

two awards by topic: The Richard Cropp award of \$100 for the best paper in military history and the Cedric Cummins award of \$100 for the best paper in institutional history. All papers submitted for competition must be read by the author. Papers presented will be published if desired. Deadline for submission of papers is January 31, 1987.

Please address all correspondence to: Herb W. Blakely, Department of History, Dakota State College, Madison, SD 57042-1799.

- - - - -

NORTH AMERICAN SOCIETY FOR OCEANIC HISTORY

The North American Society of Oceanic History will hold its annual meeting at Kingston, Ontario, on 21-23 May 1987. It will meet jointly with the Canadian Nautical Research Society.

The Society welcomes proposals for papers on the meeting's theme, which is the history of the Great Lakes and the St. Lawrence River, or other maritime and naval topics. Of particular interest will be papers on the history of merchant activities, hydrography, cartography, literature, naval affairs since the 15th century, and studies of French maritime history with respect to the St. Lawrence. Also invited are papers on the Anglo-American struggle for the St. Lawrence River Basin, 1812-1814, and the diplomatic-strategic consequences of that struggle. More current themes on aspects of marine archaeology, canals, seaborne commerce, port development, and international waterways diplomacy and management also are welcome.

Please send abstracts of proposed papers, with a curriculum vitae, by 1 January 1987, to Professor Barry M. Gough, History Department Wilfrid Laurier University, Waterloo, Ontario, N2L 3C5, Canada.

* * * * *

COMPUTER-ASSISTED HISTORY AT UNIVERSITY OF OTTAWA

Over the past three years a number of professors and graduate students at the University of Ottawa have been involved in computer-assisted historical research, including statistical analysis of nominal census returns, shipping lists and library-circulation records as well as the creation of complex-data linkage systems and natural-language analysis. Drawing from our experiences, this paper endeavours to aid those embarking on computer-assisted research by discussing the problems encountered in the creation and exploitation of data sets.

thèmes précis: le prix Richard Cropp de 100\$ pour la meilleure communication en histoire militaire et le prix Cedric Cummins de 100\$ pour la meilleure communication en histoire des institutions. Toutes les communications proposées devront être présentées par les auteurs; elles seront publiées si l'auteur le désire.

Veuillez communiquer à cet effet avec Monsieur Herb W. Blakely, Département d'histoire, Dakota State College, Madison, SD 57042-1799, Etats-Unis, avant le 31 janvier 1987.

- - - - -

NORTH AMERICAN SOCIETY FOR OCEANIC HISTORY

La North American Society of Oceanic History tiendra sa réunion annuelle à Kingston (Ontario) les 21, 22 et 23 mai 1987, conjointement avec la Canadian Nautical Research Society.

On invite les chercheurs à proposer des communications sur le thème retenu pour la conférence, soit l'histoire des Grands Lacs et du Saint-Laurent ou sur tout autre thème propre à la navigation. Un intérêt tout particulier sera accordé aux communications qui traitent de l'histoire des activités marchandes, de l'hydrographie, de la cartographie, de la littérature, des affaires navales depuis le XVIe siècle et de l'histoire maritime française relative au Saint-Laurent. Seront également prises en considération les communications sur les luttes anglo-américaines dans le bassin du fleuve Saint-Laurent (1812-1814), et sur les conséquences diplomatiques et stratégiques de cet affrontement. Des communications sur des thèmes d'actualité, tels que l'archéologie maritime, la canalisation, le commerce maritime, le développement portuaire, la diplomatie et l'administration des eaux navigables internationales sont aussi les bienvenues.

Veuillez faire parvenir un résumé de votre communication et votre curriculum vitae avant le 1 janvier 1987 à Monsieur Barry M. Gough, Département d'histoire, Université Wilfrid Laurier, Waterloo (Ontario), N2L 3C5, Canada.

* * * * *

L'ORDINATEUR AU SERVICE DE L'HISTOIRE A L'UNIVERSITE D'OTTAWA

Au cours des trois dernières années, un certain nombre de professeurs et d'étudiants aux études supérieures de l'Université d'Ottawa ont utilisé l'ordinateur au cours de leurs recherches en histoire, pour faire, par exemple, l'analyse statistique de formulaires généraux de recensement, de listes d'envois et de registres de prêts bibliothécaires, ainsi que pour créer des systèmes d'assemblage complexes de données et des analyses du langage naturel. Cet article, qui est le fruit de notre expérience, est écrit à l'intention de ceux qui s'engagent dans la recherche assistée par ordinateur; il aborde les problèmes auxquels nous nous sommes heurtés lors de la création et de l'exploitation des ensembles de données.

