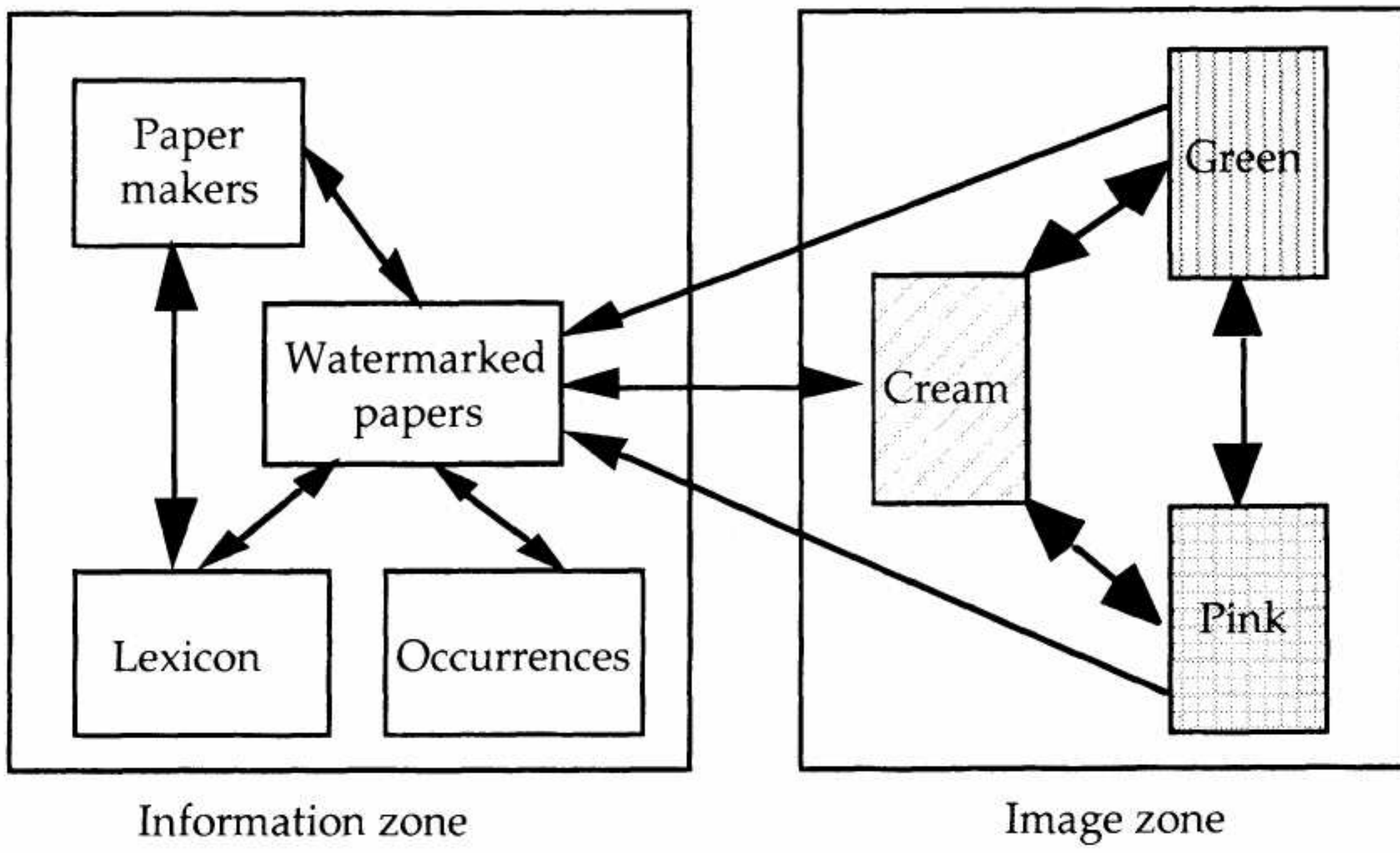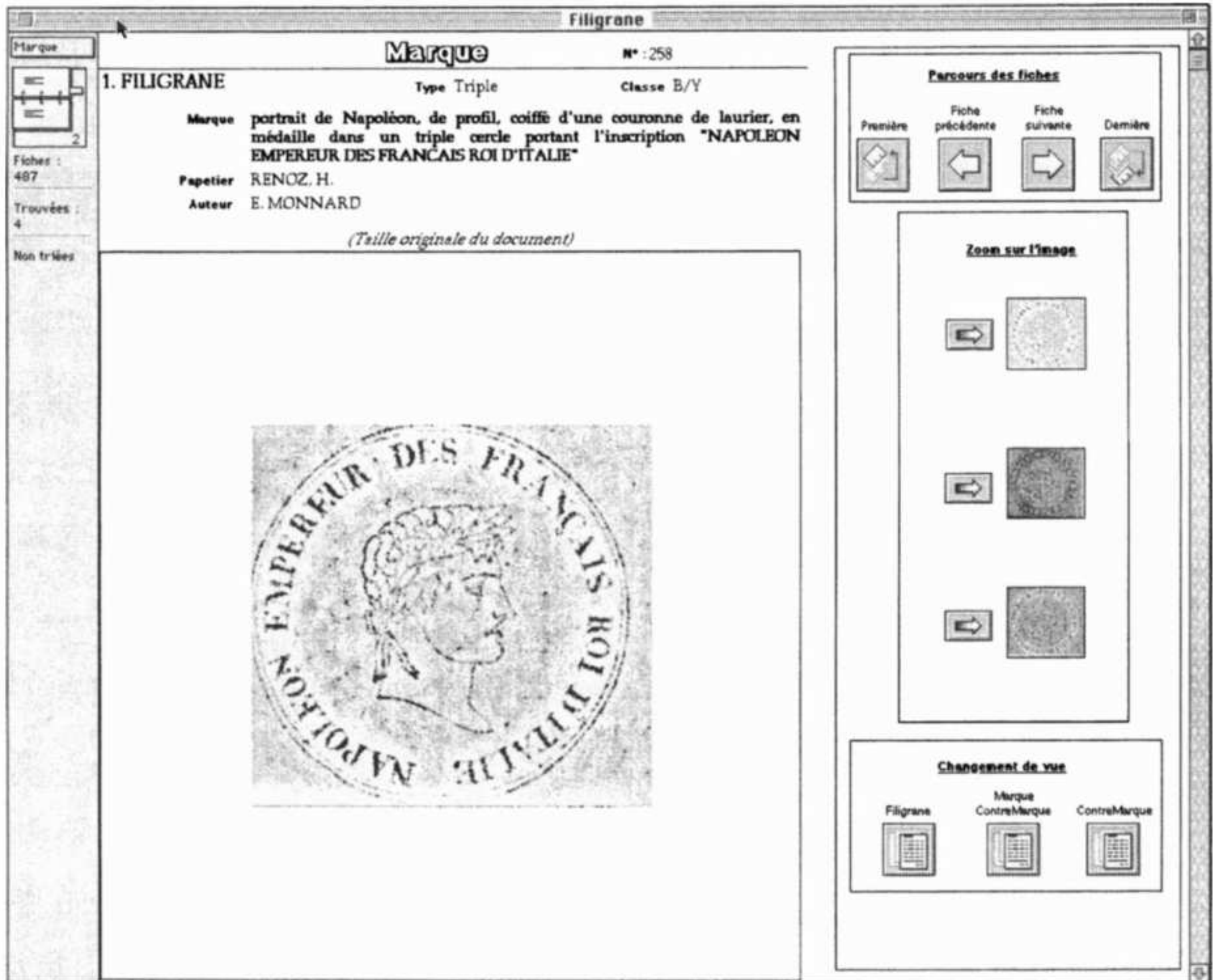
Figure 5. Navigation in *Profil*.

Figure 6. Navigation interface in the image zone.

3.2. THE IMAGE-RELATED FUNCTIONS

3.2.1. Search/Display

Among all the possibilities at the disposal of the user, the function 'Search/Display' (cf. Figure 7) is designed to search for a set of forms and display the associated marks and countermarks.

The result of the search is displayed as a list of images; this makes it easier to compare watermarks with common features. It is particularly useful for studying, for example, the various aspects and shapes taken by the '*griffon*' motif through time and in different regions, or to see which are the countermarks associated with the '*Hollandia*' (cf. Figure 8).

