ANNALES BRETAGNE PAYS DE L'OUEST

Annales
de Bretagne et des Pays de l'Ouest
Anjou. Maine. Poitou-Charente. Touraine

108-4 | 2001
La Chambre des comptes de Bretagne

Base de données prosopographique : de la conception à la réalisation

Bernard Courbot

revues.org

Éditeur
Presses universitaires de Rennes

Édition électronique
URL : http://abpo.revues.org/1671
DOI : 10.4000/abpo.1671
ISBN : 978-2-7535-1484-3
ISSN : 2108-6443

Édition imprimée
Date de publication : 20 décembre 2001
Pagination : 19-29
ISBN : 978-2-86847-674-6
ISSN : 0399-0826

Référence électronique
Bernard Courbot, « Base de données prosopographique : de la conception à la réalisation », *Annales de Bretagne et des Pays de l'Ouest* [En ligne], 108-4 | 2001, mis en ligne le 20 décembre 2003, consulté le 04 octobre 2016. URL : http://abpo.revues.org/1671 ; DOI : 10.4000/abpo.1671

Base de données prosopographique : de la conception à la réalisation

Bernard COURBOT
enseignant en statistiques,
CRHMA, Université de Nantes, IRMAR, Université de Rennes 1

Créer une base de données regroupant des informations sur des individus très variés et sur une longue période pose de délicats problèmes qui ne peuvent être résolus que par une étroite collaboration entre le programmateur et l'historien ; ceci exige de longs efforts réciproques pour traduire en langage binaire informations et aspirations formulées en langage littéraire. De même, présenter en quelques pages la conception d'une base de données prosopographique nécessite des compromis ; il ne saurait être question d'adopter un langage qui soit commun à l'historien peu averti des contraintes de l'outil informatique et à l'informaticien ignorant des contraintes de la collecte d'informations historiques. Qu'il me soit ainsi pardonné quelques passages trop techniques pour certains et trop simplistes pour d'autres.

Intentions

Le but assigné à cette base de données est de regrouper sous une structure unique les informations collectées par divers historiens de l'équipe de recherche, avec le minimum de perte d'information possible, et ceci avec trois buts principaux :
• éditer la liste des officiers connus, avec le maximum de renseignements familiaux, de carrière… sous forme de dictionnaire ;
• permettre aux chercheurs une consultation rapide sur le Web de ces mêmes renseignements ;
• étudier statistiquement la structure et les passages de pouvoir au sein de la Chambre des Comptes.

Cette universalité de buts ainsi que la particularité des données à stocker comportent des contraintes spécifiques :
• variabilité des renseignements par chercheur : chaque historien ayant son idée propre de la fiche informative, il était nécessaire d'uniformiser au

maximum ce travail de collecte afin que la transcription informatique ne soit pas source de perte d'informations ou d'erreurs ;
• variabilité des informations dans le temps : la longue période d'exercice de la Chambre implique des variations dans les titres, intitulés, dénominations, patronymes…
• incomplétude des données : en particulier dans les premiers temps, nombre d'informations ne sont pas disponibles ou sont entachées d'incertitude ; il est essentiel que la base reflète ces zones d'ombre et ne soit pas encombrée de champs vides.

Ces considérations et ces particularités furent déterminantes pour la conception de la base. Le choix d'Access 2000 de chez Microsoft est principalement d'ordre pratique : d'un prix abordable, universellement répandu et d'usage enseigné à l'Université, ce logiciel offre en outre dans sa dernière version la possibilité de créer très aisément des pages Web. Avant de préciser comment ces difficultés ont été gérées, quelques explications techniques s'imposent pour le néophyte en informatique, que pourra sauter tout lecteur tant soit peu au fait des structures de données.

Base de données : monotable *vs* relationnelle

	Champs		
Enregistrements :	Champ 1	Champ 2	…
Enregistrement A			
Enregistrement B			
…			

Les bases de données manuelles regroupent des fiches, chacune regroupant des informations sur un individu donné. Sur le plan informatique, la base de données monotable reflète cette structure : elle est formée d'« enregistrements » (les individus) et de « champs » (les caractéristiques). Une telle structure a l'avantage de la simplicité mais devient très vite inutilisable dès que les données deviennent importantes ou complexes.

