

Le point de vue de l'utilisateur

Aline Durand, Élisabeth Lalou

Citer ce document / Cite this document :

Durand Aline, Lalou Élisabeth. Le point de vue de l'utilisateur. In: Le médiéviste et l'ordinateur, Numéro spécial, 1990. Actes de la Table ronde (Paris, CNRS, 17 novembre 1989) pp. 117-124;

doi : <https://doi.org/10.3406/medio.1990.1251>

https://www.persee.fr/doc/medio_0223-3843_1990_hos_1_1_1251

Fichier pdf généré le 08/01/2020

LE POINT DE VUE DE L'UTILISATEUR

Elisabeth LALOU - Aline DURAND

Les deux travaux qui vont faire l'objet de ce développement n'en sont pas au même stade d'avancement. Dans le cadre d'une thèse de doctorat, A.Durand a entrepris le traitement informatique des chartes médiévales languedociennes avec le logiciel TEXTO depuis 1985-1986. Pour sa *Gallia regia philippica*, E.Lalou vient de terminer l'étape des essais de traitement avec le logiciel ISIS depuis un an. Dans les deux perspectives, l'aide des informaticiens de l'IRHT a été précieuse. Pour A. Durand, elle a même été irremplaçable, car la base n'aurait sans doute jamais vu le jour. La construction des deux fichiers reste celle de bases individuelles. Cependant, la création d'une banque européenne sur les chanoines¹ offre à E. Lalou la possibilité future d'intégrer une partie de ses données dans une entreprise plus vaste et d'articuler ainsi son travail au sein d'une équipe.

I.- Les sources : La description des données

Pour E. Lalou, il s'agit de construire une base de données prosopographique sur les officiers royaux du règne de Philippe le Bel dénommée, sur le modèle de celle de Dupont-Ferrier, "Gallia regia philippica". Soit à peu près 4000 à 5000 fiches lentement élaborées et sédimentées depuis une trentaine d'années.

Il s'agit de la refonte d'un travail traditionnel qui doit mener à terme à un dictionnaire. En plus du dictionnaire traditionnel dont elle devrait avoir la perfection de présentation, cette base fournira des index multiples. Il sera en outre toujours possible d'enrichir la base au fur et à mesure que de nouvelles sources seront dépouillées ; la difficulté majeure est le manque d'homogénéité des sources.

Le corpus TEXTO élaboré par A. Durand est constitué de 5400 actes diplomatiques des Xe-XIIe siècles, principalement des chartes et quelques censiers et testaments. La problématique de la recherche étant de comprendre l'évolution des campagnes bas-languedociennes durant le Moyen Age central, l'objectif principal du projet informatique consiste en une analyse diachronique des structures agraires. L'organisation du fichier s'articule autour des cellules économiques et sociales de base du monde rural : parcelles, manses, *villa*, *castrum*... Les choix opérés dans la sélection des documents et, dans le document, de l'information à prendre en compte procèdent directement de cette problématique.

1. Cf. la communication d'É. Mornet, 2e. session.

II.- La construction d'une métasource

1) CHOIX DU LOGICIEL

Les raisons pour lesquelles nous n'avons pas choisi le même logiciel sont multiples et complexes. Le temps, les rencontres interviennent et se mélangent. La seule certitude est que ni l'une ni l'autre ne voulions ni ne pouvions coder nos données. La saisie en langage clair et naturel, le plus proche de la source originelle, est l'une des premières raisons qui nous ont orientées vers un logiciel de type documentaire.

Partant du principe que le logiciel idéal n'existe pas et que l'outil informatique doit se plier à notre démarche scientifique propre, l'une de nous a sélectionné TEXTO. Les contacts pris avec les responsables d'expériences méthodologiques menées avec ce logiciel¹ l'ont convaincue des avantages de cet instrument de travail.