3.2.2. Lexicon

Another way of exploring the image base is to use the lexicon which actually includes a *technical lexicon* and an *iconographic lexicon*. The former has all the specific terms related to techniques of paper manufacture including, for instance, the definitions of '*chain-line*' and '*laid line*'. The latter is more related to iconographic motifs and graphic features of watermarks and their interpretation (cf. Appendix 2).

To this end a window can be selected in the 'Lexicon' function, displaying the set of listed words (cf. Figure 9). Selecting a word sends the user to its definition and, if it is a word in the iconographic lexicon, to a list of other possibilities for displaying images. One image being worth a thousand words, it is considerably more useful to actually see the watermarks corresponding to a '*Britannia*' than merely to give a description, however detailed.

Among the new types of display possible from the lexicon, the function 'Examples' displays an array of four images selected as prototypes by an expert which allows either to illustrate the different variants of a motif or to show common features (cf. Figure 10).

The user may, however, be curious to see not just a few examples, but all of the images corresponding to a given motif. To enable this, the 'Display' function gives a dynamic and exhaustive display of all the marks and countermarks relating to an entry in the lexicon. This function can also be seen as a tool for the expert enabling him to see at once all the images that correspond to a word, and hence to be able to compare them and select those he wishes to use as examples. As for the 'Search/Display' function mentioned above, all updates to the database are immediately effective since both functions rely on interrogating the database. The expert can thus be sure that he is working with the current version of the database and therefore with the whole of the available documents.

Particular care was taken in designing the navigation feature of the image base, since the image is central to the present project.
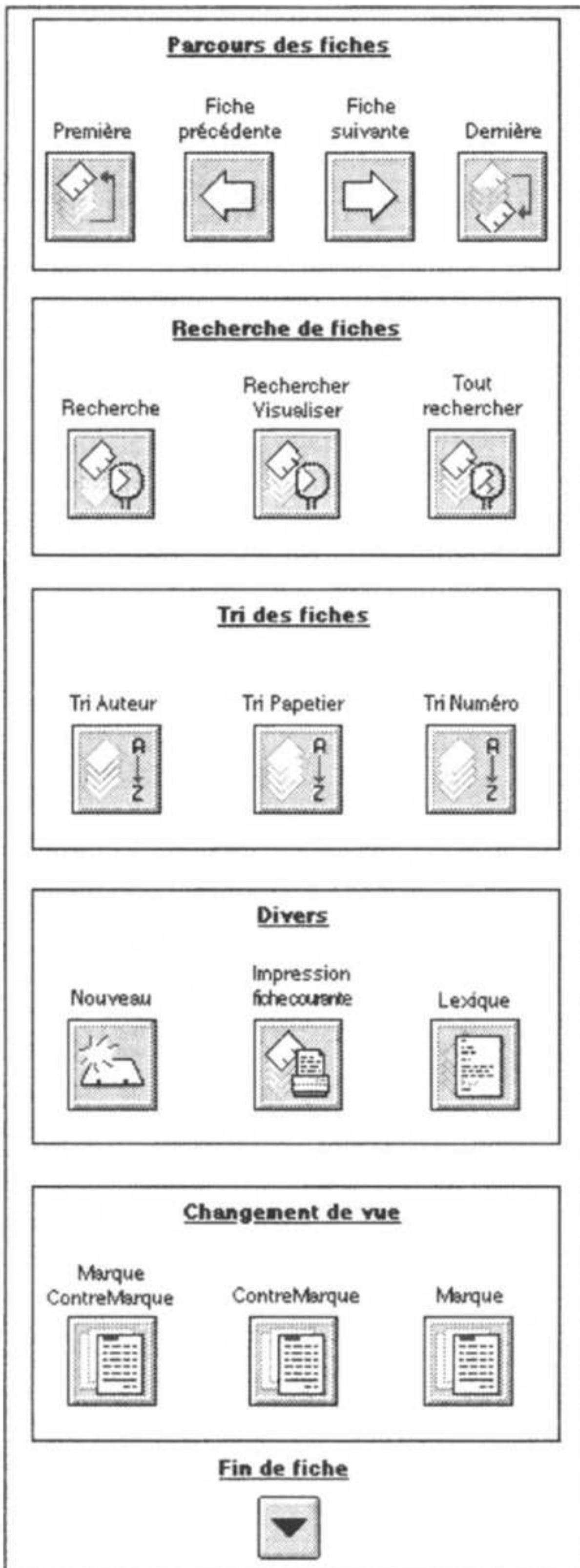
Figure 7. Function panel.

4. Construction of the Database

The *Profil* database was initially developed at the ITEM with *Pantheon*,[6] and recently transferred to a multi-platform database application. During the transfer a number of the options that were adopted over the years, and which are detrimental to overall functional coherence, were reviewed. In this section, we discuss

Figure 8. Display of countermarks related to the *Hollandia* (extract).

the choice of software, the various problems related to the setting up of databases in the field of human sciences in general, and in this application in particular.

4.1. CHOICE OF SOFTWARE

ITEM is a laboratory for research into genetic literary criticism and codicology; it is equipped with computers, but at the present time there is a lack of computer specialists to develop new applications or to meet the needs of the various research teams. Because of this, the smallest change in the initial database required a computer specialist to be called in to modify the appropriate lines of code written in Pascal. It therefore became clear to us that the new version should be in such a form that the researchers themselves could handle the changes. This is why we chose a commercially available package. This has several advantages: it is highly likely that new versions will be developed incorporating technical advances, available for various operating systems such as PC, Macintosh and Unix.

Given the use to which the software was to be put, and the context of its use, the chosen system had to meet some extra requirements. In particular it had to enable images to be incorporated easily, while at the same time being very user-friendly both as regards the interface and the setting up. We also wanted it to have effective full-field sorting and search capabilities. We wanted software that came, not with a

Figure 9. Lexicon.

Figure 10. Illustrative examples of a word from the lexicon.