Two major factors limit the use of a computer in historical research: suitability of the source material and the financial and time costs. First, it is essential to know the limits of using the computer and the kinds of questions that a computer can help to answer. Second, using the computer can be costly in terms of computer time, learning time, data entry and the purchasing of hardware and software. There is no strict rule by which to judge the cost-benefits of computer-assisted research, but a clear idea of expected results is helpful in making the decision in whether to use the computer.

The major problem facing historians is data collection. The transforming of information found in non-circulating manuscripts or microfilm into machine-readable form can be time-consuming and error-laden. A data-collection program makes coding sheets obsolete, speeds up the data-collection process and reduces the number of errors. Program such as WATFILE, PC-FILE and DATCOL automatically format the data, eliminating the prime source of data-entry error. More sophisticated database programs and dedicated data-entry program include data-verification modules, checking, for example, if a value falls within a given range of values or if a value is one of a defined set of key words. Unfortunately, these more elaborate systems are expensive and difficult to learn.

An important concern is the method of entering the data into the computer. Two options are available. One option is to enter the data as it exists in the source, that is, to type the name, occupation, religion etc. as it appears in the original document. Subsequently, the computer would transform the raw data into numeric codes. A second option is to put the data into numeric or coded values on entry, which entails the creation and updating of a codebook. Our experience suggests that the second option is preferable for two reasons. First, for many types of data -- book titles, addresses, occupations, for example -- the entering of data as it appears greatly increases data-entry time. Second, translating uncoded data into codes by computer is not a simple task. Variation in spelling and typing errors cause machine errors.

We recommend the entering of data in unclassified form, assigning a discrete code to each different piece of data. Leaving the regrouping of data until after the completion of the data-entry stage avoids the pitfall of formulating a classification system which later may prove to be too limiting or completely unsuitable.

Once in machine-readable form the data is ready for any number of operations including data-preparation routines such as nominal-record linkage, data-verification programs,

Deux facteurs majeurs limitent l'utilisation de l'ordinateur dans la recherche historique: la convenance du matériel-source, et la dépense en temps et en ressources financières. En premier lieu, il est essentiel de connaître les limites d'utilisation de l'ordinateur et le genre de questions auxquelles celui-ci peut aider à trouver une réponse. En second lieu, l'utilisation de l'ordinateur peut être onéreux en termes de temps-ordinateur, de temps d'apprentissage, d'introduction des données, et d'achat de matériels et de logiciels. Il n'existe aucune règle qui permet de juger les coûts-avantages d'une recherche assistée par ordinateur, mais il est bon d'avoir une idée précise des résultats escomptés pour décider si l'utilisation de l'ordinateur s'avère utile ou non.

La collecte de données est le problème majeur auquel les historiens sont confrontés. La transformation en forme exploitable par l'ordinateur de l'information trouvée dans des manuscrits privés ou dans des microfilms, peut prendre du temps et peut constituer une source d'accumulation d'erreurs. Un programme de collecte de données rend les feuilles de programmation désuètes, accélère le processus, et réduit le nombre d'erreurs. Des programmes comme WATFILE, PC-FILE et DATCOL mettent les données automatiquement en forme, ce qui élimine la source d'erreurs principale. Des programmes de base de données plus perfectionnés, et d'autres dédiés à la saisie des données, comprennent des modules de vérification de données vérifiant, par exemple, si une valeur fait partie d'une classe ou si elle est l'une de celles qui figure dans un ensemble défini de mots-clés. Malheureusement, ces systèmes plus élaborés sont très onéreux et difficiles à apprendre.

La méthode d'introduction des données dans l'ordinateur est importante. Deux possibilités sont offertes. La première est d'introduire les données telles qu'elles existent à la source, ce qui veut dire taper le nom, la profession, la religion, etc., comme ils apparaissent dans le document original. Par la suite, l'ordinateur transformera les données brutes en codes numériques. La seconde possibilité est d'introduire les données sous forme de valeurs numériques, ou codées, ce qui entraîne la création et la mise à jour d'un livret de codes. Notre expérience suggère que la seconde méthode est préférable pour deux raisons. Premièrement, l'introduction de certains types de données telles qu'elles apparaissent -- comme les titres de livres, les adresses, les professions -- nécessite beaucoup de temps. Deuxièmement, faire traduire par l'ordinateur des données non-codées en codes n'est pas une tâche simple. Les variations orthographiques et les erreurs de frappe peuvent causer des erreurs de programmes.

