

Les fichiers-réseaux de population : Un retour à l'individualité

Gérard Bouchard*

À la lumière d'une expérience concrète qui s'étale sur plus de quinze ans, ce texte vise à présenter : a) une conception des banques de données historiques; b) une réflexion critique sur la place qu'elles ont occupée jusqu'ici; et c) des éléments d'un plaidoyer en faveur de leur développement.

I — LES FICHIERS-RÉSEAUX DE POPULATION COMME BANQUES DE DONNÉES

En premier lieu, il est important de rappeler qu'en lui-même, le principe de la banque de données informatisée n'est pas vraiment synonyme de rupture. On peut dire qu'il est au contraire en continuité avec le métier d'historien, dans le prolongement direct de ses méthodes. De tout temps, en effet, le chercheur isolé a constitué son propre fichier (format, classement des données, etc.), en a élaboré un système de gestion (ajouts, corrections, retraits) et une méthode d'exploitation (comptage, repérage, transcription des données). Ces procédés accusent des lacunes importantes, soit sur le plan de l'organisation de la recherche, soit sur le plan de l'analyse elle-même. Signalons en particulier :

- Le caractère partiel des reconstitutions.
- Un manque de transparence en vertu duquel il est difficile, voire impossible, de connaître, de contrôler et de critiquer les procédés de traitement et d'analyse.
- Le caractère non systématique des méthodes employées d'un chercheur à l'autre.
- Le gaspillage entraîné par l'impossibilité fréquente de transférer et de réutiliser à volonté les *corpus* constitués dans un cadre artisanal, en accord avec la tradition du métier¹.

Le propre de la banque de données informatisée ne réside pas dans la nature des opérations et des fonctions inhérentes à la construction, à la gestion et à l'exploitation du fichier, puisqu'en réalité, elle n'y introduit aucun changement de nature. Il réside plutôt dans les modalités d'exécution de ces opérations, auxquelles il ajoute cinq caractères essentiels : la permanence, la systématisation, la transparence, l'accès universel et la vitesse de traitement. Tous ces traits sont résumés dans la notion d'infrastructure informatique.

Les fichiers de population représentent un cas particulier de banque de données historiques. On peut les définir par rapport à cinq composantes :

* SOREP, Université du Québec à Chicoutimi.

1. On notera que ce facteur essentiel n'est jamais pris en considération lorsqu'on fait le procès des banques de données informatisées en termes de coûts-bénéfices.

- Des informations nominatives (individuelles, familiales, généalogiques ou autres).
- Un système de jumelage automatique des données.
- Un système de gestion informatique (Data Base Management System; à l'avenir : D.B.M.S.).
- Des instruments d'évaluation critique des données et des opérations.
- Des programmes d'exploitation flexibles, assortis d'interfaces qui rendent la tâche le plus facile possible aux usagers qui ne sont pas des professionnels de l'informatique.

Le fichier-réseau BALSAC, construit par notre centre de recherche, a été conçu selon ce modèle. Contenant près de 9 000 dossiers nominatifs, il recouvre présentement la région du Saguenay et une partie de la région de Charlevoix (Annexe 1); il est géré par le biais du D.B.M.S. INGRES; il permet des analyses à l'échelle des individus, des familles ou des descendances; et il assure l'accès à des données dites sectorielles à caractère démographique, culturel, social, médical, etc. (Annexe 2). Du point de vue de l'exploitation, il donne lieu à des analyses dans deux volets, l'un social et historique, l'autre génétique et médical². Enfin, une interface a été mise au point, INGRID, qui permet à l'utilisateur non professionnel d'effectuer lui-même ses requêtes après une brève initiation.

II — PERTINENCE DES BANQUES DE DONNÉES HISTORIQUES

De diverses façons, on peut montrer que le développement de banques de données informatisées en histoire n'obéissait pas uniquement à la volonté de suivre l'évolution technologique; il correspondait aussi à la nécessité de prolonger des changements profonds, déjà en cours dans la pratique du métier d'historien. À partir de la décennie 1960-1970, en effet, dans presque tous les domaines de la science historique, on observe une prise de conscience à l'endroit de la dimension sociale qui entoure d'une façon plus ou moins immédiate les péripéties ou les réalités étudiées, même les plus événementielles et les plus individuelles³. Mais qui dit social ou collectif dit souvent multitude, c'est-à-dire, pour l'historien, extension de l'aire d'observation et, presque nécessairement, accroissement du volume des relevés, intégration du quantitatif, gestion des grands nombres. Il faut, en outre, souvent reconstruire des sources trop partielles, défaillantes ou tout simplement inexistantes. En ce sens, la banque de données informatisées est un héritage naturel de l'histoire sociale. Or, il est important de rappeler que celle-ci, au moment de sa conception, ne disposait pas de l'infrastructure technique et méthodologique requise pour l'atteinte de ses objectifs fondamentaux. On peut montrer aussi que les fichiers de population sont une conséquence logique de développements méthodologiques et techniques importants survenus entre 1955 et 1975. C'est en premier lieu la méthode de reconstitution des