E.Lalou a également testé très rapidement TEXTO qui l'a rebutée par sa difficulté d'accès. Elle a continué à chercher le logiciel de rêve, en attendant PROSOP. A été testée la version micro de CLIO, KLEIO, qui n'était pas encore au point et seulement partiellement implantée sur micro² ; puis DBASE III, mais la différence qui existe entre un système documentaire —de type TEXTO ou ISIS— et un système de gestion de base de données est vite apparue. Après quelques tests d'ISIS, il a été décidé de continuer plus avant avec ce logiciel. D'autant plus qu'une base de données sur le personnel de la Curie de Boniface VIII est en cours de création sur ce même logiciel.³

Evoquons en trois mots la différence entre TEXTO et ISIS. TEXTO présente à l'entrée des données un aspect rébarbatif (à présent compensé par l'utilisation d'éditeurs). Certes, les possibilités de tris sont plus vastes que dans ISIS, puisqu'on peut interroger sur tout le fichier avec une troncature gauche, ISIS dans sa version 2.3 ne permettant d'appliquer la troncature qu'à un seul champ. Mais maîtriser les interrogations sur TEXTO semble assez délicat. ISIS fournit très rapidement un "dictionnaire" très réconfortant, qui donne des idées et permet de travailler à la standardisation du vocabulaire utilisé dans les fiches, et même à un thesaurus pour lequel des données de programmation avancée sont fournies et seront probablement développées dans les versions futures du programme.

L'aspect matériel n'est pas à négliger. Les deux logiciels tournent aussi bien sur des AT avec disque dur que sur gros système. TEXTO est distribué pour un prix assez élevé par son fabricant. ISIS dans sa version micro est fourni gratuitement par l'UNESCO. Mais le suivi de TEXTO est bien assuré par l'entreprise-mère qui répond volontiers à tous les problèmes posés.

1. A. Pralong, Répertoire informatisé des sites byzantins ; A. Vallée, Inventaires et index analytiques informatisés des registres de chancellerie du règne de Jean le Bon conservés aux Archives nationales ; A. Sturla, Informatisation du chantier archéologique du Louvre.

2. CLIO et son adaptation sur micro KLEIO ont été créés par M. Thaller (Göttingen), avec la collaboration de J. Smets (Montpellier).

3. Thérèse Montecchi travaille sur ce projet à l'Ecole française de Rome.

2) STRUCTURATION DU FICHER

Les deux logiciels fournissent des champs répétitifs non limitatifs¹. Les sous-champs n'existent que dans ISIS. Cette dernière notion, tant sur le plan conceptuel que sur le plan pratique, est difficile à maîtriser. Dans quelle mesure tel sous-champ fournira-t-il des réponses intéressantes à l'interrogation ? Il faut tenir compte aussi de la contrainte — légère mais contrainte malgré tout — qu'il y a à entrer des données avec des sous-champs (un accent circonflexe et une lettre prédéfinie à mémoriser avant les données).

a) Description du fichier ISIS

La structuration du fichier de la *Gallia* ressemble à celle de beaucoup de fichiers prosopographiques et nous renvoyons à l'exemple fourni en annexe (fig.1). Bien des questions se posent pourtant : la simple question de savoir comment nommer les personnages n'est pas encore vraiment résolue (nom, prénom, surnom, titre, variante de nom, désignation). Pour résoudre le problème de l'hétérogénéité des fiches, il existe deux solutions : on peut créer plusieurs bordereaux qui s'additionnent ou bien utiliser "les champs déroulants" illimités.

b) Description du fichier TEXTO

Conception théorique d'ensemble

La base du fichier est logiquement constituée par le ou les biens faisant l'objet de l'acte étudié, chaque bien répertorié ouvrant une nouvelle notice dans le fichier TEXTO. Cette conception implique a priori la répétition d'informations indispensables et communes à tous les biens d'un même acte (la date, la référence du texte, les parties en présence, etc...) qui alourdissent la saisie et la gestion. Pour résoudre ce problème, TEXTO autorise le chaînage de fichiers : il s'agit de mettre en correspondance un document d'un fichier avec un document d'un autre fichier. Ici, chaque bien du fichier objet ou fichier chaînant est associé au document unique du fichier texte ou fichier chaîné stockant les informations communes à tous les biens d'une même charte (fig. 2). Sous TEXTO, le chaînage est conçu comme une simple juxtaposition de fichiers indépendants et non pas comme une véritable base relationnelle. Néanmoins, avec l'aide d'un informaticien et de LOGOTEL, le langage de programmation maison, cet obstacle de taille peut être contourné (voir présent texte d'A. Vallée).