Un simple exemple éclairera ces propos. Si l'on désire entrer les informations d'ordre familial, aucune difficulté ne se posera pour les champs « père » et « mère » puisque la correspondance entre personne et parents est « un à un ». Par contre, l'entrée des frères et sœurs pose problème : puisqu'un champ ne doit contenir qu'un seul individu pour être utilisable, le concepteur devra créer des champs multiples (par exemple « frère 1 », « frère 2 »…) au risque de ne pas en prévoir suffisamment et de perdre ainsi des informations. De plus si deux frères doivent figurer dans la base, la liste des frères et sœurs figurera en doublon pour chacun, la complexifiant et l'alourdissant et en rendant délicate voire impossible l'exploitation statistique. La solution réside en une conception de type relationnel. Reprenons l'exemple ci-dessus avec cette nouvelle structure. Les individus figurent

tous dans une table, sans liste de frères et sœurs ; des relations sont créées entre un individu et ses parents de sorte que pour obtenir la liste désirée, il suffit de demander quels enregistrements ont mêmes pères et mères. La question des doublons est ainsi résolue ainsi que celle de la limitation du nombre de champs puisque le nombre de relations n'est pas limité.

Précisons les notions essentielles des bases relationnelles :

• elle est structurée en plusieurs **tables** qui regroupent des informations homogènes ; par exemple une table pour les officiers, une table pour les offices, une table pour tous les individus rencontrés…

• deux tables A et B peuvent être liées entre elles par des **relations**, qui peuvent être de plusieurs types :

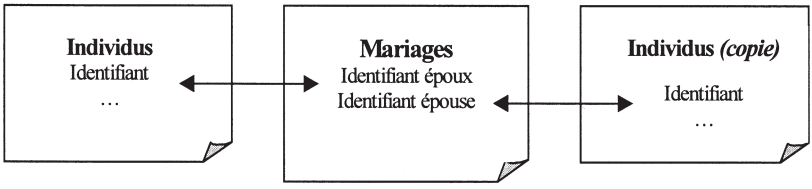
– 1 ↔ 1 : à chaque enregistrement de A correspond au plus un enregistrement de B (individu ↔ officier par exemple) ;

– 1 → ∞ : à un enregistrement de A peut correspondre un nombre quelconque d'enregistrements de B (officier → offices exercés) ;

– ∞ ↔ ∞ : tout enregistrement de A peut être relié à un nombre quelconque d'enregistrements de B et vice-versa (offices ↔ officiers).

Les **requêtes** permettent d'obtenir des informations précises et de créer les structures plus complexes nécessaires à l'exploitation de la base ; par exemple la question « quelles furent les épouses successives de tel personnage ? » se résout par la requête fournissant la structure suivante :

Figure n° 1



En fixant un individu dans la table de gauche, la requête fournira la liste des conjointes lue dans la table de droite.

Cette liberté de conception a bien sûr un prix ! L'entrée des données doit être, elle aussi, structurée. Pour qu'une relation de mariage puisse être construite, il est évidemment nécessaire que les deux conjoints figurent au préalable dans la table des individus. Une des tâches les plus délicates rencontrées lors de la conception fut la création de **formulaires** qui, tout en respectant l'ordre d'entrée des informations, soient suffisamment clairs pour faciliter cette tâche fastidieuse. Détaillons quelque peu ces notions dans notre cadre prosopographique.

