

*REUNION DU COMITE DE L'INFORMATIQUE DU C.I.A.
WASHINGTON DC SEPTEMBRE 1976*

*NOTE SUR LES FICHIERS NOMINATIFS SUR SUPPORT INFORMATIQUE
INTRODUCTION*

Les fichiers nominatifs constituent une source particulièrement importante d'informations de toute nature. Ils peuvent intéresser les domaines les plus divers de l'activité humaine et sont particulièrement nombreux dans le cadre de ce qu'il est convenu d'appeler les sciences humaines. Depuis l'Etat-civil, la liste de baptêmes, la liste d'électeurs, la liste des membres d'une association, qui constituent le stade élémentaire des fichiers nominatifs, en passant par les fichiers des recensements établis par les Etats, les Provinces, les Communes selon les modalités propres à chacun d'eux, pour aboutir aux dossiers médicaux, aux dossiers administratifs les plus complets, les fichiers nominatifs offrent un éventail de renseignements que pourront utiliser les spécialistes des disciplines les plus diverses : du généalogiste au généticien en passant par le démographe, l'historien, le philologue ou le sociologue.

Nous allons tenter dans cette note d'envisager les problèmes que les Archivistes peuvent rencontrer lorsqu'ils sont confrontés avec les questions qui se posent lors de la création dans le cadre de leur Institution, lors de l'archivage, lors de la mise sur support informatique. L'organisation, la consultation et les types de liaison à établir entre ces fichiers afin d'en créer éventuellement des banques de données feront également l'objet d'un examen. Nous n'oublierons pas, bien entendu, d'essayer

d'approcher le problème, capital dans ce domaine, du matériel et du personnel nécessaires à ces travaux.

ORIGINE DES FICHIERS NOMINATIFS

Nous pouvons *a priori* déceler deux origines différentes pour les fichiers nominatifs dont peut disposer un archiviste.

La première, et celle qui normalement doit poser le moins de difficultés, est la création par l'archiviste lui-même d'un ou de plusieurs fichiers nominatifs. Nous nous trouverons alors en présence de fichiers qui seront par exemple des tables onomastiques d'une publication, des tables de noms de lieux et/ou de personnes, de séries de documents conservés, d'instruments de travail relatifs à des collections, etc... Dans ce cas, l'archiviste, qui est l'auteur de ces fichiers, a pris la responsabilité lors de leur création de leur donner un format déterminé, de les conserver sur un support bien défini, généralement tributaire du matériel dont il dispose, de les organiser en fonction des besoins des travaux en cours et des fichiers qui existent déjà. Il a également prévu la consultation de ces fichiers et résolu les questions relatives à l'environnement de ces fichiers : confidentialité, protection, etc...

Les fichiers nominatifs peuvent également avoir une origine étrangère aux Archives. Dans ce cas, outre le processus normal et traditionnel d'archivage : estimation de la valeur documentaire présente et future, catalogage, possibilité de consultation, mode et qualité de conservation, d'autres questions doivent se poser pour lui : reformatage éventuel, insertion dans des ensembles plus grands, etc... Ces questions doivent évidemment recevoir une solution rapide et satisfaisante.

UNITE DES FICHIERS NOMINATIFS

L'intérêt primordial de conserver des fichiers nominatifs est de mettre à la disposition des intéressés des listes de noms de personnes, éventuellement de noms de lieux, munies de toute une série de renseignements. Il semble bien dès lors que l'ensemble des opérations visant à conserver, à consulter, à relier ensemble un certain nombre de fichiers nominatifs a un point commun essentiel, c'est le NOM de la personne de l'institution, du lieu, ... C'est donc bien au départ le caractère nominatif du fichier qui devra retenir l'attention de celui qui le conservera ou le mettra à la disposition des utilisateurs.

Un certain nombre d'opérations va donc être nécessaire lors de la création et de l'acquisition de tels fichiers. Nous allons envisager l'une ou l'autre de ces opérations.

CONTROLE DES PATRONYMES

Nous envisagerons ici le cas des patronymes proprement dits, les mêmes considérations peuvent naturellement être émises pour des noms de lieux, des noms de firmes ou d'institutions.