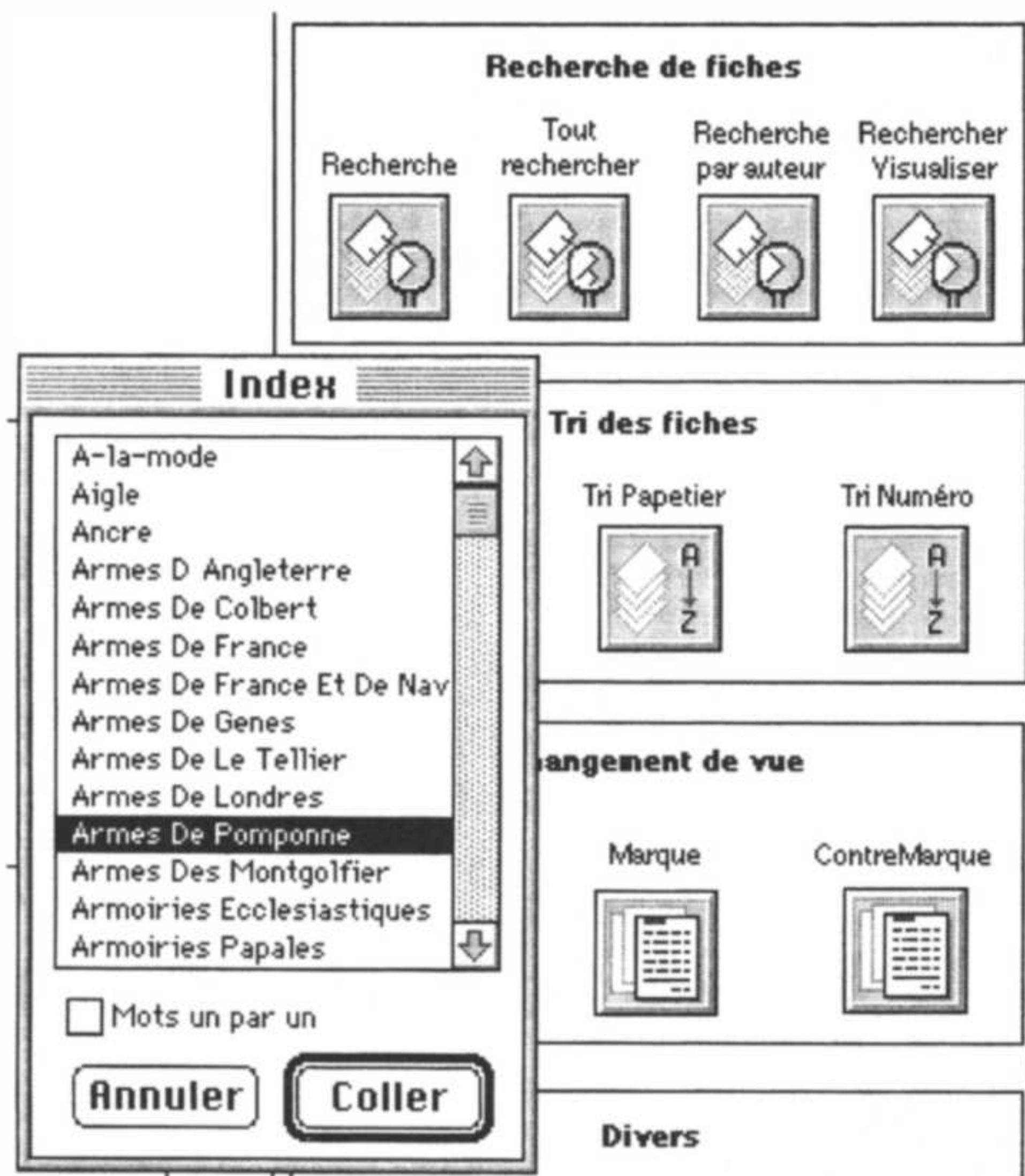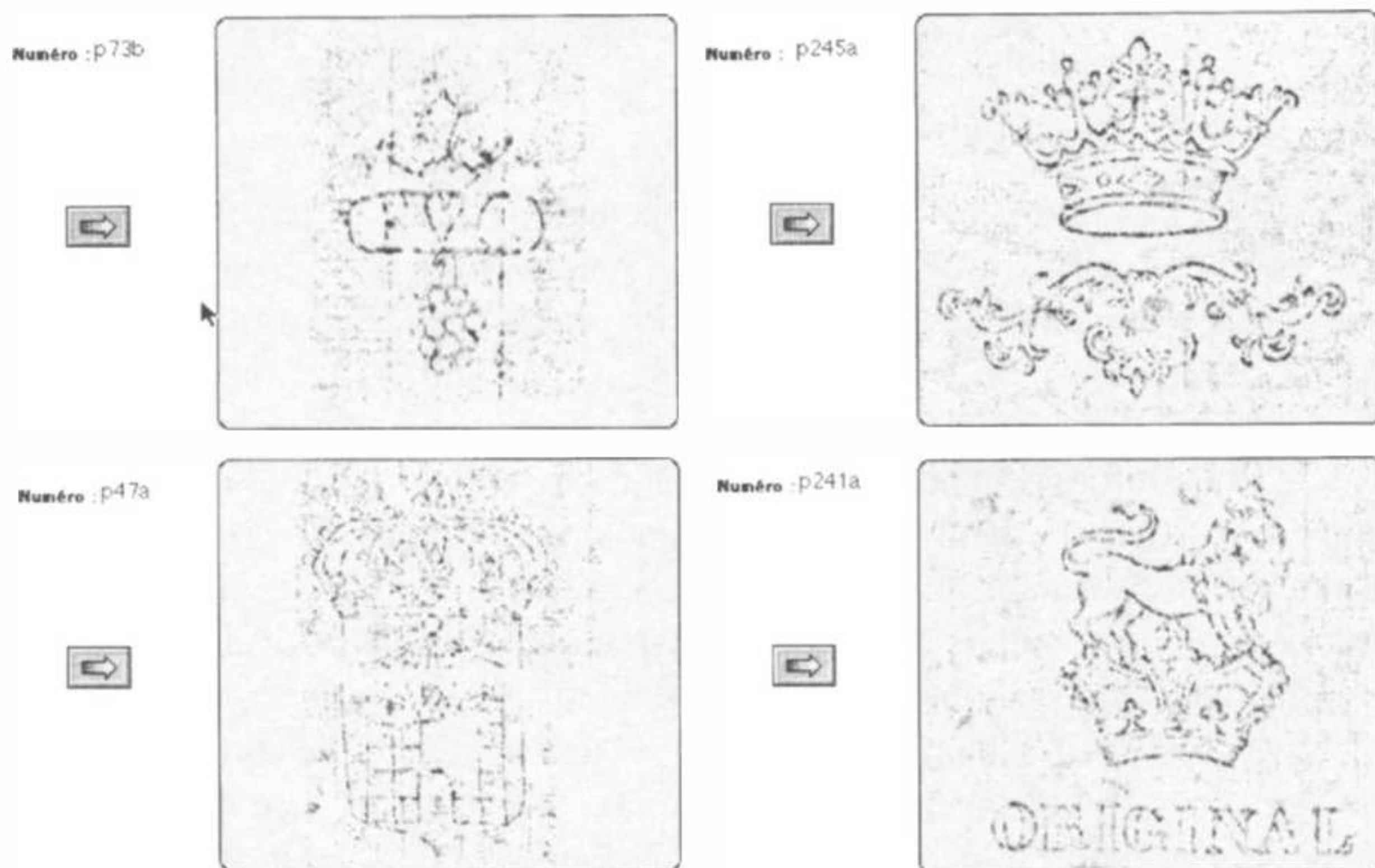complete programming language, but at least with a script language for programming the standard functions. These various considerations led us to the choice of FileMaker Pro® as our development software.[7] Since *Profil* was intended firstly as a reasonable scale tool for local resarch rather than an encyclopedic archive, the existing limitations of the software seemed acceptable, as long as it offered easy access. We only recently discovered that the Thomas L. Gravell Watermark Archive, available on the web, was using the same software.

4.2. DIFFICULTIES IN DOCUMENT TRANSFER

First of all we had to transfer data that did not correspond to any standard, that were not recognized by any commercially available application and that contained numerous and variable nested fields. As in all database design, we had to identify all the *fields* and *rubrics* necessary and to set down the possible values. Here we were faced with two different viewpoints: that of an expert in the field for whom it is vital not to lose information and whose reasoning is mainly descriptive, and that of the programmer who, bearing in mind subsequent automatic data processing, tries to eliminate as far as possible any special cases. A compromise was found between the requirements of both, in the form of a set of files structuring the information in such a way as to facilitate searches or sorting, and a set of complementary information containing all those details needed by experts which cannot be coded in any other way and cannot be included in a field of predefined values (cf. Figure 1).

Using a range of automatic procedures for re-arrangement and preliminary conversion we were able to retrieve all of the data in a form compatible with the newly-defined fields and with the new software. There were however several fields that were too composite and that had to be dealt with manually.

For the images, graphic converters had to be used to carry out the changes necessary to obtain a format compatible with current standards.

4.3. DIFFICULTIES RELATING TO THE HUMAN SCIENCES

Beyond this preliminary work of field definition and data transfer there remained a number of unsolved problems directly connected with the human sciences and with the difficulties of processing information which is inherently difficult to codify.

As we have already said, the expert needs to store a whole range of details describing the object under study, together with comments on the validity or incompleteness of the information, and even subjective impressions awaiting formal verification; such items are not always easy to model.

We give here a few examples of some difficulties arising.

irregular pulp (visible fibres); smoothed surface
smoothed aspect, irregular fibres
woven aspect; irregular pulp;
woven aspect when held up to light, smooth, almost shiny surface
woven aspect under glancing light
slightly woven aspect
slightly woven aspect when held up to light
thin, shiny, translucid paper
superb quality
thick, fairly smooth, good quality
high quality paper
superior quality paper
vellum weave visible and fairly coarse
coarse wove with regular shading
medium thickness, fairly stiff, not very smooth
thin paper, surface aspect of woven cloth, but smooth

Figure 11. Examples of paper support descriptions.

a) The 'lumber room'

For the expert in the field, the important thing is to have as precise as possible a description. However, to simplify the entering of data and allow specialist to access the database, he may have to gather various pieces of information under one. This was the case for example for information on the characteristics of the paper support where we found grouped data on the quality of the pulp, the roughness of the paper and its thickness. Moreover, as there is not yet in certain domains any standard vocabulary that can be applied without resorting to sophisticated measuring instruments, different observers may use different terms to refer to similar things or to describe similar visual or tactile characteristics.

The effect of such a lack of formalization, which is inevitable during the development stage of a methodology, makes searching for information difficult. Mindful of the need to be able to process various findings at a later stage, we have separated the contents of some rubrics in the initial database into several fields. Taking the previous example, we thus redefined more precisely what are the quality characteristics of a paper and drew up a set of appropriate qualifying terms (cf. rubrics 2.5, 2.6, 2.7 and Appendix 1).