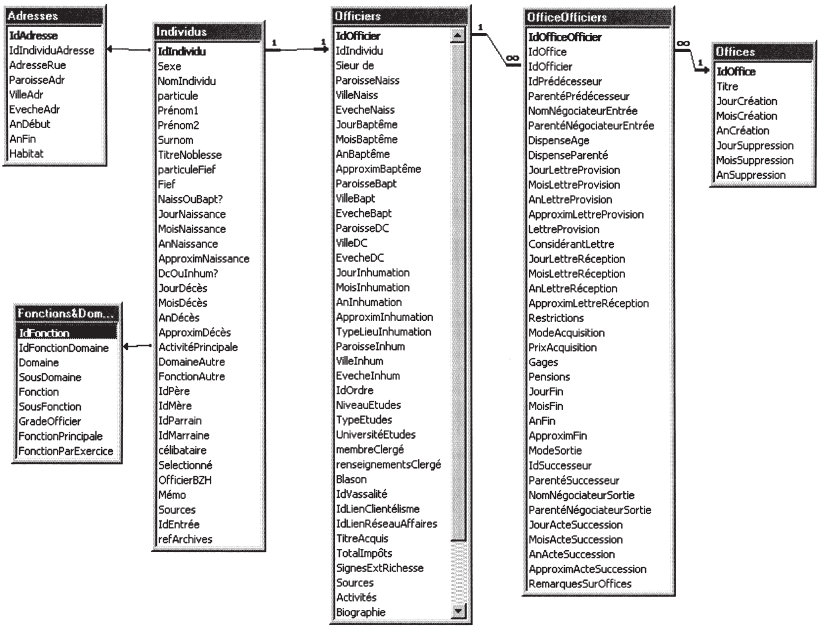
Les tables

Comme ceci a été dit précédemment, les informations sont regroupées en tables spécifiques. Pour des raisons techniques qui seront précisées ultérieurement, notre base comporte un important nombre de tables. Nous ne considèrerons ici que les plus importantes.

Les principales tables

La table **Individus** regroupe les informations de base (état-civil, fonction principale,…) sur tous les personnages liés à la Chambre (officiers eux-mêmes, apparentés, relations d'affaires,…) rencontrés lors de la collecte des données. Sa structure sera précisée dans la suite.

Figure n° 2



La table **Officiers** stocke les informations complémentaires sur les seuls officiers de la Chambre. Elle est reliée à d'autres tables annexes permettant d'entrer sans limitation de nombre les adresses successives, les fonctions exercées chronologiquement ; tout l'itinéraire, toute sa carrière peuvent ainsi être reconstruits. La table **Offices** regroupe les informations sur chaque office indépendamment des officiers ayant occupé ce poste (date de création,…). Ces deux dernières tables sont reliées en liaison ∞↔∞ par l'intermédiaire de la table annexe **OfficeOfficiers** qui détaille chaque office

exercé par les officiers (dates de début et fin, mode d'acquisition et de sortie, parenté avec le prédécesseur…). Ce choix de liaison permet de stocker un nombre non limité de charges d'office pour un même officier.

Dans la figure ci-dessus, une liaison 1→∞ entre la table **Individus** et la table **Adresses** permet d'entrer les adresses successives d'un individu sans limitation de nombre. D'autres tables comme celle **Fonctions&Domaines** de cette même figure permettent de modifier *a posteriori* les intitulés des fonctions, domaines d'activité, etc. sans devoir modifier l'ensemble de la base. En effet lors de l'entrée des informations, l'opérateur doit choisir une valeur parmi un ensemble de possibilités. Cet ensemble doit parfois être réajusté (modification dans le temps d'un intitulé par exemple) et il est alors essentiel de minimiser le travail de re-programmation qui en découle. Il est plus aisé et moins risqué de modifier une simple table qu'une série de formulaires ou d'états. Ceci explique l'abondance de tables annexes dans la base de données.

La table Individus : une table généalogique

La table **Individus** a un statut quelque peu particulier de par sa structure généalogique. En effet, pour reconstruire un arbre généalogique, les relations parentales et matrimoniales suffisent. Toute autre relation découle de ces deux fondamentales ; il suffit de construire les requêtes pertinentes qui, bien évidemment, se compliquent lors de recherches de parentés éloignées. La Figure n° 1 donnait déjà un exemple de telle requête pour la recherche des épouses d'un individu. Les figures n° 3 et 4 ci-dessous donnent des exemples où la table de droite fournit respectivement les frères et sœurs (même père, même mère) et le grand père paternel (père du père).