2. Pour des exposés plus détaillés : G. Bouchard *et alii*, 1980; G. Bouchard, R. Roy, B. Casgrain, 1985; G. Bouchard, 1987; Rapports annuels de SOREP. À certains égards, le volet génétique, créé en 1979 (alors que le projet d'ensemble a été mis en marche en 1972), constitue une sorte de boni, puisqu'à l'origine, la banque devait d'abord appuyer des enquêtes à caractère historique et social. Mentionnons aussi que l'aire géographique couverte par la banque a été sans cesse étendue. L'objectif actuel vise tout l'est du Québec, du XVII^e siècle jusqu'à aujourd'hui. Quelques lecteurs ou lectrices aimeront peut-être aussi apprendre que ces travaux de collecte, de saisie et d'élaboration des données ont été financés dans une proportion de 15 p. 100 environ par des subventions provenant du C.R.S.H. (Ottawa) ou du Fonds F.C.A.R. (Québec). Cette partie de nos travaux a donc pesé assez peu sur les budgets de la recherche historique proprement dite au Québec et au Canada.

3. De ce point de vue, l'un des meilleurs exemples reste sans doute l'ouvrage précurseur que Lucien Fèbvre (1952) a consacré à Martin Luther.

familles dont l'exploitation a été d'abord confinée à la démographie historique alors qu'elle contenait virtuellement les principaux éléments du fichier-réseau. Il lui manquait l'automatisation, le croisement avec des données non démographiques et l'extension des aires supraparoissiales. En verticale, allaient aussi s'ajouter les filiations généalogiques, déjà présentes, du reste, dans l'enquête pionnière de Louis Henry (1956) sur les anciennes familles genevoises. Enfin, du côté de l'informatique, le développement de logiciels de gestion très sophistiqués (D.B.M.S.), hiérarchiques, relationnels et autres, allait donner forme concrètement à l'idée de fichier-réseau de population en ouvrant la voie à l'intégration de fichiers sectoriels et à l'observation des itinéraires individuels et familiaux.

De la conjugaison de ces apports sont nées des infrastructures scientifiques qui constituent une sorte d'archivistique de deuxième génération, à mi-chemin entre la donnée brute et l'analyse. C'est ici l'exemple concret d'une idée fondamentale, surgie directement de l'évolution de la science historique et poussée jusqu'à son terme. On nous permettra d'évoquer ici certains aspects spectaculaires de ce qui est désormais acquis : a) la possibilité de construire et d'analyser en quelques heures quelques milliers de généalogies, quelques dizaines de milliers d'itinéraires individuels ou familiaux; b) l'établissement d'un cadre souple et efficace pour l'enquête véritablement longitudinale; c) des moyens de croisement de données nominatives qui fournissent à l'interdisciplinarité une technologie originale; d) des possibilités d'interrogation et d'observation qui dépassent l'échelle de la paroisse et de la ville; e) des conditions de traitement des données qui permettent à volonté les simulations, les comparaisons et les reprises.

Il est utile également de rappeler que les innovations reliées à l'informatisation et à la construction des banques de données ne sont pas étrangères au renouvellement du savoir en histoire sociale durant les quinze dernières années. M'en tenant aux thèmes qui me sont les plus familiers, je mentionnerai : la mobilité des populations préindustrielles, rurales aussi bien qu'urbaines; le calendrier et la mesure précise des événements démographiques; la formation et la reproduction de la famille; les itinéraires individuels dans une perspective intergénérationnelle; les rapports sociaux liés au travail ouvrier; la critique générale du modèle de la modernisation et de la synchronisation qu'il postule entre les évolutions économiques, sociales et culturelles.