This large effort to clarify terms next required us to review all the descriptions to ensure they were coherent with the new definitions.

b) Dates

Recognition and processing of dates is not without its problems either. How can dates in widely different formats be processed and interpreted? It may be the date of fabrication of the paper or the date which a document was written, and the date may

Profil: An Iconographic Database for Modern Watermarked Papers

BRIGITTE DE LA PASSARDIÈRE[1] and CLAIRE BUSTARRET[2]
[1]*Université Paris VI-LIP6, 4 place Jussieu, 75252 Paris, Cedex 05, France*
E-mail: Brigitte.de-la-Passardiere@lip6.fr
[2]*CNRS-ITEM, 61 rue Richelieu, 75002 Paris, France*
E-mail: bustarre@ens.fr

Abstract. The database Profil has been set up to offer readers studying modern literary manuscripts a reference tool to identify watermarked papers. In the study of writers' drafts as in artists' sketches, the different kinds of papers used provide valuable information on the genesis of a work of art and watermarks, when they exist, are the best visible hint allowing us to identify paper. A multimedia database, with digitized images more precise than usual traced design, seems to be appropriate to register, visualize, and compare modern watermarked papers. Besides its usefulness for specialists, such a database bearing on modern manuscripts should also be conceived in a didactic perspective, as it is oriented towards literary scholars who are not particularly familiar with the history of modern paper. In this paper we present the database Profil which includes a set of digitized images from a collection of betagraphies made by the reproduction service of the National French Library. Then we explain problems of database normalization when human sciences are involved.

Key words: betagraphy, codicology, database, image, watermarked paper

The *Profil* database has been designed at the Institute of Modern Texts and Manuscripts (ITEM) of the Centre National de la Recherche Scientifique in Paris, to offer researchers in modern literary manuscripts a reference tool to identify watermarked papers (Bustarret, 1994a). When studying a writer's drafts, valuable information on the origin of the work and its phases of composition can be obtained from the different kinds of paper used, as is the case for an artist's sketches (Grésillon, 1994; Bower, 1990). A watermark, when there is one, allows us to identify the paper easily and quickly (Briquet, 1907), yet it is not the only characteristic available to describe a sheet. The idea that a detailed physical analysis of the document can help in dating and localizing or identifying the author's hand is now widely accepted among scholars of ancient manuscripts, who have developed an auxiliary discipline for that purpose, namely 'Codicology' (Dain, 1949; Stevenson, 1962; C.N.R.S., 1974; Hay, 1976; Irigoin, 1980; Spector, 1987; Lemaire, 1989). However, the application of this method to modern and contemporary manuscripts – which consist of loose handwritten draft or typed sheets and notebooks of all kinds – is still at an experimental stage (Hay, 1993; Bustarret, 1994b).

be truncated or otherwise incomplete. It may come in such varied forms as 1840, 18**, **21, [18]26, 16?3, 1825–1828, 01.03.16–, 'around 08.1644', 2/3.07.1913 and '4 floréal an IV'.

We could opt for a textual format that simply accepts the date as it is printed, but in that case how could we obtain a list of watermarked papers in chronological order, determine the mean deviation between the date of fabrication and the date of use, extract a subset of papers used during the life time of Victor Hugo or to find all the watermarks in all works dated 1827? This problem could easily be solved using a programming language, but databases are not really designed for this, and our goal is that non-computer specialists should be able to adapt the database to their requirements. For this reason, so as not to lose any information vital to the specialist, we decided to keep the data in their original form (specifying the meanings of any special characters), but also to create equivalent dates (hidden) to allow value-based processing and hence to resolve some of the indetermination. If a manuscript by Victor Hugo only mentions an incomplete date, the specialist will copy the information as it appears on the autograph document, but this will in no way prevent the date in question from being easily identified. The same is true for dates of the revolutionary calendar, which simply need to be translated.

c) Intervals

Intervals pose quite a different problem. In the listed occurrences of a watermarked paper (cf. Figure 12) are found long lists of intervals corresponding to consecutive sequences of sheets (e.g.: ff. 1–8), for very often a writer will mainly have used the same paper to write on for an entire manuscript, with the intermittent use of different papers. Now, this source of information is very difficult to exploit when we wish to use the database to find out whether such and such a sheet has been described (e.g.: f° 7); such a presentation does not enable retrieval of a value which does not actually feature in the list. To remedy this we had to split up the various intervals and for each one locate the upper and lower bounds, in a way that is completely clear to the user. This technique gives the advantages of both a concise overview of listed occurrences of a watermarked paper and the possibility to search on any folio.

In comparable domains in human sciences, one of the goals of designing a database is to try to standardize items without losing information. In other words, the descriptive richness required by the human sciences has to be kept while making it possible to ask questions about any value. What could be more frustrating for a researcher than to be refused an information search on the very subject in which he is interested!

4.4. IMPLEMENTATION

The various tools and options available in FileMaker Pro® were used to implement the functions described in this article.

4. OCCURRENCES REPERTORIEES

Auteur	Cote	Folio	Date	Lieu
DIDEROT	NAF 15680	50-115, 148-197	[176?]	
	NAF 13729	124-125, 140-141, 142-143, 144-145,	1768	
	NAF 24930	90, 91-92	1772	
		123-126	1776	
		160-161	1767	
	NAF 24931	40-43	1768	Paris
		84-85	Paris	
		143-144	(1771)	
		188-189	(177?)	
	NAF 13724	1-84	1775	
	NAF 13725	2-4, 6-11, 13-15, 17-23, 25-28, 34-35,		
		39-40, 43-44, 49-50, 55, 61, 66, 70-72,		
		74-76, 78-79, 81, 83, 93, 95-96,		
		99-101, 104-112, 114-121		
	NAF 13727	1-32		

Figure 12. Examples of listed occurrences.
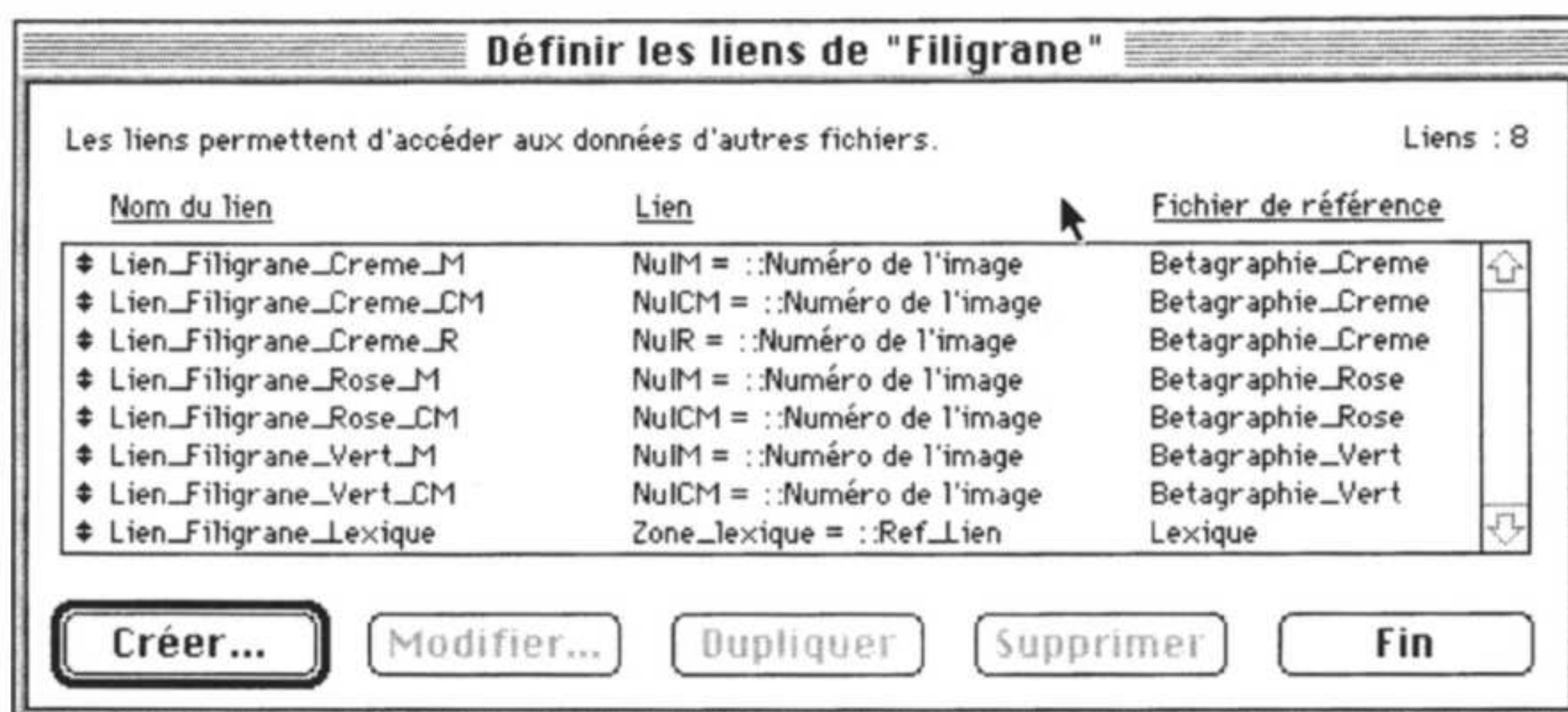
Figure 13. Linked files.