Figure n° 3

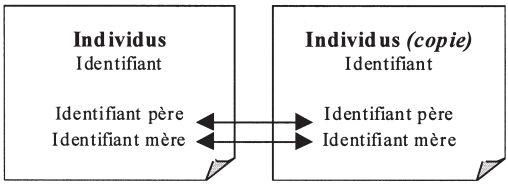
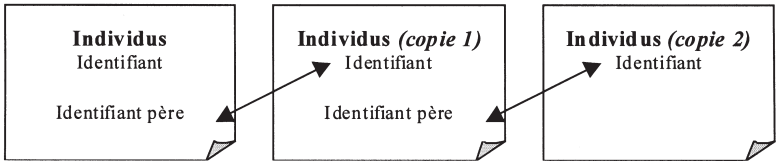


Figure n° 4



Il n'y a ainsi aucune perte d'information sur les divers parents connus des officiers et le fichier peut conserver une taille raisonnable.

Choix des champs

Généralités

Le logiciel offre un grand éventail de structures des champs. Un choix pertinent est indispensable pour assurer une saisie rapide et sûre des données et pour permettre les études statistiques ultérieures. On distingue principalement les champs de types :

• booléen : deux valeurs possibles (oui/non) ; cela se traduit le plus souvent dans un formulaire par une case à cocher (l'individu est ou n'est pas officier de la Chambre) ;

• numérique : tout nombre avec de multiples choix de longueur et présentation ;

•date : ce type, très pertinent pour des bases classiques, l'est beaucoup moins ici de par la spécificité historique et de l'imprécision fréquente des données ; ce point sera précisé plus loin ;

•texte : tout ensemble d'au plus 255 caractères ;

•mémo : tout texte de grande longueur, mais sans possibilité de tri ou de sélection ; ce type se prête bien aux remarques, commentaires additionnels… sur lesquels aucune analyse n'est prévue ;

•lien OLE : toute sorte d'objets reconnus par le système, en particulier dessins (blasons), photographies (demeures, portraits, etc.).

Les entrées de texte se font dès que possible par choix dans une liste déroulante. Par exemple, l'opérateur n'a pas à taper la particule d'un patronyme mais la choisit dans une liste préétablie ; cette technique permet de limiter les risques d'erreurs par dysorthographie et hétérogénéité des termes. Il en résulte de sérieuses contraintes lors de la création de la base puisqu'il faut choisir toutes les possibilités dignes d'intérêt qui seront offertes à l'opérateur, par exemple la liste des divers titres de noblesse pour tous les individus. Cette méthode peut sembler contraignante à l'historien néophyte en informatique puisqu'elle bride la liberté d'entrée de données ; cependant comment imaginer sélectionner selon un critère donné si les divers opérateurs ont utilisé des graphies variées pour le même objet ?

Problèmes plus spécifiques

Le problème de l'entrée des dates vient d'être évoqué. En particulier pour les cas les plus anciens, nombre d'informations manquent de précision. Considérons l'exemple de l'entrée en fonction dans l'office. Il arrive assez fréquemment que la date exacte ne soit pas connue ; seules sont mentionnées quelques dates d'interventions. Ainsi, dans ce cas, la seule information rigoureuse est une prise de fonction antérieure à la première intervention retrouvée. Pour ne pas perdre d'information, l'entrée d'un tel champ date est précédée d'un champ « précision » comportant les choix :

« = », « ≤ », « ≥ » et « ≈ », selon que la date est exacte, antérieure, postérieure ou approximative. Il sera impératif de tenir compte de ce degré de précision lors de l'étude statistique des données mais l'utilisation d'un champ de type mémo (dans lequel aurait été tapé en toutes lettres les indications sur la première intervention) aurait empêché cette analyse. Un champ **Commentaires** de type mémo permet de fournir ces renseignements qui pourront par exemple figurer dans le dictionnaire.