Nous savons que l'orthographe des noms de personnes aussi bien que l'orthographe de leurs prénoms est facilement l'objet de distorsions. Une première obligation, outre l'assurance, prise par le créateur du fichier, d'une concordance parfaite entre le document original et l'enregistrement informatique serait de procéder à un contrôle de vraisemblance des patronymes repris dans le fichier; de procéder éventuellement, par un ajout

d'informations au record, à une unification sinon même à une codification des patronymes en présence.

Si le fichier est originaire de l'Institution archivistique, le problème peut déjà se poser. Citons pour mémoire les nombreuses graphies différentes qu'un même patronyme peut se voir attribuer dans des séries de documents tels que les registres paroissiaux, les documents d'Etat-civil, les recensements anciens ou contemporains, les actes de justice, etc... Ces différentes graphies ont plusieurs causes : non fixation de l'orthographe à certaines époques, méconnaissance de l'orthographe par celui qui a rédigé le document, méconnaissance aussi (analphabétisme ?) de l'orthographe de celui qui décline son identité et qui ignore parfois comment s'écrit correctement son patronyme. Ces erreurs peuvent aussi trouver leur origine dans un changement de domicile ou de région, citons pour mémoire la latinisation ou la romanisation de certains patronymes germaniques ou l'inverse. Relevons enfin les erreurs de copie.

Il semble donc bien nécessaire à un premier stade d'envisager le contrôle formel des patronymes. Ce contrôle devrait se faire initialement par une collation avec le document original. Toutefois, il n'est pas toujours possible, pour toutes sortes de raisons, de pratiquer ainsi. Et si la masse des documents augmente considérablement, il faudra passer à un stade où l'utilisation des méthodes du traitement automatique de l'Information s'avérera nécessaire.

Il y a toutefois un préalable à procéder de la sorte. A quoi se référer pour opérer un tel contrôle ? Ceci pose immédiatement la question de la constitution d'un *thesaurus* de patronymes.

La création et le développement d'un *thesaurus* peuvent être envisagés aisément par des procédés automatiques. Les records d'un bon *thesaurus* devraient comporter un certain nombre d'éléments : le patronyme lui-même sous la forme où on l'a trouvé, un élément chronologique (éventuellement les *termini* chronologiques où la forme est attestée), un élément géographique, éventuellement une référence archivistique générale identifiant le type de documents où le patronyme a été découvert. Un autre élément qui devrait figurer dans ce record et qui me semble essentiel est une forme standardisée du patronyme. Cette forme standardisée devrait permettre ultérieurement les regroupements de patronymes, les liaisons entre fichiers nominatifs et devrait aussi permettre certains modes rapides de consultation. Comment envisager cette forme standardisée ? Divers travaux ont été conduits à bon terme en ce domaine; nous pensons ici au système canadien SOUNDEX; en Belgique, les responsables du *Registre national* ont développé un fichier de patronymes à partir des patronymes existants afin de posséder le plus grand nombre possible de formes. Les procédures d'enrichissement d'un *thesaurus* ne semblent pas poser de problèmes. Des moyens automatiques peuvent être rapidement mis en oeuvre; le travail manuel qui est en l'occurrence un travail de décision, serait limité au contrôle effectué par le responsable du fichier des ajouts à y faire. On pourrait donc disposer ainsi assez rapidement d'un bon instrument de travail et de référence pour un contrôle formel et un contrôle de vraisemblance des formes patronymiques que l'on rencontre.

Ce type de *thesaurus* devrait, selon toute vraisemblance, être établi pour des aires géographiques qu'on limiterait en se basant sur des critères d'ordre administratif, d'ordre linguistique, d'ordre culturel afin d'aboutir

à la constitution d'un dictionnaire relativement uniforme d'une part, spécifique à la région; enfin, sa taille ne devrait pas être telle qu'elle empêcherait une exploitation rentable. Une autre question que l'on peut se poser à ce propos est une division chronologique des *thesauri*. Cela dépend évidemment du nombre de fichiers dont on dispose et de leur répartition chronologique.