Nous recommandons d'introduire les données dans une forme non-classifiée en désignant un code discret à chaque unité de donnée. En regroupant les données après l'introduction, on évite le piège d'avoir à reformuler un système de classification qui s'est révélé trop limité ou complètement inadéquat.

Une fois dans une forme exploitable par la machine, les données sont prêtes pour toute opération incluant la saisie des données telle que la liaison de rubriques générales, les programmes de vérification et

data reformatting and selection of data subsets. One important aspect in generating complex data sets is the linking of information from more than one source. Comparable to the process of amassing pertinent information on an individual or a single event from private papers, published material or newspapers, the computer allows the creation of relational data sets which link together two or more files sharing such variables as family name, social insurance number or age. For example, one study underway at the University of Ottawa links nearly 13,000 individuals, using the names, addresses and occupations listed in three kinds of routinely-generated sources, including eight pollbooks, five census returns and five assessment rolls.

The decision to link by hand or by computer involves a number of considerations. The time and resources involved in linking the files must be weighed against the size and complexity of the data set. Once the choice is made to use an automated linkage system, the researcher must also consider the kind of system to be implemented. Historians have little difficulty in deciding whether two different sets of information refer to the same person. Unfortunately, programming computers to recognize misspelled names, subtle differences in occupation or changes in address is a complex process. The program must translate the historian's "judgement" into simple logical steps. Not a simple task. There is no standard algorithms for the recognition of names, although social scientists frequently use language-specific coding systems such as SOUNDEX or FONEM. To complicate matters, there is no package on the market specifically designed for historical record linkage. This requires the researcher writing a linkage program or effectively communicating a set of rules of judgement to a programmer.

Once data is in machine-readable form, further modifications, such as adding more data to a case or reformatting the data into more effective structures, are easily accomplished. Nominal record linkage is only one way of creating complex records. The collection of data in stages is also possible. For example, in a book-circulating study, common code numbers permitted the adding of new bibliographical information to previously collected circulation data. High-level languages such as SNOBOL also allow the writing of error-detection and correction programs. For example, to "clean" large data sets we created a program for the detection and correction of data-type, justification and record-length errors. This saved considerable time and effort.

la sélection de sous-ensembles. Un aspect important de la production d'ensembles complexes de base de données est la liaison d'information de plus d'une source. Comparable au processus de collecte d'information pertinente sur un individu ou sur un simple événement à partir de documents privés, de documents publiés ou de journaux, l'ordinateur permet la création d'ensemble de données relationnelles lesquelles lient ensemble deux ou plusieurs fichiers qui partagent des variables comme le nom de famille, le numéro d'assurance sociale ou l'âge. Par exemple, une étude à l'Université d'Ottawa relie actuellement près de 13 000 personnes, en utilisant le nom, l'adresse et la profession selon trois types de sources créées par programme, incluant huit livres de scrutin, cinq recensements et cinq listes d'évaluation.

La décision de faire ces liaisons à la main ou par l'entremise de l'ordinateur implique un certain nombre de considérations. On doit évaluer le temps et les ressources engagés dans la liaison des fichiers en fonction de la taille et de la complexité de l'ensemble des données. Une fois qu'il a décidé d'utiliser un système de liaison automatisé, le chercheur doit alors choisir quel type de système il va mettre en place. Les historiens ont peu de difficulté à déterminer si deux différents ensembles d'information se rapportent à la même personne. Malheureusement, programmer des ordinateurs pour reconnaître des noms mal orthographiés, des différences subtiles dans la profession ou des changements d'adresse, est un processus complexe. Le programme doit traduire le "jugement" de l'historien en simple démarche logique. Ce n'est pas une sinécure. Il n'existe aucun algorithme de référence pour la reconnaissance des noms, même si les chercheurs en sciences sociales utilisent fréquemment des systèmes de codage de langages sélectifs tels que SOUNDEX ou FONEM. Pour compliquer les choses, il n'existe pas sur le marché de programmes spécialement conçus pour relier des données historiques. Le chercheur est obligé d'écrire lui-même un programme de liaisons ou bien de communiquer efficacement à un programmeur un ensemble de règles de jugement.