Un autre problème, plus délicat encore, est celui des graphies multiples des patronymes et des homonymies. Inutile sur ce point d'espérer que l'ordinateur résoudra automatiquement ce que le cerveau humain ne sait pas faire! Une même personne nommée Harrouys ici et Harruys là sera comptée comme deux individus distincts. La plus grande rigueur s'impose aux opérateurs et des règles précises doivent être édictées par les historiens pour limiter voire éliminer ces erreurs. Lors de l'entrée des noms patronymiques, une indexation appropriée interdit la saisie de doublons : un message d'erreur signale à l'opérateur qu'un individu ayant même nom et mêmes premier et second prénom figure déjà dans la table. Les cas d'homonymie se résolvent alors en adjoignant un numéro au second prénom, qui pourra éventuellement être retiré lors de l'édition du dictionnaire. Le nom de l'opérateur ainsi que la date de saisie sont enregistrés pour chaque individu afin de permettre la résolution rapide des problèmes qui pourraient se présenter *a posteriori*.

Mentionnons enfin les difficultés soulevées par les dénominations des fonctions exercées par les individus étudiés. Afin d'éviter les fluctuations d'entrées des opérateurs, une table a été construite, contenant quatre champs : le domaine d'activité avec le plus souvent un sous-domaine (par exemple : administration–province–), et pour chaque groupe une fonction avec éventuellement une sous-fonction (par exemple : gouverneur–ville–). Cette division permet de faciliter le choix des activités en sélectionnant le domaine puis la fonction parmi les possibilités liées au domaine choisi. La construction de cette table fut délicate par suite de la variété et des variations dans le temps des fonctions. Cette table est bien sûr susceptible de modifications et une profession rencontrée exceptionnellement peut toujours être entrée manuellement ; cependant cette situation doit rester rarissime pour permettre une exploitation statistique rationnelle des données.

Entrée des données

Les formulaires

Les formulaires permettent de faciliter la saisie des données qui sont dans le cas présent nombreuses et relativement délicates. Dans ce but, une présentation avec force boutons d'actions, menus déroulants, cases à cocher a été élaborée pour accélérer et sécuriser au maximum les entrées.

La figure n° 5 présente comme exemple le formulaire d'entrée des individus sur lequel on retrouve les diverses notions précédemment exposées.