En ce qui concerne les fichiers acquis par les Archivistes ou déposés par des Administrations, le problème du contrôle ne se pose guère au moment du dépôt puisqu'en principe, ces fichiers ont déjà fait l'objet de procédures semblables là où ils ont été créés.

L'utilité des *thesauri* ne se limite cependant pas uniquement au contrôle des formes patronymiques des fichiers. Ils peuvent aussi avoir un rôle important à jouer dans la recherche documentaire à travers les fichiers. Ils peuvent aussi être considérés comme des fichiers éclatés des fichiers originaux et exploités, statistiquement par exemple, en tant que tels.

CONTENU ET FORMAT DES FICHIERS

Le contenu des fichiers nominatifs peut varier considérablement. Ils peuvent être élémentaires ou fournir, au contraire, une abondance de renseignements dans chaque record. Il est évident qu'on aura toujours avantage lors de la création d'un fichier nominatif à le pourvoir du plus grand nombre possible d'informations.

Le type de format utilisé pour l'enregistrement d'un fichier nominatif est lui aussi fort important. La rubrique-vedette sera le patronyme (rappelons aussi le nom de lieu, le nom d'institution, etc...). Mais, il est évident que

les autres rubriques doivent être agencées en fonction du rôle documentaire du fichier et de ses éventuels enrichissements ou liaisons avec d'autres fichiers.

Du point de vue strictement informatique, nous nous trouvons dès lors devant une série de choix. Le record peut se présenter en zones de longueur fixe ou en zones de longueur variable. Le type de définitions des zones doit être choisi à bon escient. En ce qui concerne la notation même des informations, on aura le choix entre la mention en clair des renseignements, même sous une forme condensée, ou la mention sous forme d'un code. Les codes, nous le savons fort bien, sont nombreux; ils peuvent avoir toutes les provenances possibles, on peut les créer soi-même. Leur usage est une question de bon sens.

L'expérience nous conduirait à déclarer qu'il est préférable de choisir le système des zones de longueur variable et la mention en clair. Pourquoi ? Il semble bien que la solution des zones de longueur variable soit la meilleure parce qu'elle permet la plus grande souplesse de corrections et d'ajouts. La mention en clair est préférable, car il semble aussi que l'usage de codes ralentisse une partie du travail lors du décodage et de la présentation des résultats. En outre, le problème se pose aussi de l'unité et de la bonne application des codes à des ensembles parfois disparates. La question de la liaison entre codes se pose également. Les erreurs dans ce domaine sont évidemment fort possibles.

En ce qui concerne le contenu, un record étant rubriqué par un patronyme, toutes les autres informations contenues dans les différentes

zones du record doivent réellement servir d'informations. Ces informations sont des plus diverses, répétons-le. Il s'agit dès lors pour l'archiviste, que ce soit dans le cas de la création d'un fichier par lui-même ou que ce soit dans le cas d'un fichier qu'il a archivé, de posséder une fiche signalétique complète du fichier. Cette fiche aura à la fois un volet purement archivéconomique classique : origine, nature, destination, objet, ainsi qu'un volet purement informatique : origine, type de record, format, support et renseignements annexes, date de création, date de back-up, type de software, éventuellement existence d'un software particulier ou de plusieurs softwares. Il est également utile de savoir si un reformatage est possible et relativement aisé afin de pouvoir transformer un fichier en un format susceptible d'être traité par un software déterminé.

CONTROLE, UNIFORMISATION ET FUSION DE FICHIERS NOMINATIFS

Nous avons déjà envisagé la question du contrôle des patronymes d'un fichier. D'autres contrôles peuvent encore être effectués sur un fichier semblable. Il s'agit, par exemple, de contrôles internes visant à déceler et éliminer des informations aberrantes de type chronologique, géographique, contrôle des codes, etc... On peut également penser au contrôle formel des records eux-mêmes. La question à se poser dès lors est : la forme du record répond-elle exactement à la forme prévue ?