Une fois les données dans une forme exploitable par la machine, il est facile de faire des modifications, comme pas exemple ajouter des données à un cas ou recomposer des données en structures plus efficaces. La liaison de dossiers généraux est seulement une des façons de créer des dossiers complexes. Il est aussi possible de faire la collecte de données par étapes. Par exemple, lors d'une étude sur les emprunts à la bibliothèque, on a pu, grâce au numéro de code commun, ajouter de nouvelles informations bibliographiques à des données déjà recueillies sur les emprunts. Des langages de haut niveau comme SNOBOL permettent également d'écrire un code détecteur d'erreurs ainsi que des programmes pour la détection et la correction des erreurs dans le type de données, dans la justification et dans la longueur des dossiers. Nous avons donc épargné bien des efforts et un temps considérable.

Simple data changes are possible with such statistical packages as SPSS and SAS, but major modifications to the data set, including translating alphanumeric to numeric values, adding or eliminating variables, compressing or performing calculations to the data, are also feasible. For example, a program written in SNOBOL for the pollbook study reduced the number of cards per case from 22 cards to 7 cards by totalling assessment values, calculating ages, and copying information collected from one source to cards from other sources. It is also possible to create programs that write large sections of statistical programs. In the book-circulation study, a SNOBOL program reformatted the author, title and borrower codebooks into the appropriate SPSS format-value labels. The end result was a 3000-line SPSS program that was fully machine-generated and error-free.

Analysis of data need not be restricted to a particular packed program or one type of machine; collecting the data in standard-data format provides flexibility in terms of program and machine preference. We favour the 80-column-fixed format because it is readily usable during the data-entry stage, it is transmittable via modem to other machines and is easily transformable to less common formats.

For analyzing the data, there are two different types of packages available: statistical and database. The strength of statistical packages lies in numeric analysis where there is less concern for individual cases. Database programs, like dBASE III or SPIRES, deal with data on a case by case basis, where the data is typically more meaningful to humans. Names, occupations, and even text can be stored, queried and retrieved using a database system. Properly formatted data allows the historian to use both methods of analysis if the data contains both numeric and alphanumeric values.

The researcher should be aware of the flexibility of modern computer systems. In the past few years, the appearance of more powerful microcomputers has given the historian direct access to computers that rival the mainframes of past decades without the intimidating surroundings that used to accompany computer-assisted research. In particular, text and database processing have opened new kinds of computer applications to the historian. At the same time, the cost of using this research tool, while decreasing, remains still considerable. Furthermore the researcher may have to acquire

Des changements simples de données sont possibles avec des progiciels statistiques tels que SPSS et SAS, mais les modifications majeures à l'ensemble des données, comme la traduction de valeurs alphanumériques en valeurs numériques, l'addition ou l'élimination de variables, la compression ou le calcul, sont aussi possibles. Par exemple, un programme écrit en SNOBOL pour l'étude des livres de scrutin, a réduit le nombre de cartes par cas de 22 à 7 cartes, en totalisant les valeurs de l'évaluation, en calculant les âges, et copiant l'information recueillie d'une carte-source aux autres. Il est aussi possible de créer des programmes qui écrivent de grandes sections de programmes statistiques. Dans l'étude sur l'emprunt des livres à la bibliothèque, un programme SNOBOL a recomposé le nom de l'auteur, le titre, et le code de l'emprunteur dans la structure propre au SPSS. Le résultat fut un programme SPSS de 3 000 lignes créé entièrement par l'ordinateur et sans erreur.

L'analyse des données n'est pas limitée à un type de programme en particulier ou à un type de machine; recueillir les données dans une forme normalisée procure une certaine flexibilité en termes de programme ou de machine. Nous préférons la structure fixe en 80 colonnes parce qu'elle est déjà prête à être utilisée lors du stade d'introduction, elle est transmissible, via modem, à d'autres machines et elle est facilement transformable en structures moins communes.