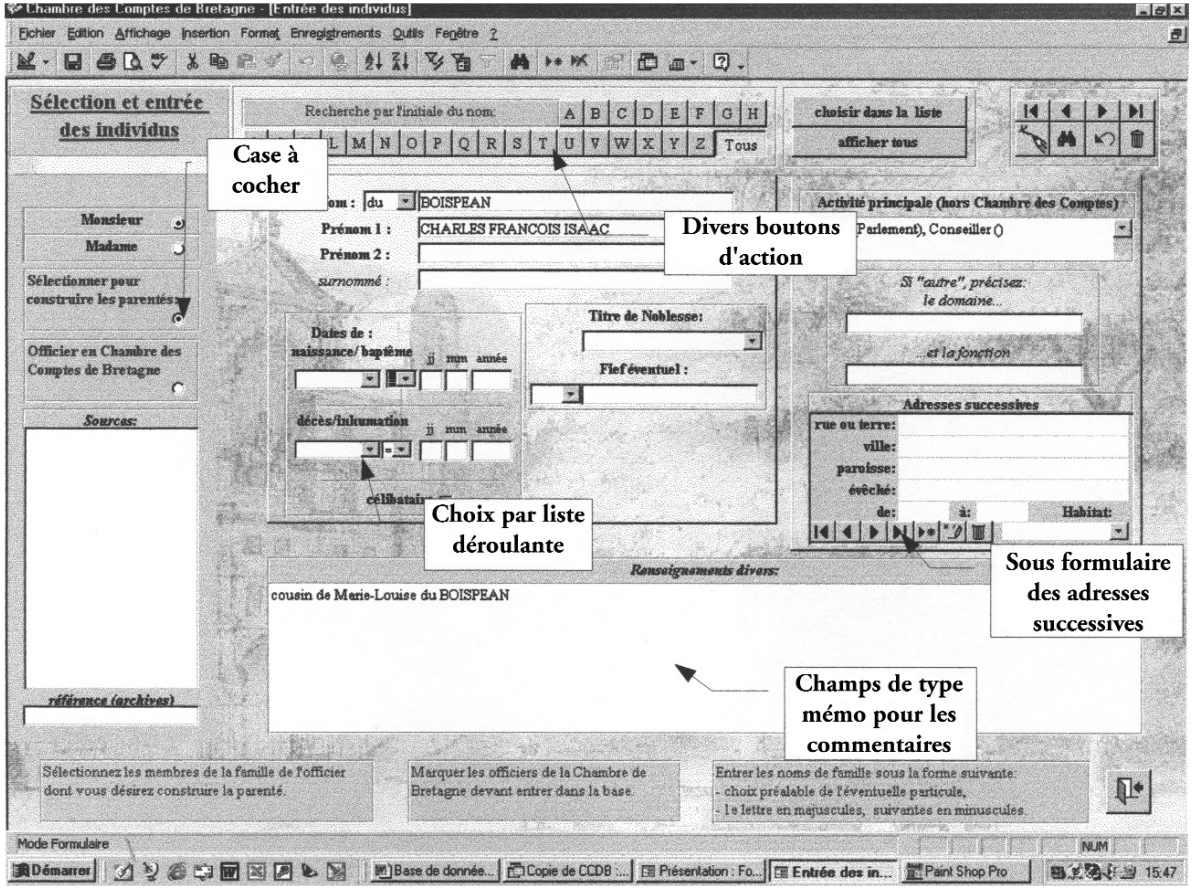

Chambre des Comptes de Bretagne - [Entrée des individus]
Fichier Edition Affichage Insertion Format Enregistrements Outils Fenêtre ?
Sélection et entrée des individus
Recherche par l'initiale du nom:
A B C D E F G H
L M N O P Q R S T U V W X Y Z Tous
choisir dans la liste
afficher tous
Case à cocher
Divers boutons d'action
Monsieur
Madame
Sélectionner pour construire les parentés
Officier en Chambre des Comptes de Bretagne
Sources:
référence (archives)
Prénom 1 : CHARLES FRANCOIS ISAAC
Prénom 2 :
surnommé :
Activité principale (hors Chambre des Comptes)
Parlement), Conseiller ()
Si "autre", précisez: le domaine...
...et la fonction
Titre de Noblesse:
Fief éventuel :
Dates de : naissance/baptême
décès/inhumation
célibataire
Adresses successives
rue ou terre:
ville:
paroisse:
évêché:
de:
à:
Habitat:
Choix par liste déroulante
Sous formulaire des adresses successives
Renseignements divers:
cousin de Marie-Louise du BOISPEAN
Champs de type mémo pour les commentaires
Sélectionnez les membres de la famille de l'officier dont vous désirez construire la parenté.
Marquer les officiers de la Chambre de Bretagne devant entrer dans la base.
Entrer les noms de famille sous la forme suivante:
- choix préalable de l'éventuelle particule,
- le lettre en majuscules, suivantes en minuscules.
Mode Formulaire
NUM
Démarrer
Base de donnée...
Copie de CCDB ...
Présentation : Fo...
Entrée des in...
Paint Shop Pro
15:47

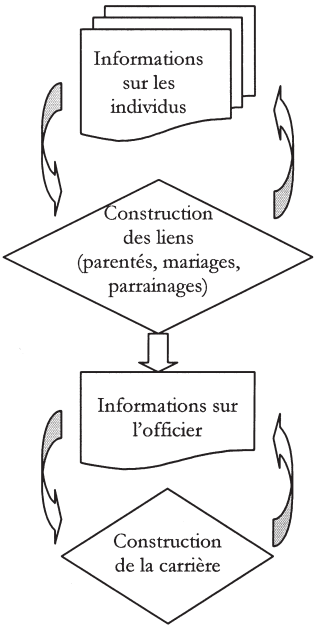
Figure n° 5

Stratégie d'entrée des informations sur un officier

Afin d'accélérer et de sécuriser la saisie des données, un minimum d'ordre et de rigueur s'impose, tant à l'historien lors de la collecte des informations qu'aux opérateurs sur machine.

