

ACTION GROUP REPORTS

DATA ARCHIVE REGISTRY

- Canada- Lisa Lasko, Canadian Consortium for Social Research, Institute for Behavioral Research, York University, 4700 Keele Street, Downsview, Ontario M3J 1P3
- Europe- Joseph Bonmariage, Belgian Archives for the Social Sciences, University of Louvain, SH-2, 1348 Louvain-la-Neuve, Belgium
- United States- John Kolp, Regional Social Science Data Archive, University of Iowa, Iowa City, Iowa 52242

Report of the Joint Canadian-United States Action Groups on
Data Archive Registry Meetings in Toronto, May 11-12, 1977 (DAR AG)

Submitted by

Lisa Lasko
Institute for Behavioral Research
York University

Members Present

Canada

Lisa Lasko, Institute for Behavioral Research, York University
Sharon Henry, Data Clearing House for the Social Sciences
Gerald Prodrick, University of Western Ontario
Jana Prokop, University of Toronto

The Canadian Data Archive Registry Group made substantial progress when it met for a working session at the Canadian IASSIST meetings, May 11 and 12, in Toronto. The working sessions were attended by four Canadian IASSIST members. No American members were present.

The first issue that was addressed by the group was whether or not the Data Archive Registry Group still held a viable mandate. The presence of two new data archive directories-- the upcoming 1977 September edition of Vivian Sessions' Directory of Data Bases in the Social and Behavioral Sciences and the Unesco sponsored Directory of Data Services now underway under the direction of Jean Meyriat, Secretary General of the International Committee for Social Science Information and Documentation (ICSSD) forced consideration of this issue. In addition, the Canadian Group had to consider that the Data Clearing House for the

Social Sciences intended to produce a hard copy Directory of Social Science Data Centres for Canada. The group agreed that the first two directories were less than satisfactory, but that the soon-to-be-produced directory of the Data Clearing House posed a real conflict. Sharon Henry, Executive Director of the Data Clearing House, reported that although she had to have the Directory published by the end of the current year, in fact, no work had yet been started on the publication and no questionnaire had yet been designed. It was suggested that a formal liaison between the IASSIST Data Archive Registry Group and the Data Clearing House be established to produce the directory. More specifically, the questionnaire utilized for surveying Canadian data centres would be jointly designed by both groups, and the data elements covered in the questionnaire would allow the results to be used by both groups for their various needs. One publication, the Directory of Social Science Data Centres, would be published cooperatively by the Data Clearing House for the Social Sciences and the IASSIST Data Archive Registry Group. However, it was clearly understood that when and if an IASSIST International Data Archive Registry did get underway, the relevant Canadian entries would be pulled from the data base for inclusion in the IASSIST registry. The Canadian directory, then, was to be viewed as a pilot project for the international IASSIST data archive registry. It was agreed that in order for this joint activity to work, a formal representative from the Data Clearing House for the Social Sciences should become a permanent member of the IASSIST Data Archive Registry Group.

Once this basic issue was resolved, the group immediately started to work on designing the questionnaire. Both the Session's questionnaire and the Meyriat questionnaire were consulted. By the end of the two days, most of the necessary elements had been agreed upon and Pierre Lacasse, coordinator of the Data Acquisition Group, made some valuable suggestions for elements concerning acquisition policies of archives. However, as work on the questionnaire did not get completed, it was decided that Lisa Lasko and Jana Prokop would meet in Toronto to complete it, after which time, a meeting would be held with Sharon Henry to discuss it further.

The group spent a substantial portion of its time compiling a list of broad subject headings to be used for describing categories of data in the questionnaire. During the conference, it was pointed out that the lack of a controlled vocabulary for descriptions of categories or holdings of data, was a major factor in the lack of good subject access to data archives. The group felt that its work in this area would be a significant and important contribution to useful descriptions of data archives. The group built upon the list of 105 subject headings compiled by David Gerhan and Loretta Walker, and also consulted with Vivian Sessions' 26 broad subject categories. In addition, unique Canadian subject terms were chosen. The final version differed substantially from the aforementioned lists, and, as it was still very rough, Lisa Lasko and Jana Prokop agreed to refine the list further at a later date in Toronto. Sue Dodd, Coordinator of the Classification Action Group, generously offered to assist in the compilation of subject terms. It was agreed that the refined list would be sent to Sue Dodd for comments and suggestions as soon as possible.