a) Linked files

Navigation in the information zone and the image zone (described in Section 3.1) works through links that specify the possible paths between the various components of the two zones (cf. Figure 13) or that set up reference between the various files that make up the application. There will thus be a link between the watermarked paper descriptions and the set of cream-colored reference images.

b) The various models

To differentiate each user's working space, a set of models or views specific to each user, and another specific to each expert/author has been defined. For the former we find the standard view (cf. Figure 4) that we have already described, and the interrogation interfaces which overall are identical to those described, but in which some information is refined to facilitate searches and to take into account

the solutions to problems handled in the previous section. For the latter, it allows recording of data and of specific information such as references to images of the mark and countermark of a watermark. Various printing options allow to be kept a record of the consultation and in particular of the result of research carried out on the database.

c) Scripts

The *Profil* application includes complex functions such as 'Search-Display' (cf. Section 3.2.1) which require programming. This is done through a script language[8] enabling tasks to be automated. If necessary a script can call up sub-scripts or even external scripts, i.e. scripts defined in another part of the application. The resulting programming is thus highly modular (Figure 14). A script can be called up by a button. All the complex functions of the application have thus been gathered into a function panel (cf. Figure 7).

d) Work carried out

At the present time, the *Profil* database covers modern watermarked papers from 1650 to 1950 with 500 types described and digitized, reproduced from the manuscripts of Bossuet, Diderot, Chateaubriand, Flaubert, Proust, Gide and various authors whose manuscripts are kept in the *Département des Manuscrits* of the *BNF*. It features among other things a *Lexicon* of 51 technical and iconographical terms and a database of *paper manufacturers* with 209 entries each with 32 fields (name, genealogy, associates, mill, location, dates, production, bibliography, etc.).

The *Profil* database is now operational. The interface has been approved by specialists both for its ergonomy and its contents, and it is considered to be one of the first of its type in Europe (DAP, 1995).

5. Conclusion

Although this database represents a real advance for specialists in the field, in comparison to watermark tracings and indexes, it is certainly still insufficiently large for it to be used as an instrument of research for public use. The next task will thus be to expand this still embryonic collection with many more descriptions of papers used in manuscripts from the eighteenth to the twentieth century and in doing so to systematically reveal a new corpus. Such a massive acquisition of data is closely dependent on means of reproduction (betaradiography or others), that are not at present available at the Bibliothèque Nationale de France, to obtain new prints. Once this is launched, it would be reasonable to envisage placing the resource on the Internet. Assuming some funding is provided for the acquisition of data and the translation into English of both data and navigation instructions, this presents no major difficulties given that the latest version (FileMaker Pro 4) of

```
Activer modèle : ["Profil"]
Mode Recherche [Pause]
Exécuter recherche []
Si ["ETAT (Fiches trouvées) = 0"]
    Tout rechercher
    Activer modèle ["Profil"]
    Activer fiche/requête/page [Première]
    Signal sonore
    Afficher message :
        ["Aucune fiche ne correspond à votre recherche."]
        Boutons : "Recommencer","Annuler",""
    Si ["ETAT (Message) = 1"]
        Exécuter script [Sous-scripts, "Rechercher_Visualiser"]
        Fin de script
    Sinon
        Fin de script
    Fin de si
Fin de si
Figer fenêtre
Activer fiche/requête/page [Première]
Boucle
    Si ["NuIM = """]                        pas d'image de Marque
        Ignorer fiche
    Fin de si
    Activer fiche/requête/page [Fin de script après dernière, Suivante]
Fin de boucle
    Si ["NuICM = """]                       pas d'image de ContreMarque
        Ignorer fiche
    Fin si
    Si ÉTAT (FichesTrouvées) = 0
    Signal sonore
    Afficher message [
        "Les fiches sélectionnées n'ont ni marque, ni contremarque."]
    Si ETAT (Message) = 1
        Exécuter script ["Sous-scipts, "Rechercher_Visualiser"]
        Fin de script
    Sinon
        Fin de script
    Fin de si
Finde si
Activer fiche/requête/page [Première]
Activer modèle ["Visual.M et CM"]
...
```

Figure 14. Extract from the script search/display.

the development software used to create the application offers the tools to launch a database on to the web.

Beyond its immediate use to identify manuscripts, it is easy to imagine more didactic applications using the image base allowing thematic exploration. From the point of view of computer use, the next direction of research is towards semi-structured databases (structures which would appear well suited to this type of domain) and searching for information starting from elements in the image description (a technique that requires sophisticated indexing methods). This would be a

useful tool for the researcher in literary genetics who has to establish the origin of a fragment of paper sometimes showing only a quarter, or less, of a watermark.

Yet the proportion of watermarked paper used by writers has grown smaller as the variety of products designed for writing purposes tremendously expanded, during the last century. As far as the knowledge of paper is concerned, further acquisition should take into account quantified data such as thickness and smoothness, to be collected with specific instruments, in order to allow the identification of unwatermarked papers as well, and printed characteristics typical of post-1870 school notebooks or hotel stationary, such as lines, cross-lines, margin lines, or letterhead. Besides enlarging the corpus beyond watermarked papers, ITEM has undertaken the development of a more refined and complex database for expert study of the modern manuscripts, using the structure of *Profil* as a basis. This new project aims at integrating a full material description (including data on handwriting) and a detailed analysis of the "genetic" process, dealing with each document page by page to trace the stages of creation. A prototype named "MUSE" is to be tested in 2000–2001 on manuscripts of Montesquieu and Stendhal. Such projects open up an entire field of future research.

Appendix 1

List of the data fields used in Profil

1. WATERMARK TYPE
Type: structure of the watermark (simple/double/triple/repetitive)
Iconographic class: reference to the IPH codes
Mark (M): brief description of the mark
Detail mark: details about shapes of letters, ornaments
Countermark (CM): brief description of the countermark
Detail Countermark: details about shapes of letters, ornaments
Position M, Position CM: number of chain lines supporting the watermark (sign "/" separating successive written lines)
Initials: as they appear in the watermark
Papermaker: name and first name of the papermaker, either as they appear in the watermark, or completed by historical research
Origin: placename, name of mill, town, region of making either as they appear in the watermark, or completed by historical research
Date of production: as it appears in the watermark
Reference to watermark: bibliographical ressources allowing to identify, localize or date the watermark
Remarks on watermark: notes on specific aspects related to the watermark, such as the existence of twins, broken or twisted wires, etc.

2. PAPER TYPE
Dimensions of fragment: as observed on the document, in case it is not a full sheet (height by width, in mm)

Dimensions of full sheet: from observation or calculation when the cutting of the fragment allows an approximate reconstruction (height by width, in mm)

Trimmed/untrimmed

Type: laid/woven/woven with chain lines

Number of laid lines per cm

Technique: hand or machine-made

Colour: estimated, not measured, description of colour (never used as a single criterium to distinguish two kinds of paper)

Quality: most prominent aspects can be chosen to qualify texture, opacity, surface or content homogeneity

Thickness: a scale of 8 degrees (from extra-thin to very thick) corresponds to average measured or estimated thickness

Smoothness: a scale of 8 degrees corresponds to measured or estimated smoothness

Tranchefils: distance between the last two chain lines on each side of the sheet: 2 entries of 2 figures (minimum-maximum) in mm

Distance between chain-lines: 2 figures (minimum-maximum) in mm, give an estimation of the distance as well as of the regularity (ex: "23–24 mm" is more regular than "28–32 mm")

Place of mark, of countermark: centered in half sheet, close to lower side of the sheet, etc.

Remarks on place of M/CM: specific cases (sideways, parallel to chain-lines, etc.)