Considérons le cas d'un officier X. Quel sera l'ordre à respecter ?

Figure n° 6



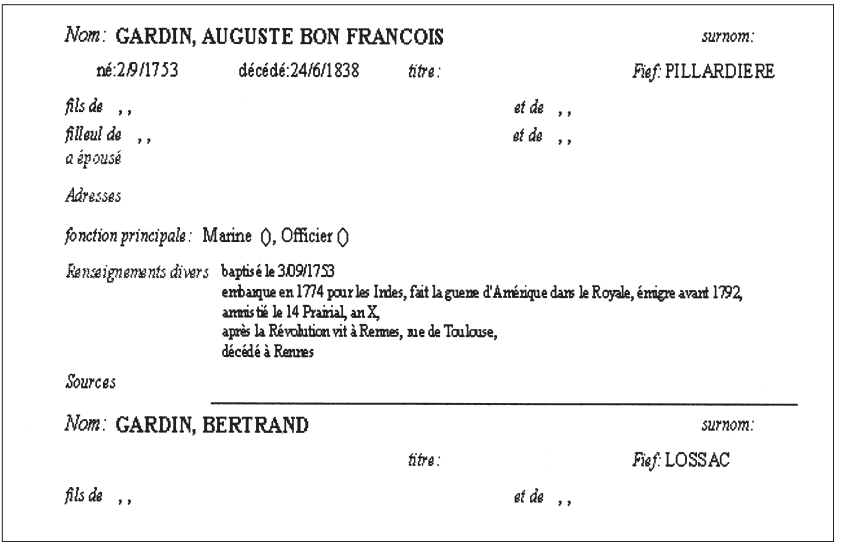
1- Collecte des informations : les données issues des archives ou d'ouvrages antérieurs sont rassemblées sur des fiches selon des consignes préétablies. Doivent y figurer les renseignements sur l'officier, sa carrière et sur toute personne mentionnée.
2- Entrée de tous les individus concernés par X : ceux-ci sont soit sélectionnés s'ils figurent déjà dans la base (rappelons que le logiciel interdit tout doublon donc signalera ce problème le cas échéant), soit entrés s'ils sont nouveaux. Les informations (*cf.* Figure n° 5) sont dans ce cas minimales et bien sûr peuvent (et sont souvent) incomplètes. Cette table comportera à terme un très grand nombre de personnages, aussi de petits utilitaires permettent d'accélérer la recherche dans la liste sur des critères alphabétiques ou d'époque.
3- Construction des relations maritales entre les précédents individus.
4- Construction des relations de filiation.
5- Construction des relations de parrainage.
6- Entrée des informations sur l'officier X : compléments d'état civil, construction de sa carrière dans et hors de la Chambre des comptes, documents iconographiques le cas échéant.

Ainsi saisies, les informations sont stockées dans les tables convenables. Une phase de vérification est alors inévitable, soit manuelle en éditant et imprimant les données, soit logicielle par recherche de doublons ou de certaines incohérences. La base sera alors prête à être interrogée.

Interrogations de la base

Une interrogation de la base consiste à construire la requête correspondante, qui peut être élémentaire en ne portant que sur une seule table, soit très sophistiquée selon la question posée. Le résultat est donné à l'écran sous forme d'une liste de réponses, la plupart du temps délicate à lire et à interpréter.