Organisation des fichiers

L'organisation des fichiers suit le déroulement logique de l'acte diplomatique. Le fichier objet comprend d'abord la description de l'objet de l'acte, sa localisation médiévale et actuelle, l'énumération des parties constitutives du bien, ses confronts, les redevances et droits afférents, le type de contrat d'exploitation ou de liens d'homme à homme et éventuellement sa valeur marchande, sa superficie et les renseignements concernant les cycles et façons culturelles. Le fichier texte, lui, est constitué des champs référence du document, date, parties en présence, rapport entre les deux parties, type juridique du bien, témoins, lieu de rédaction de l'acte.

1. Dans ISIS 2.3, il semble que la limite du champ soit de 8000 caractères

III. L'exploitation de la métasource

1) *UN FORMAT STANDARD*

TEXTO comme ISIS sont des systèmes d'exploitation ouverts c'est-à-dire que des transferts de fichiers vers l'extérieur et inversement sont possibles. De telles options ouvrent de séduisantes perspectives :

— Ne pas mobiliser pour la saisie le poste de travail où est implanté le logiciel, avantage non négligeable lorsqu'on travaille sur serveur.

— Exploiter une même base de données sous d'autres machines et d'autres logiciels pourvus d'un format éditeur.

— Fusionner et échanger les données avec celles d'autres banques.

Plusieurs tables rondes et colloques ont récemment souligné l'importance d'une standardisation des bases de données. En effet, la micro-informatique a multiplié les entreprises individuelles, les faisant sortir des circuits traditionnels. Cette recherche d'un format standard est capitale dans les années à venir.

2) *TRIS ET INDEX*

a) *La philosophie des deux logiciels*

La conception d'ensemble des deux logiciels conditionne la gestion des questions.

Sous ISIS, version 1, la construction du dictionnaire exigeait, dès sa mise en oeuvre, que l'ensemble des traitements fussent tous pensés à l'avance. Il n'y avait pas moyen de revenir dessus par la suite sans une modification fondamentale de la base. Dans le cas où l'on travaillait avec des mots ou des groupes de mots, l'information était déjà sélectionnée et hiérarchisée. La version 2.3 d'ISIS a assoupli considérablement cette méthode de travail qui demeure cependant l'essentiel de sa mise en oeuvre.

La philosophie de TEXTO est radicalement différente : la souplesse du logiciel permet de ne pas prévoir à l'avance toutes les questions possibles, par conséquent de ne pas pousser l'analyse préalable jusque dans ses dernières extrémités, ce qui laisse une marge de manoeuvre appréciable.

Sous ISIS, la forme des questions s'oriente vers une perspective conceptuelle du type thesaurus. Sous TEXTO, elle reste très proche du texte, des mots, du contexte, même si elle peut être revue par la suite dans l'optique d'un thesaurus.

b) *Gestion des questions*

L'exploitation de la base passe par l'indexation de champs-clefs. ISIS autorise une indexation sur tout le champ, les sous-champs, certains mots (par marquage <>) ou encore tous les mots d'un champ, avec possibilité de créer ensuite un fichier de "mots vides". Dans TEXTO, les champs date et lieu sont systématiquement associés à la question posée. L'interrogation est séquentielle.

Dans les deux logiciels, la syntaxe autorise tous les opérateurs logiques dont le "sauf", les parenthèses à plusieurs niveaux, les symboles "supérieur à" et "inférieur à" et les troncatures droite et gauche. Des critères de tris très sélectifs sont ainsi définis.