III — APERÇUS CRITIQUES

Cela dit, il est paradoxal de constater que, dans la conjoncture actuelle, les banques de données informatisées demeurent ou sont devenues impopulaires auprès de bon nombre d'historiens — pour ne pas parler des bailleurs de fonds, généralement réfractaires à l'idée de nouvelles banques à construire. Des problèmes réels, importants, ont nourri et, jusqu'à un certain point, justifient ces attitudes. À ma connaissance, et à des degrés divers, les grands projets de banques de données historiques mis de l'avant au Canada depuis vingt ans — ceci inclut le nôtre — doivent presque tous plaider coupables aux accusations suivantes :

- Des coûts supérieurs à ce qui avait été estimé.
- Des délais de réalisation qui ont excédé les calendriers initiaux — sans compter les échecs purs et simples.
- Un déséquilibre entre les ressources investies et la production scientifique réalisée à ce jour.

- L'impossibilité de généraliser à partir d'un objet ou d'une fenêtre d'observation qui est de toute manière circonscrite dans le temps et dans l'espace. (Sont ici visées les banques définies par référence à des aires géographiques ou à des populations régionales.)

Ces critiques sont fondées, mais il faut les mettre en perspective et bien considérer le contexte difficile dans lequel les premières tentatives ont dû s'élaborer — pensons ici au fichier des démographes Charbonneau-Légaré, au *Canadian Social History Project* de M. Katz, au *Peel County Project* de David Gagan, à la banque de données de Darroch et Ornstein ou au fichier de SOREP. Tous ces projets ont dû se réaliser dans des circonstances précaires de financement à court terme et une insuffisance d'appuis techniques et matériels qui compromettaient constamment la continuité des opérations. Dans ces conditions, il était extrêmement difficile d'assurer la stabilité des équipes et de recruter les meilleures expertises. Il faut ajouter à cela que la recherche dans ces domaines neufs a dû surmonter des problèmes considérables, inconnus jusque-là, en particulier : la mobilité élevée des populations rurales aussi bien qu'urbaines; les lacunes relevées dans les grandes séries d'archives nominatives (recensements manuscrits, état civil); la fluidité des données servant de support au jumelage automatique (cf. les fameuses variations nominatives); l'apparition tardive de systèmes de gestion informatique appropriés⁴. À tout prendre, le milieu universitaire, comme cadre de recherche, n'était sans doute pas le plus approprié à ce genre d'entreprises à long terme. Il est remarquable du reste que la technologie des banques de données s'est développée plus efficacement et plus harmonieusement dans la sphère gouvernementale, surtout aux États-Unis, où elle a trouvé de meilleures conditions de financement et de stabilité. Il faut tenir compte de ces circonstances si l'on veut éviter de faux procès. De même, s'il paraît exister une disproportion entre l'ampleur des développements et les résultats d'analyse, c'est que, dans la plupart des cas, les travaux d'exploitation comme tels ont débuté depuis peu. C'est là un déséquilibre qui est appelé à disparaître très rapidement, le rapport développement-exploitation s'inversant en faveur du second terme. Il faut aussi souligner que, par rapport à des secteurs mieux établis de notre métier, l'histoire sociale à base de *corpus* informatisés opère à une micro-échelle et poursuit des interrogations différentes. Sa contribution à l'histoire politique ou nationale n'est donc pas immédiate. Encore là, il faut l'évaluer en référence avec son champ d'action. Quant au grief lié à l'impossible généralisation, il est assez inattendu de le trouver sous la plume de l'historien. Plus que tout autre instrument, les banques de données augmentent l'aire d'observation et de comparaison des phénomènes, elles étendent l'éventail de la singularité chronologique et spatiale, mais elle ne la supprime évidemment pas, ce qui est parfaitement conforme aux orientations traditionnelles de l'historiographie.

Selon d'autres critiques, les banques de données accouchent trop souvent de résultats peu originaux, presque connus d'avance. En fait, ceci met en cause non pas les banques de données elles-mêmes mais l'imagination scientifique de leurs usagers. Enfin, les banques de données auraient le tort de privilégier une démarche quantitative qui réduirait l'objet historique à ses composantes les plus mécaniques. Dans la mesure où les fichiers de population sont concernés, cette impression naît d'un malentendu ou d'une méconnaissance de l'outil. Les techniques d'accès et de croisement des données qui sont à la base du fichier-réseau consacrent précisément la possibilité de retourner selon l'échelle désirée à

4. Le concept des D.B.M.S. relationnels n'a été explicité pour la première fois qu'en 1971 (cf. E.F. Codd, 1971).

l'individu, aux acteurs, à leurs itinéraires et à leurs particularités. Chacun peut évidemment exploiter cette possibilité à ses propres fins, qu'il s'agisse du généalogiste, du biographe ou du praticien de l'histoire sociale en général.