Reference to Paper type: bibliographical ressources allowing to identify, localize or date the type of paper

Remarks on paper: notes on specific aspects related to the type of paper

3. REPRODUCED DOCUMENT

Shelfmark or number: volume where the watermark has been reproduced

Folio: folio number(s)

Folding: (of folio) in-2°, in-4°, in-8°

Author's name, first name: e.g. FLAUBERT, Gustave

Title of the work: e.g. *L'Education sentimentale*

Identification of text: sub-title or incipit, addressee of a letter, etc.

Type of medium (support): loose sheets, note-book, note-pad, etc.

Type of document: holograph manuscript, copy, copy corrected by the author, etc.

Genetic status: phase of the work concerned: draft, scenario, reading notes, fair copy, etc.

Date of composition: as given in autograph form, or completed by research

Clue for date: evaluation of authenticity (autograph, post mark, content of text)

Place of composition: as given in autograph form, or completed by research

Clue for place: evaluation of authenticity (autograph, post mark, content of text)

Remarks on the document: notes on textual content or complement of information

4. LIST OF OCCURRENCES

List of folios (beginning/end of sequences) featuring the same type of paper with the shelfmark of volumes, name of author, date and place of use

Appendix 2

Lexique Iconographique (conventionnal terms requiring explanatory comments)

A-la-mode (couple)	Eagle
Anchor	Four or Hermes cross
Bell	GR monogram
Bird	Grapes
Britannia	Griffin
Cartouche	Hand
CC monogram	Heraldry
Chain	Hj (letters)
Coat of arms of Colbert	Hollandia
Coat of arms of England	Horn
Coat of Clergymen	I (letter)
Coat of France	IHS (Jesus)
Coat of France and Navarre	Initials
Coat of Genoa	J (letter)
Coat of Le Tellier	Le Tellier monogram
Coat of London	Monogram
Coat of Montgolfier	Pelmet
Coat of Pomponne	Pot
Coat of the Pope	Shell
Crawling animal	Shield
Cross	Three circles
Crown	Vryheit
Date of making	Written parts
Dove	

Notes

[1] Daniel Charraut and Philippe Jourdain, *Laboratoire d'Optique P.M. Duffieux*, Université de Franche-Comté (1994–1996).

[2] Collaborative program between CNRS and BNF entitled 'Automatic classification of watermarks', under the leadership of C. Bustarret, originally from the GDR 34 CNRS/Bibliothèque Nationale. Data acquisition: R. Pezone, I. Gaëtan, C. Feuillerat, D. Kaiser, whom we wish to thank for their collaboration.

[3] β source 13×18 cm.

[4] There is an extra icon at the end of the form for rare cases of triple structure.

[5] The percentage reduction is shown.

[6] Application developed by E. Grandjean (ITEM/CNRS) between 1988 and 1994.

[7] Claris software.

[8] A script is a command which executes a series of actions.

References

Bockelkamp, M. "L'analyse bétaradiographique du papier appliquée aux manuscrits de Diderot". *Studies on Voltaire and the Eighteenth Century*, 254 (1988) (Oxford: Voltaire Foundation), and "Wasserzeichen in neueren Handschriften". *Editio*, 4 (1990).

Boutaine, J.-L., J. Irigoin and A. Lemonnier. "La radiographie dans l'étude des manuscrits". In CNRS, 1974, pp. 159–175.

Bower, P. *Turner's Papers, A Study in the Manufacture, Selection and Use of his Drawing Papers* (1787–1820). Tate Gallery, London, 1990.

Briquet, C.-M. *Les filigranes. Dictionnaire historiques des marques du papier . . .* (éd. originale 1907) éd. fac-similé, 4 vol. Paper Publications Society, Amsterdam, [1907] 1968.

Bustarret, C. "L'histoire du papier appliquée à l'étude des manuscrits littéraires modernes: la base de données Profil". *Livre des Congrès IPH*, 10 (1994a), pp. 39–43.

Bustarret, C. and A.-M. Basset. "Les cahiers d'Impressions d'Afrique: l'apport de la codicologie à l'étude génétique". *Genesis*, 5 (1994b).

Bustarret, C. "Saisir les filigranes, les enjeux de la description". *I moderni ausili all'Ecdotica*. Univ. di Salerno, Edizioni scientifiche italiane, 1994, pp. 385–402.

CNRS. *Les techniques de laboratoire dans l'étude des manuscrits*, no. 548. Editions du CNRS, 1974.

DAP (Deutsche Arbeitskreis für Papiergeschichte). "Online Datenbank zur Wasserzeichenerschliessung". Bad Homburg, 1995.

De La Chapelle, A. and A. Le Prat. *Les relevés de filigranes, Watermark records, I rilievi di filigrane*. La Documentation française, 1996.

Dulac, G. "L'édition des textes politiques de Diderot". *Studi settecenteschi*, vol. 14. Bibliopolis, Pavia, 1994, pp. 303–327.

Gaudriault, R. *Filigranes et autres caractéristiques des papiers fabriqués en France aux XVIIe et XVIIIe siècles*, Editions du CNRS. J. Telford, 1995.

Grésillon, A. *Eléments de critique génétique. Lire les manuscrits modernes*. P.U.F., 1994.

Hay, L. "Eléments pour l'étude des manuscrits modernes". *Codicologica*, 1 (1976), pp. 91–108.

Hay, L. "Les manuscrits au laboratoire". in *Les Manuscrits des écrivains*, CNRS éditions. Hachette, 1993.

Heawood, E. *Watermarks, Mainly of the 17th and 18th Centuries*. Hilversum: Paper Publications Society, 1950.

Irigoin, Jean. "La datation par les filigranes du papier". *Codicologica*, 5. (1980), pp. 9–36.

ITEM. *Standard descriptif pour Manuscrits modernes/Descriptive Standard for Modern Manuscripts/Beschreibungsmodell für Neuere Handschriften*. ITEM (CNRS), 1986.

ITEM/Louvre. International colloquium organized in the Bibliothèque nationale of Paris, by ITEM (CNRS) and the Département des Arts graphiques of the Musée du Louvre. *Papier et filigranes: documentation, images, informatisation*, november 1993.

IPH (International association of Paper Historians). *International Standard for the Registration of Papers with or without Watermarks*, 1996.

Jourdain, P. *Traitements numériques d'images: détection et métrologie des parcellaires – restauration d'empreintes de filigranes*. [PhD Dissertation] Thèse de l'Université de Franche-Comté, Besançon, July 1995.

Lemaire, J. *Introduction à la codicologie*. Louvain-la-neuve, 1989.

Matsiuk, O. *Filigrani Archivnykh Dokumentiv Ukrayni*, XVII–XX st. [*Watermarks from Ukrainian Archive Documents, 18th–20th C.*]. Naukova Dumka, Kiev, 1992.

Piccard. *Die Wasserzeichen Piccard im Hauptstaatsarchiv Stuttgart*, Findbuch II–XV. Verlag W. Kohlhammer, Stuttgart, 1966–1987.

Ramstein, G. and M. Raffy. "Restauration d'image par variogramme". *Traitement du signal*, 5(4) (1988), pp. 263–272.

Schmidt, F. "Überlegungen zu einer Datenbank Papiergeschichte". *IPH*, 2(1) (1992).

Spector, S., ed. *Essays in Paper Analysis*. Folger Books, Washington, Associated University Presses, London/Toronto, 1987.

Sonka M., V. Hlavac and R. Boyle. *Image Processing, Analysis and Machine Vision*. Chapman & Hall, 1993, pp. 443–457.

Stevenson, A.H. "Paper as Bibliographical Evidence". *Library*, 5th ser., 17 (1962), pp. 197–212.

Tschudin, P., C. Rauber et al. "Archival and Retrieval of Historical Watermark Images". *IEEE, International Conference on Image Processing*. Lausanne, Sept. 1996.

Web 1. http://jefferson.village.virginia.edu/gants

Web 2. http://abacus.bates.edu:80/Faculty/wmarchive

Web 3. http://www.oeaw.ac.at/~ksbm et http://www.iinform.oeaw.ac.at/~weng

Web 4. http://ada.cath.vt.edu:591/DBs/Gravell/default.html

Wurm, H. "Eine papiergeschichtliche Datensammlung, Italienische Architektur-Zeichnungen der Renaissance als Mode". *Gutenberg Jahrbuch* (1986).