Figure n° 7



Nom: GARDIN, AUGUSTE BON FRANCOIS *surnom:*

né:2/9/1753 décédé:24/6/1838 *titre:* *Fief:* PILLARDIERE

fils de , , *et de* , ,

filleul de , , *et de* , ,

a épousé

Adresses

fonction principale: Marine (), Officier ()

Renseignements divers baptisé le 3/09/1753
embarque en 1774 pour les Indes, fait la guerre d'Amérique dans le Royale, émigre avant 1792,
amnistié le 14 Prairial, an X,
après la Révolution vit à Rennes, rue de Toulouse,
décédé à Rennes

Sources

Nom: GARDIN, BERTRAND *surnom:*

titre: *Fief:* LOSSAC

fils de , , *et de* , ,

Des **états** (figure n° 7) permettent une présentation claire et agréable ; ils se présentent comme des formulaires destinés à l'impression, avec éventuellement documents iconographiques. Ils peuvent être également très aisément exportés dans un traitement de textes où ils pourront être éventuellement modifiés et formatés à des fins de publication. L'interrogation pourra se faire *via* Internet sous forme de pages Web de format classique. La création de requêtes nécessite une bonne compréhension de la structure de la base et des outils informatiques. Ainsi, il est illusoire d'espérer pour un néophyte obtenir facilement une réponse fiable à une question quelque peu élaborée dans une base de cette importance. Les interrogations les plus élémentaires seront d'abord stockées, et progressivement enrichies selon les demandes plus complexes des chercheurs tant sur machine que sur le Web.

Quelles sont les questions auxquelles pourra répondre la base ? De par la volonté de stocker le maximum d'informations, la quantité et la qualité des réponses sont vastes. Bien évidemment, les renseignements complets sur tel officier seront disponibles de même que, pour un office donné, la liste chronologique des officiers connus ayant occupé cette fonction. Mais il est aisé d'imaginer un faisceau de renseignements sur le mode de transmission de ces charges : combien d'officiers ont succédé à un parent, quels offices comptaient le plus de nobles selon les époques, etc. La seule limitation aux questions est le temps disponible pour programmer les requêtes, la mémoire machine ainsi que l'exactitude et la complétude des données. Ce dernier point est essentiel. Tout logiciel n'a comme fonction que d'automatiser et d'accélérer (ô combien !) une recherche ou un travail répétitif ; il est clair que déterminer manuellement la proportion d'officiers nés

hors de la Bretagne représente une tâche éprouvante alors qu'une simple requête donne quasi-instantanément la réponse. L'analyse informatique et statistique des données n'aura de sens qu'après une collecte et une saisie rigoureuse des informations. De même, toute question avant d'être soumise doit être clairement et rigoureusement réfléchie pour pouvoir être transcrite en langage logique sous forme de requête. Cette confrontation entre deux mondes au départ si distants a été et est encore permanente pendant toute la période de création logicielle. Il n'est pas aisé pour un historien de devoir transcrire en binaire des notions perçues comme multiples, il n'est pas facile pour un informaticien de percevoir toutes les finesses de l'étude prosopographique. Ce n'est pas une des moindres qualités de la réalisation de cette base que d'avoir fait se côtoyer deux disciplines aux caractéristiques et méthodes très différentes et de les avoir, je l'espère, enrichies mutuellement.

RESUME

La construction d'une base de données informatisée à caractère prosopographique soulève des difficultés très particulières, qui ne peuvent être résolues que par une étroite coopération entre informaticiens et historiens. Le présent article expose les éléments qui ont conduit au choix d'une base relationnelle comportant un noyau de type généalogique ainsi que les solutions adoptées pour les problèmes spécifiques de données manquantes ou incomplètes. L'accent est mis sur les techniques mises en œuvre pour tenter de concilier les modes de pensée et de travail des historiens avec la réduction indispensable à l'informatisation de ce travail.

ABSTRACT

The construction of a computerized prosopographical data base raises very particular difficulties, which can be solved only by a narrow cooperation among computer and history specialists. The present paper explains the elements which led to the choice of a relational base containing a genealogical type core as well as the solutions adopted for the specific problems of missing or incomplete data. The stress is laid upon techniques used to try to reconcile the ways of thinking and work of the historians with the reduction indispensable to data computerizations.