Pour analyser les données, il existe deux types de progiciels: statistiques et base de données. Le point fort des progiciels statistiques est l'analyse numérique lorsque l'on se soucie peu des cas individuels. Des programmes de base de données comme dBASE III ou SPIRES, traitent les données cas par cas, lorsque les données sont pertinentes à l'aspect humain de l'analyse. Les noms, les professions, et même des textes peuvent être mémorisés, consultés et retrouvés par un programme de base de données. Des données proprement structurées permettent à l'historien d'utiliser les deux méthodes d'analyse si les données contiennent à la fois des valeurs numériques et alphanumériques.

Le chercheur doit être conscient de la flexibilité des systèmes d'ordinateurs modernes. Au cours des dernières années, l'apparition de micro-ordinateurs plus puissants a donné à l'historien un contact direct avec des systèmes comparables aux unités centrales de traitement des décennies passées, tout en évitant d'imposer à ce genre de recherche le milieu intimidant qui auparavant caractérisait la recherche informatisée. En particulier, le traitement de texte et les bases de données ont ouvert à l'historien de nouvelles possibilités. Par contre, le coût d'utilisation de ces outils de recherche, bien qu'il diminue, demeure important. De plus, le chercheur peut être obligé d'apprendre un langage d'ordinateur de haut niveau

knowledge of a high-level language and should expect to devote a large amount of time learning to use the computer efficiently and effectively.

Mark Olsen and David De Brou

* * * * *

CHA WORKSHOPS ON THEMATIC ISSUES

The Canadian Historical Association is pleased to announce that Paul-André Linteau and John Taylor have been awarded a grant for "Women's History and Urban History: Integration or Separation?", a colloquium which will be held at the University of Quebec in Montreal on October 3-4. The CHA is currently reviewing the terms of reference for this program. They will be published in the next Newsletter/Bulletin. In the meantime, for further information, please contact Professor Judith Fingard, Department of History, Dalhousie University, Halifax, Nova Scotia. B3H 3J5 Tel.: (902) 424-2011.

* * * * *

ANNOUNCEMENTS FOR THE NEWSLETTER

We are asking our members, whenever possible, to forward their material in both official languages, (conference announcements, calls for papers, etc) before the 31 October 1986. The CHA's translation costs are already quite large. In addition, the editors also spend a lot of their time on translation

* * * * *

MAIL BALLOT

As you can see from the following statistics, the mail ballot resulted in a substantial increase in votes participation.

ELECTIONS RESULTS, 1977-1986/RESULTATS DES ELECTIONS, 1977-1986

<u>Year/Année</u>	<u>Votes</u>	<u>Proxies/Par procuration</u>
1977	189	28
1978	189	?
1979	200	61
1980	193	63
1981	165	42
1982	128	20
1983	174	59
1984	173	37
1985	198	58
* 1986	357	17 (Meeting/Réunion)

* Mail ballot/Scrutin par la poste

* * * * *

et doit s'attendre à consacrer beaucoup de temps à apprendre à se servir de l'ordinateur d'une manière efficace.

Mark Olsen et David De Brou

* * * * *

COLLOQUES SUBVENTIONNES PAR LA SHC

La Société historique du Canada est heureuse d'annoncer que les professeurs Paul-André Linteau et John Taylor ont reçu une subvention pour le colloque intitulé Histoire des femmes et histoire urbaine: intégration ou séparation?, qui se tiendra à l'Université du Québec à Montréal, les 3 et 4 octobre prochain. La SHC est en voie de revoir les critères de sélection de ce programme. Ils paraîtront dans le prochain numéro du Bulletin. Entre-temps pour plus de renseignements, prière de communiquer avec Professeure Judith Fingard, Département d'histoire, Université de Dalhousie, Halifax, Nouvelle-Ecosse. B3H 3J5 Tél.: (902) 424-2011.

* * * * *

ANNONCES POUR LE BULLETIN

Nous demandons à nos membres individuels ou collectifs de faire leur possible pour expédier au secrétariat, dans les deux langues officielles du pays, toute annonce (avis de congrès, appels de communications, etc.) qu'ils désirent faire paraître dans le Bulletin, avant le 31 octobre 1986. La traduction à contrat coûte quelques centaines de dollars par année à la SHC; la traduction faite par les éditeurs consomme beaucoup de leur temps.

* * * * *

SCRUTIN PAR LA POSTE

Comme le suggère les statistiques qui suivent, le scrutin par la poste a accru de façon substantielle la participation aux élections.

* * * * *