Nous ne pouvons que signaler ici l'existence de certains programmes qui, pour des fichiers de format déterminé, prennent en charge le contrôle formel et la correction de la plupart des erreurs rencontrées.

fichiers archivés par les Administrations d'Archives et en provenance des différentes branches de l'Administration générale. La spécificité de ces fichiers et, parfois, le nombre élevé des informations qu'ils contiennent peuvent constituer un obstacle à leur uniformisation et à leur fusion. Toutefois, après un dernier contrôle interne et formel, si les Archives en ont le moyen, on peut penser à une fusion générale de ce matériel documentaire.

CONSERVATION ET CONSULTATION DES FICHIERS NOMINATIFS

Le mode de conservation des fichiers nominatifs est identique à celui de tous les autres fichiers. Toutefois, tout comme en archivéconomie traditionnelle, on peut penser à éliminer éventuellement après les divers travaux de regroupement certains fichiers pouvant faire double emploi. La procédure conduisant à la destruction de ces fichiers tiendra évidemment compte des critères à observer : double emploi manifeste, fichier extrait d'un autre fichier, etc..., ainsi que des précautions administratives usuelles en cette matière : autorisation, date de péremption de l'usage des fichiers, etc...

La consultation des fichiers nominatifs peut s'envisager à divers niveaux, par plusieurs types différents d'utilisateurs et selon des modalités techniques à convenir. Cette liberté de consultation devra tenir compte, bien entendu, des règles en vigueur en ce qui concerne la confidentialité des fichiers ainsi que leur secret. Ces règles limitant la consultation seront celles de la législation en cours dans le pays où le fichier est consulté; éventuellement, il faudra tenir compte des restrictions à la consultation imposées par le déposant qu'il soit une personne ou une institution ou

fichiers archivés par les Administrations d'Archives et en provenance des différentes branches de l'Administration générale. La spécificité de ces fichiers et, parfois, le nombre élevé des informations qu'ils contiennent peuvent constituer un obstacle à leur uniformisation et à leur fusion. Toutefois, après un dernier contrôle interne et formel, si les Archives en ont le moyen, on peut penser à une fusion générale de ce matériel documentaire.

CONSERVATION ET CONSULTATION DES FICHIERS NOMINATIFS

Le mode de conservation des fichiers nominatifs est identique à celui de tous les autres fichiers. Toutefois, tout comme en archivéconomie traditionnelle, on peut penser à éliminer éventuellement après les divers travaux de regroupement certains fichiers pouvant faire double emploi. La procédure conduisant à la destruction de ces fichiers tiendra évidemment compte des critères à observer : double emploi manifeste, fichier extrait d'un autre fichier, etc..., ainsi que des précautions administratives usuelles en cette matière : autorisation, date de péremption de l'usage des fichiers, etc...

La consultation des fichiers nominatifs peut s'envisager à divers niveaux, par plusieurs types différents d'utilisateurs et selon des modalités techniques à convenir. Cette liberté de consultation devra tenir compte, bien entendu, des règles en vigueur en ce qui concerne la confidentialité des fichiers ainsi que leur secret. Ces règles limitant la consultation seront celles de la législation en cours dans le pays où le fichier est consulté; éventuellement, il faudra tenir compte des restrictions à la consultation imposées par le déposant qu'il soit une personne ou une institution ou

qu'il soit une Administration. A ce propos, il semble que l'on puisse suggérer qu'une étude sur les diverses règles et lois en vigueur dans les différents pays soient entreprises à l'avenir.

1.- LA CONSULTATION DANS LE CADRE MEME DES ARCHIVES

La consultation des fichiers nominatifs se fera comme la consultation de tout instrument de travail. En principe et dans le respect des règles de déontologie, cette consultation est libre et doit permettre une utilisation effective des fichiers. Les modalités techniques de cette consultation dépendent des moyens dont disposent les Archives : cela peut aller du listing correctement présenté, à la consultation par batch ou écran de visualisation. Le software utilisé peut être un software firme développé par les Archives, fourni par le déposant des fichiers ou acheté à une firme commerciale.