Existing reference books for the identification of watermarks are traditionally heavy volumes with hand-drawn watermarks, roughly classified by iconography, accompanied by more or less succinct historical information, and listed in an index (Briquet, 1907; Heawood, 1950; Piccard, 1966–1987). These indexes, both on a national and an international scale (mostly European), deal mainly with corpora prior to 1600, and rarely extend to nineteenth and twentieth century industrially produced watermarked papers (Gaudriault, 1995; Matsiuk, 1992). Today, the best solution for recording, describing and visualizing modern watermarked papers appears to be a multimedia database with digitized images, much more accurate than hand copies (Bustarret, 1994c). Besides its usefulness for specialists, for whom the IPH "Standard for registration of watermarked and un-watermarked Paper" (IPH, 1996) has been conceived, such a database can also have a didactic use for researchers who are not particularly familiar with the history of modern paper.

There are few such watermark databases, and most of them are too incomplete to be of real value as research and analysis tools, yet they may be referred to as experimental models. Developed on a historically limited corpus of documents (Wurm, 1986), or embracing as wide a scope as possible (Schmidt, 1992), most of these applications are still under study (ITEM/Louvre, 1993). Mainly focused on watermarks, they rarely take into account other characteristics of the paper (such as chain-lines or texture) as recommended by Historians of paper (IPH, 1996), and of the document (nature of use, nature of the text), as required by scholars in litterature. Moreover, they tend to represent each identified type of paper by only one (dated) example, as is traditionally the case since Briquet, whereas *Profil* offers up to several hundreds of occurrences for a type of paper. It thus provides ground for research far beyond the mere question of dating.

Nevertheless, such databases are starting to appear on the web. One can for example have access to all the watermarks of the papers used by William Stansby in 1616 in '*The workes of Beniamin Jonson*' (Web 1). These are itemized in three classes: 'standard' papers, 'large' papers and 'other' papers used by the author in other works. For each one we find, apart from the watermark, characteristic information such as the chain or laid lines, a quick description and a mention of legible items (date, initials, etc.). This site is in fact an example of what could be done, but for the moment it only contains fifteen watermarks in the 'standard' paper class, and nothing in the other classes.

Another site (Web 2), called '*The Archive of papers and watermarks in Greek manuscripts*', contains an image database (watermarks), a descriptive database for papers (identification, description of the manuscript and of the paper), a biblio-graphic database, and the possibility of submitting new watermarks or descriptions for storage. As the name of the site suggests, we find there all the ingredients of what such an archive could or should contain, including a catalog of the different watermark classes described in the archive. For the time being, it is above all a suggestion of site architecture, with instructions and tools so that future users can

enter data. But regretably, the navigation is not very intuitive, images have not been enhanced and are therefore hard to identify, and image downloading time is prohibitive.

The (Web 3) site contains watermarks of papers from the Middle-Ages. It does not describe them, but gives the process of treatment that can be applied to the watermarks to make them more easily readable. This process has been applied to three images (a tower, a corsair, and a fleur de lys) that are presented in a first version which is barely readable (very dark, with no details), in a second version where one can guess the motives, and in a third where the watermarks are clearly identifiable. For these examples, the demonstration is clear, but it has not been applied to the 350 other images available only in their original form and barely readable.

Valuable insights are found in the (Web 4) site, which shows careful conception of the various prerequisites for a research instrument in the field. Developed on a FileMaker Pro application with the support of the Virginia Polytechnic Institute and State University, the *"Gravell Watermark Archive"* presents an important collection of digitized images obtained by Dylux reproduction process from manuscripts and books dating from 1400 to 1835. Although we have developed our applications separately, it seems we share a number of options concerning the type of information such a database should provide and how it should be presented.

In this article we shall be presenting the *Profil* database, developed by the 'Supports et Tracés' team of the Institute of Modern Texts and Manuscripts (ITEM, CNRS), the images of which have been digitized[1] from a collection of betaradiographic photographs from the Bibliothèque Nationale de France (BNF).[2] We shall also tackle the problems arising from the requirement for normalization in a database in the discipline of human sciences where the management of exceptional situations is necessary, and in particular the problem of conserving the accuracy of details during automatic processing of information.

1. The Descriptive Standard for Watermarked Papers

The ITEM descriptive standard for watermarked papers is based on the codicological data from modern literary manuscripts accumulated over a twenty-year period (ITEM, 1986; Bockelkamp, 1988, 1990). The data structure has evolved with innovations in computer hardware and software, especially with the possibility of incorporating digitized images in an application, which has obviated the need for lengthy descriptions. A first version of the *Profil* application was presented in the international workshop organized in Paris by ITEM and the Musée du Louvre in 1993.

Although members of ITEM were not personally involved in the realization of the IPH Standard, which matured during the same period, we did take part in the discussion over the years, and it makes sense that almost all of our descriptive entries were later found to match the "basic" items marked with a "*" in the IPH

1 - Watermark description
     Type
     Iconographic code
     Mark description and mark detail
     Countermark description and countermark detail
     Position of the mark and countermark
     Images of the mark and countermark
     Initials
     Paper maker
     Origin
     Date of manufacture
     Bibliographic watermark reference
     Additional information on the watermark

2 - Paper support
     Dimensions of the fragment and of the full sheet
         (trimmed/ untrimmed)
     Type of paper
     Technique
     Color
     Quality
     Thickness
     Smoothness
     Number of laid lines per cm
     Distance between chain-lines, Tranchefil (mini/ maxi)
     Position of the mark and countermark on the sheet
     Bibliographic paper reference
     Remarks on paper

3 - Reproduced Document
     Manuscript location number
     Folio
     Folding
     Author's name and first name
     Title of the manuscript
     Text content identification
     Type of support
     Type of document
     Genetic status
     Writing date and source of writing date
     Writing location and source of writing location
     Remarks on the document

4 - List of Occurrences
Author's name, location number, folios, writing location and dates

Figure 1. Description of a watermarked paper (see Appendix 1 for further details).

issue of 1996. This publication led to several minor revisions in the presentation of our data, yet we did not fully adopt the IPH Standard for two major reasons: a number of recommended elements of description are not easily accessible on fragments or folded sheets; moreover, we do not usually have any access to chemical or optical analysis. We purposefully chose to accept some limitations in our contribution to paper history, considering as our priority that our descriptions should enable untrained readers to take a much closer look at the material features of a manuscript to make use of paper evidence in their research.

Our medium level of expectation in this first database project does not entail an approximate collection of data: as many fields as possible in our form have been completed for each type of paper identified. On the other hand, *Profil* does provide a much wider collection of occurrences (see below Section 4) for each type of watermarked paper recorded than is usually available in current watermark publications.

Each watermarked paper type is described by about fifty data entries (see Figure 1), divided into four sections: 1. Watermark; 2. Paper; 3. Reproduced document; 4. List of occurrences.

a) Watermark

Depending upon the manufacturing technique (manual or mechanical) and the period, the structure of the watermark can be *simple* (single sign), *double* (associated mark and countermark, most common between 1650 and 1800), sometimes *triple* (mark, countermark and additional symbol), or *repetitive* (modern structure due to the manufacturing process, for example the repetitive 'Extra Strong' stripe on contemporary A4 size typewriter paper). The watermark can be spotted by its position on the sheet of paper – a double watermark divides into a '**mark**' and a '**countermark**', generally set in each half of the sheet (which was sold folded *in-folio*). The watermark includes characteristic elements of the papermaker and mill where it was produced, so in that way it can contribute to the dating of the document.

b) Paper

Laid paper differs from **wove** paper (which first appeared by the end of the 18th c.) by the presence of regularly spaced lines (distant by 18 to 35 mm) called the **chain-lines**, which form a network with the perpendicular laid-lines (8 to 12 lines per cm). When the watermark is only partially visible on a fragment of sheet, these structural elements can help identify a type of paper. On the other hand, their absence in wove paper make it more difficult.

c) Reproduced document

This section gives information on the document containing the occurrence chosen for reproduction in the database. The user can find the **location number** (shelf-mark) of the manuscript (in the present case, that of the BNF) and the reproduced

folios (if the watermark is double, the mark and countermark appear at the center of two in-folio pages; it can even require two photographs for each mark if the document is folded into four, as the watermark then appears along the crease between the two pages). Therefore the indication of the **folding** of each page (in-folio, four-fold, eight-fold) is important: in principle it allows the initial manufacture format of a whole sheet to be reconstituted, from just a fragment.