En outre, on a tort d'assimiler informatique et quantitatif. Il est beaucoup plus juste de concevoir l'ordinateur comme un instrument d'accès rapide et flexible à des données et comme un outil de systématisation des opérations d'analyse. Il en résulte de nouvelles possibilités d'interrogations : non pas une quantification mais une extension de l'aire de recherche.

IV — UNE CONCEPTION DE L'HISTOIRE

Les pages qui précèdent avaient pour but de faire voir la pertinence des banques de données informatisées en histoire et le potentiel scientifique qu'elles représentent comme archives de deuxième génération. Elles visaient aussi à mieux fixer le cadre d'une discussion critique équitable, laquelle doit absolument tenir compte des conditions d'instabilité dans lesquelles ces banques de données ont dû se développer.

Pour le reste, tout est affaire d'orientation historiographique ou d'épistémologie, où il revient à chacun de se situer. Pour ma part, je suis partisan d'une conception très libérale de la science historique. Je crois qu'une pluralité et une diversité de démarches, de styles et d'écoles produisent une connaissance plus riche, plus nuancée et plus vivante des sociétés passées. Dans cet esprit, les banques de données, et plus particulièrement les fichiers de population, me paraissent une sorte de don que la technologie moderne a fait aux historiens pour les aider à mieux comprendre certains aspects de leur objet. Mais la connaissance des autres aspects nécessitent bien évidemment d'autres apports, « qualitatifs », bibliographiques et autres à conjuguer dans une perspective plus large.

En définitive, je crois que l'historien doit toujours raconter une histoire, quelle que soit sa spécialité, parce qu'il vise à reconstituer un espace, des acteurs, un temps mort en référence avec le présent. Mais la singularité ou l'individualité qui fait l'objet de tout récit doit aussi précisément, pour en faire ressortir la spécificité, être projetée dans ce qu'elle a de multiple, dans ce qui la prolonge, parfois la déborde même, et qui est le propre du social. Je crois aussi qu'à long terme, la recherche de l'homme vivant va se faire dans la multitude des octets autant que dans la poussière des archives. Il serait étonnant en effet que la technologie que représentent les banques de données ne puisse s'imposer comme un support original pour un nouveau modèle de gestion de la mémoire collective, nourri des apports indispensables de pratiques plus classiques associées à l'histoire dite « qualitative ». Mais faut-il réaffirmer que l'ordinateur, dans notre esprit tout au moins, ne vise qu'à enrichir et à mieux servir l'idéal commun d'une histoire qui ne peut être qu'humaniste.

Notre défense des banques de données informatisées n'a pas rallié tous les sceptiques sans doute. Mais elle a peut-être contribué à repousser le spectre d'un clivage et d'une opposition stérile entre soi-disant quantitatifs et qualitatifs. Car ce sont là deux termes d'une même exigence de notre métier : rassembler dans des constructions cohérentes les segments multiples et complexes de l'objet social. Après deux ou trois décennies d'expansion et de fractionnement, la science historique a peut-être maintenant besoin d'une génération de rassembleurs. Et au lieu de mettre en procès une technologie qui va sans cesse s'étendre quoi qu'il advienne, il sera sans doute plus avisé de réfléchir désormais aux moyens d'exploiter le plus efficacement possible cette nouvelle fenêtre sur le passé.