2.- LA CONSULTATION PAR DES PERSONNES EXTERIEURES A L'INSTITUTION

Cette consultation ne sera permise qu'en fonction des règles de confidentialité ou de secret générales édictées par la loi ou par les déposants. Au point de vue technique, la consultation peut se faire dans le cadre de l'Institution si celle-ci dispose du matériel nécessaire à cet effet et si elle dispose d'un service assurant la consultation des archives informatiques. Une autre possibilité est la mise à la disposition des utilisateurs de copies entières ou partielles. Nous ne nous étendrons pas plus sur ce sujet qui est en voie d'étude par M. Carroll.

LES ASPECTS TECHNIQUES DE LA CREATION, DE L'ARCHIVAGE ET DE LA CONSULTATION DES FICHIERS NOMINATIFS

Les considérations qui seront émises ici seront forcément limitées. En effet, l'équipement que requiert le traitement automatique de l'Information dépend de nombreux facteurs. Trois de ces facteurs semblent pourtant déterminant : la politique générale du pays concerné en ce qui concerne l'équipement de ses Administrations en matériel informatique, le volume des crédits qu'il peut allouer aux services d'Archives pour son équipement et l'utilisation de l'Informatique, le personnel qu'il veut bien consacrer à ces tâches,

En ce qui concerne la politique générale des Etats, il y a peu de choses à dire. Certains Etats laissent à leurs Administrations le choix du constructeur, certains l'orientent, d'autres l'imposent.

En ce qui concerne les crédits et le personnel, il s'agit d'une question propre à chaque Administration. Il est toutefois nécessaire de disposer d'un personnel minimum, difficile à évaluer, car il dépend du volume des travaux informatiques en cours. Toutefois, il devrait comporter au minimum un archiviste connaissant suffisamment l'Informatique que pour assurer la direction de cette équipe, un informaticien professionnel et du personnel d'exécution. La collaboration d'autres archivistes présentant leurs problèmes sera la bienvenue.

En ce qui concerne le matériel, que ce soit pour la création ou le traitement de fichiers dus à des archivistes, ou pour le traitement de fichiers reçus d'autres Administrations, la tendance actuelle étant de concevoir d'assez vastes fichiers, il sera nécessaire de disposer d'un

équipement relativement puissant. En effet, plus un équipement est puissant, plus il est diversifié dans ses différentes composantes et plus il est performant. Cela ne veut certainement pas dire que les Archives doivent se procurer cet équipement pour elles-mêmes. La solution la meilleure est l'utilisation des ressources en matériel d'un grand centre informatique dépendant d'une Administration centrale, d'une institution scientifique ou même d'une firme commerciale. Cela permet d'utiliser toutes les facettes d'un potentiel technique sans devoir soi-même en assurer les lourdes charges financières. Il n'en reste pas moins qu'un minimum de matériel informatique est nécessaire. Il s'agit bien entendu du matériel de saisie des données. La configuration de ce matériel dépend de la façon dont les Archives collaborent avec le centre principal d'Informatique. Le matériel de saisie peut comporter selon le cas : des perforatrices, des encodeuses sur cassette magnétique et même des unités OCR. C'est un minimum. A l'autre bout, si les Archives disposent de suffisamment de crédits, on peut envisager l'installation de terminaux légers de consultation des fichiers : imprimantes, écrans de visualisation, etc... Il est essentiel en ce domaine de proportionner l'effort aux buts à atteindre.

Un dernier problème d'ordre technique est celui du développement et de l'utilisation des softwares. Plusieurs solutions sont encore ici possibles : développement de softwares par les Archives elles-mêmes pour ses propres fichiers ainsi que pour les fichiers qu'elles conservent, utilisation des softwares prévus pour les fichiers déposés, achat ou location de packages (type STAIRS ou GOLEM). Toutes les solutions sont acceptables de mon point de vue. Les avantages et les

inconvénients s'équilibrent. Toutefois, même si l'utilisation des logiciels commerciaux ou packages offrent des possibilités énormes, il est intéressant de développer dans le cadre des Archives, en vue d'aider les utilisateurs dans la consultation des fichiers, des programmes de consultation, de calculs statistiques, etc...

CONCLUSION

Nous espérons que ces quelques réflexions sur les fichiers nominatifs aideront à promouvoir et à préciser les questions relatives à ce matériel documentaire.

J. PIEYNS,
Archives de l'Etat
(Belgique)