Pour affiner davantage les traitements, il est judicieux de décomposer la question en plusieurs étapes, soit en ne travaillant que sur une partie du fichier et en obtenant ainsi un tri

arborescent sur plusieurs niveaux, ou bien en jouant sur les combinaisons de sous-ensembles et ré-assemblages partiels de sous-fichiers.

Dans l'un et l'autre cas, la gestion des sorties offre toute une gamme de possibilités : choix des champs pertinents à prendre en compte, obtention d'un simple listing ou d'une édition soignée en bonne et due forme.

Pour ce qui est des traitements post-listing, ISIS possède un logiciel statistique (IDAMS-PC), alors que TEXTO est pour l'instant particulièrement démunie, laissant à l'utilisateur chevronné le soin de programmer via LOGOTEL.

3.- L'EXPLOITATION DU FICHIER SUR PC ET SES LIMITES

A l'heure actuelle, l'exploitation d'une base de données sur micro-ordinateur comporte des limites. Dans le cas d'A. Durand, tous les traitements prévus sur PC n'ont pu être effectués. Deux raisons principales à cela : un disque dur de 20 méga-octets s'est avéré insuffisant pour gérer l'ensemble de la base et tout l'appareil des questions, index, sous-fichiers etc... et surtout les temps-réponses aux questions se sont révélés beaucoup trop longs. Le tableau (fig.3) en offre un ordre de grandeur sur AT avec co-processeur INTEL 80286. Si la durée des traitements pour des questions simples (de 0 à 2 opérateurs logiques) est acceptable, en revanche, celle nécessitée par des questions complexes (>3 opérateurs logiques différents) est évidemment beaucoup moins satisfaisante. Mais l'écueil principal sous PC reste les capacités de l'imprimante, tant en rapidité qu'en place de mémoire tampon, qui double pratiquement l'attente et qui bloque l'accès à d'autres tâches. Or, lorsque le véritable stade d'exploitation de la base est atteint (plus de 400 listings ont été obtenus dans le cas de TEXTO), le rythme de travail sous PC devient vite un véritable casse-tête, en particulier lorsque plusieurs centaines de documents sont à éditer. Le recours à l'informatique lourde, en l'occurrence le CIRCÉ, s'est imposé. Ce verrou incontournable au moment où le travail a commencé sautera probablement bientôt avec une nouvelle génération de co-processeur et de disque dur à grande capacité.

CONCLUSION

Les deux exemples montrent que si les matériaux à informatiser sont différents, les problèmes d'ordre conceptuel et analytique se rejoignent. Deux états différents d'une base de données ont été développés : un fichier très proche du document initial où le contexte prime (TEXTO), et un fichier plus hiérarchisé devant déboucher sur une maquette avec thesaurus (ISIS).

Notre travail paraît sans doute prosaïque mais il aboutit à des résultats tangibles. Si les utilisateurs de logiciels documentaires sont encore peu nombreux, et cela serait a priori un handicap, en revanche ce petit nombre facilite des échanges fructueux et actifs.

Table de définition des champs

Field Definition Table (FDT)