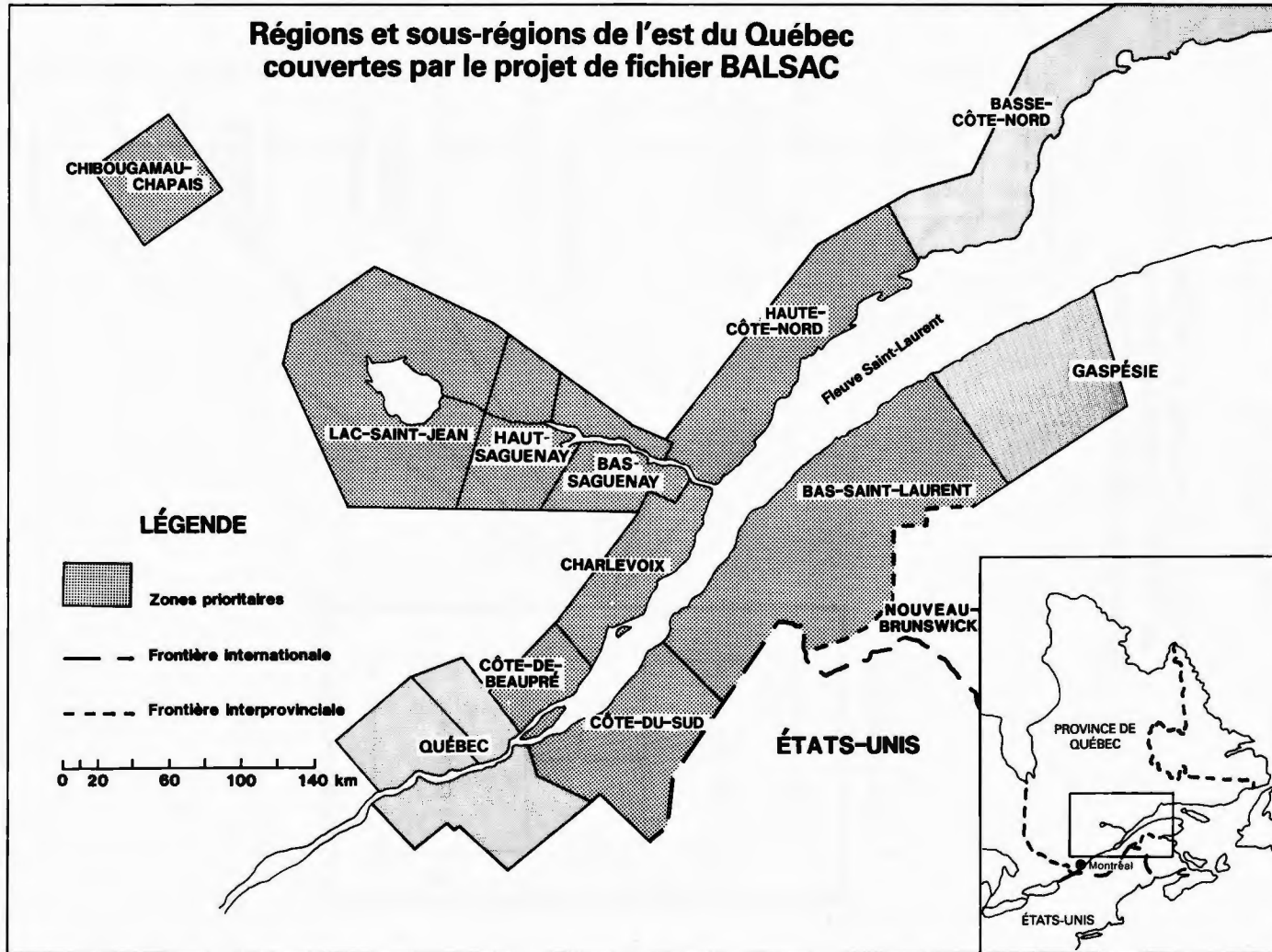
Pour le reste, dans la mesure où les banques de données informatisées sont à la fois un corollaire et un instrument des aspirations fondamentales de l'histoire sociale, on s'expliquerait mal un virage qui consisterait à l'en détourner.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Bouchard, Gérard. « Le développement de SOREP comme centre de recherche multidisciplinaire et interinstitutionnel », *Les Centres de recherche universitaires en sciences humaines*. Ottawa, Les Presses de l'Université d'Ottawa, 1987, pp. 33-47.
- Bouchard, Gérard. *La reconstitution automatique des familles : Le système SOREP*. Université du Québec à Chicoutimi, 1985, Dossier n° 2, 2 vol., 745 p.
- Codd, E.F. *Further Normalization of the Data Base Relational Model*. IBM Research, 1971, RJ 909.
- Fèbvre, Lucien. *Un destin : Martin Luther*. Paris, Les Presses Universitaires de France, 1952, 214 p.
- Henry, Louis. *Anciennes familles genevoises. Étude démographique : XVI^e-XX^e siècles*. Paris, PUF, 1956, Cahier n° 26 de l'I.N.E.D.

Annexe 1

Régions et sous-régions de l'est du Québec couvertes par le projet de fichier BALSAC



(SOREP)

Annexe 2

Structure du fichier-réseau : fichier central et fichiers sectoriels (tables) (Système relationnel INGRES)