The data concerning the author (name and first name) and the **title of the manuscript** and of the chapter or, failing that, a quotation of the first lines of the reproduced sheets (**identification of the text**), is completed by three headings that characterize the **type of support** (separate sheets, notebook, notepad, etc.), the **type of document** (autograph manuscript, corrected print, etc.) and the **genetic status** of the document (first draft, structure, rough work, fair copy, etc.). Finally, the date and place of writing of the document, when known, are added: they are essential data for the codicological classification of the manuscript, as they enable cross-checking with other documents written on the same type of paper. This is why as far as possible we selected dated documents for representation in the base.

d) List of occurrences

The list of occurrences of the paper in the same volume, and also in other manuscripts by the same author or even by contemporary authors, are recorded in a separate file, which appears as a list at the end of each descriptive form. As each type of paper is represented in only one form, this list of occurrences (which is liable to grow with research on extended corpuses) constitutes a capital source for the cross-checking of dates and places of use. Unlike classical printed collections, which only gave an isolated example, dated more or less precisely, the sheet from which the watermark has been copied has a context of use (guaranteed as far as possible by an autograph dating of the writer), which constitutes a crucial element of validation for the researcher.

All of these sections enable the user to carry out searches concerning both the history of the paper and the study of the documents. For instance we can list all the types of paper that were used for the long and laborious writing of the *Tentation de Saint Antoine*: the Britannia symbol, which indicates a British-made paper, appears under different forms, associated with the papermakers Snelgrove and J. Whatman, in 'vintage' watermarks which have the advantage of giving manufacturing dates (1831, 1844, 1846 and 1848) (Profil n. 301, 295, 296, 297). Two French papers, marked 'BFK/RIVES' (Profil n. 299, 381) can also be found. They are produced by the Blanchet Frères et Kléber company in Isère; the writer used this paper quite regularly. These different types of paper correspond to different stages in the writer's work.

If we enlarge the research, we will find either all the types of paper used by Flaubert (19 have been listed), or other occurrences, situated around the same time period, of papers manufactured by the same paper makers. Furthermore, thanks to

the vintage of the watermarks, we note a certain regularity (as well as significant exceptions) in the length of time between the manufacturing and the use of the paper. If, on the other hand, our approach is centered on the identification of the watermark, and not on written documents, we can carry out a search on a motive such as that of the *shell*, to compare the 38 identified types, all of the nineteenth century, according to their geographical characteristics, according to changes in size or so as to estimate the area and the duration of use of this motif in French or European papers. To give a precise identification example, we can quote the case of the mark '*aux armes de Pomponne*' associated with the name of the papermaker Malmenaide, which appears in the *Lettre sur le commerce de la librairie* and in the letters of Diderot (Profil n. 125): this enables us to date an undated manuscript, named *Discours au roi par les députés d'une cour souveraine* (Dulac, 1994) where it also appears.

2. The Watermarks

The descriptive database is complemented by an illustrative database. Unlike those offered for instance in the database developed by the Schweizeriches Papier Museum and the University of Geneva (Tschudin and Rauber, 1996), all images in *Profil* have been digitized from betaradiographs of original documents. Now, the displaying of watermarks requires an acceptable quality of images with all the relevant information. To obtain the highest quality, digital processing has to be used both in recording and restoring the image. The appropriate techniques have been developed for the purpose of our application by Daniel Charraut and Philippe Jourdain in the laboratory of Optics P.M. Duffieux in Besançon (Jourdain, 1995).

2.1. THE IMAGE SOURCE: BETARADIOGRAPHY

Reproduction of the watermarks is carried out by betaradiography of the manuscripts (Boutaine *et al.*, 1974). The technique consists of making a sandwich of a β-particle emitting plate,[3] the document and a photosensitive film. The transmitted intensity of radiation depends on variations in density and thickness of the document. The levels of gray at each point of the photo obtained reveal the motif impressed into the thickness of the paper. The advantage of the method is that all traces of ink on the surface are eliminated without any loss at all of the watermark image. The photo is then digitized by scanner (cf. Figure 2). The digitized image is composed of gray values coded at each point with one of 256 values. This betaradiography forms the source image for subsequent processing.

2.2. IMAGE PROCESSING

The digitized image now has to be processed to improve quality and legibility. There naturally exist numerous defects due to the nature of the support (thick-
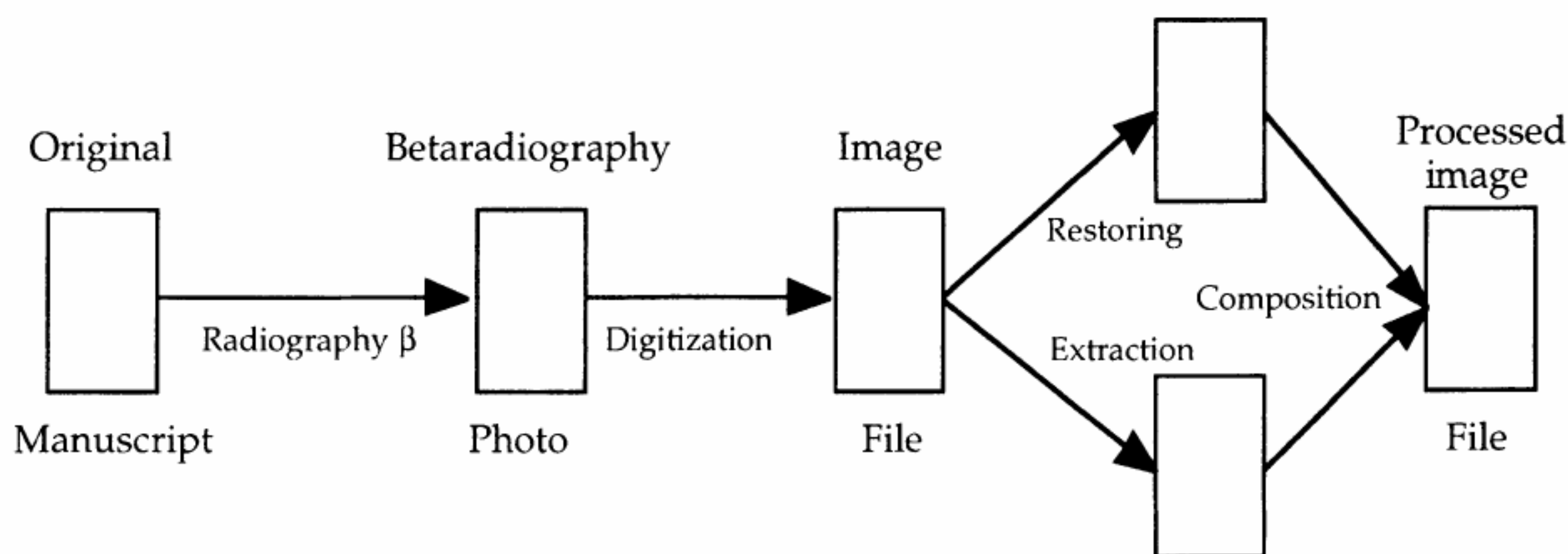
Figure 2. Image processing.

ness, texture) or to the photographic print (imperfect contact) which result in local variations in contrast. These are particularly visible in the traces left by the chain-lines and the laid lines, and interfere with the interpretation of the image. Now, the chain-lines play a vital role in recognizing watermarks and identifying papers. The laid-lines on the other hand are not a vital source of information, at least visually, for papers from the modern period.

2.2.1. Restoring the Image

The first step is thus to use an image restoring technique to increase the legibility and overall understanding of the image. In the present case, the aim is to eliminate effects due to the weave and to homogenize the image.

To do this, a 'noise map' is made by evaluating the fluctuations in contrast over the whole of the image (Ramstein, 1988). This map is obtained by low-pass filtering of the two-dimensional Fourier spectrum of the image (Sonka *et al.*, 1993). The map obtained is then subtracted from the initial image to obtain a new image upon which various operations are carried out to make the contrast uniform and enhance the image strength in areas of low brightness. Finally, the weave pattern is eliminated by spatial filtering. The resulting image is homogenous and devoid of weave pattern (Jourdain, 1995).

2.2.2. Extracting the Outlines

Apart from defects in paper thickness or texture, there may be others arising from the pulp composition and the paper manufacturing technique, which affect the legibility of the outlines. To eliminate these and bring out the watermark lines an image smoothing technique is used. This consists of segmenting the image and sorting those fragments that bear the characteristics of the background from those bearing the watermark lines. This process is then repeated on smaller and smaller fragments.