Data Base: GALIA

?	Tag	Name	Len	Typ	Rep	Delimiters/Pattern
-	10	numéro	10	X	R	
-	11	nom ou toponyme	50	X		
-	12	prénom	30	X		
-	13	Surnom ou patronyme	50	X		
-	103	Désigné comme	50	X	R	
-	101	titre	50	X	R	
-	14	variante de nom	50	X	R	
-	15	fonction	50	X	R	
-	115	fonction province	80	X	R	tld
-	200	hôtel	80	X	R	tld
-	16	attestation	14	X		pd
-	17	naissance	80	X		dml
-	18	mort	80	X		dml
-	19	origine	50	X	R	
-	20	sépulture	40	X	R	
-	51	Etat	80	X	R	
-	21	études	50	X	R	geld
-	22	livres	80	X	R	
-	220	vie intellectuelle	150	X	R	
-	23	mariages	30	X	R	
-	24	enfants	80	X	R	
-	25	ascendance	80	X	R	pm
-	26	collatéraux	80	X	R	
-	27	bénéfices	100	X	R	tldc
-	28	hiérarchie eccl.	100	X	R	tldc
-	29	commentaire	1650	X	R	
-	30	carrière avant 1285	150	X	R	
-	31	carrière	150	X	R	
-	300	carrière ailleurs	150	X	R	
-	310	carrière commentaire	150	X	R	
-	32	affaires traitées	300	X	R	
-	33	lettres commandées	300	X	R	
-	34	gages	60	X	R	
-	35	missions	300	X	R	
-	350	commissions	300	X	R	
-	40	itinéraire	100	X	R	
-	44	conseil	150	X	R	
-	48	Parlement	80	X	R	
-	49	Enquêtes 1	150	X	R	
-	490	Enquêtes 2	150	X	R	
-	491	Enquêtes 3	150	X	R	
-	492	Enquêtes 4	150	X	R	
-	493	Enquêtes 5	150	X	R	
-	50	armée	100	X	R	tldc
-	60	biens	100	X	R	
-	70	maisons	100	X	R	
-	80	dons	60	X	R	
-	84	testament	60	X	R	
-	841	obit	60	X	R	
-	842	fondations et dons	100	X	R	
-	90	affaires privées	100	X	R	
-	100	entourage	80	X	R	
-	110	bibliographie	50	X	R	
-	112	Sources	50	X	R	
-	120	commentaire	1650	X		

Figure 1

FICHER OBJET OU FICHER CHAINANT

Bien n°1

- localisation
- énumération des parties constitutives
- énumération des confronts
- redevances
- droits afférents (droit d'usage, droit banaux...)
- type de tenure
- etc...

Bien n°2

- localisation
- énumération des parties constitutives
- énumération des confronts
- redevances
- droits afférents (droit d'usage, droit banaux...)
- type de tenure
- etc...

Bien n°3

- localisation
- énumération des parties constitutives
- énumération des confronts
- redevances
- droits afférents (droit d'usage, droit banaux...)
- type de tenure
- etc...

Etc...

FICHER TEXTE OU FICHER CHAINE

Référence

Date

Parties en présences

Type de rapport entre les parties

Type juridique du bien

Temoins

Lieu de rédaction de l'acte

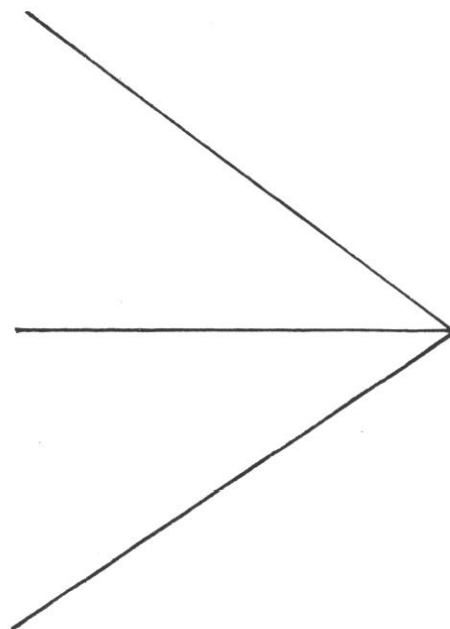


Figure 2 : le chaînage de fichiers sous TEXTO

ORDRE DE GRANDEUR DES TRAITEMENTS INFORMATIQUES

SOUS ISIS ET TEXTO

(temps du tirage listing compris)

TEXTO

pour 21 000 fiches

Indexation

Création d'un index sur 1 champs	35 mn
Création d'un index sur 2 champs	60 mn

Temps réponses aux questions

Question simple avec une dizaine de réponses	30 mn
Question simple ayant une centaine de réponses	1h
Question complexe ayant une centaine de réponses	2h
Question complexe ayant plusieurs centaines de réponses	4h et plus

ISIS

sur 2000 fiches

Indexation

Création du fichier-inverse (dictionnaire)	10 mn
--	-------

Temps réponses aux questions

En passant par le fichier ANY avec l'impression	30 mn 2 h
--	--------------

Figure 3