To summarize, then, the Data Archive Registry Group made significant headway in designing the questionnaire and developing a controlled vocabulary to be used for descriptions of data archive holdings. However, further work is still required to complete both the questionnaire and subject heading list. For this reason, Lisa Lasko and Jana Prokop both agreed to meet in Toronto in the next few weeks to take care of any work still outstanding.

CLASSIFICATION

- Canada- Mohan Sharma, Humanities & Social Science Library, University of Alberta, Rutherford North, Edmonton, Alberta
- Europe- Ekkehard Mochmann, Zentralarchiv für Empirische Sozialforschung, Bachemer Strasse 40, 5 Köln 41, Federal Republic of Germany
- United States- Sue Dodd, Data Library, Insitute for Research in Social Sciences, Manning Hall, University of North Carolina, Chapel Hill, North Carolina 27514

Report of the Joint Canadian-United States Action Groups on
Classification (C AG)

Submitted by

Sue A. Dodd, University of North Carolina

Members Present

Canada

Martha Amschutz, Canadian Radio-Television Commission
Krystyna W. Dynowski, University of Western Ontario
Sue Gavrel, Public Archives of Canada, Ottawa
Katrín Horowitz, Library of Consumer & Corporate Affairs
Hans G. Schulte-Albert, University of Western Ontario

United States

Sue A. Dodd, Ch, University of North Carolina
Gertrude Lewis, Rutgers University
Mimi Schade, Brookings Institution
Philip Sidel, University of Pittsburgh

Agenda topics: Current problems and associated tasks within the mandate of the Classification Action Group include developing (1) examples and guidelines for bibliographic references for social science numerical data files, (2) a "cataloging-in-production" scheme for major producers of social science data files, (3) a more "universally based" classification scheme for social science data files; and, reviewing (4) the cataloging efforts to date and discussing any or all related problems and (5) existing printed thesauri in the social sciences in terms of their future applications to data files.

The primary focus of the C AG at the Canadian Working Conference centered around the first problem of how to cite properly a social science numerical data file in the published literature.

Problem: Currently, there are no standards or guidelines for compiling such bibliographic references and the quality and amount of information provided when citing social science data files varies greatly among individual scholars. In many cases, the information is not sufficient to allow for direct access, and consequently, an interested party has to spend a considerable amount of time determining additional information that could easily be provided by a full bibliographic reference.

Complicating this problem is the American National Standards Institute's forthcoming work entitled "American National Standards for Bibliographic References." This ANSI Standard will provide detailed information on the theory, principles, and definitions underlying the technique of preparing bibliographic references for both print and nonprint materials, including machine-readable data files (MRDF). The problem with this forthcoming document is that it does not include examples of social science numerical data files which make up the vast majority of MRDF; nor does it attempt to be compatible with the forthcoming second edition of the Anglo American Cataloging Rules (AACR II), which also deals with bibliographic standards for MRDF.

Role of IASSIST C AG: The Classification Action Group, in its continuing role to encourage uniform standards for MRDF and to see that social science data files are represented in the best possible light will provide a written critique on the forthcoming ANSI Standard, and at the same time provide more serviceable examples of how to cite social science numerical data files in the published literature. The C AG does not intend to create a new standard but rather to work within the framework of the ANSI Standard. The intent is to supplement it and recommend revisions where it is viewed as either necessary or helpful.

The critique will focus on the omission of social science numerical data files and the lack of compatibility with other standards and related terminology dealing with the same medium. Arguments to be presented in the paper are:

that outside the field of quantitative research, social science numerical data files are probably the least understood category of MRDF and, consequently, are in the most need of attention and clarification;

that the terminology and definitions used to describe MRDF are confusing and often misleading and, therefore, a uniform standard for this medium is necessary;

that social science numerical data files make up the largest on-going collection of MRDF and, consequently, this body of information is in immediate need of bibliographic control;

that social science numerical data files are cited more frequently in scholarly journals than the two MRDF examples listed in the ANSI Standard (i.e., computer programs and bibliographic data files), and therefore, the need for relevant social science examples seems justified;

that this justification is magnified by the fact that the manner in which social science data files are cited in the literature plays an important role both in terms of access and in terms of secondary analysis; and,

that the two examples given (i.e., computer programs, and bibliographic data bases) are significantly different from the characteristics of social science numerical data files as to be less than helpful.

The first draft of this paper was reviewed by the participants of the Canadian Working Conference and will likewise be reviewed by those C AG participants who were unable to attend the conference. The final work will be reviewed by the C AG's Standards and Quality-Control Review Board. The resulting critique, along with an accompanying letter signed by the appropriate officers of IASSIST, will be sent to the American National Standards Institute, Inc. in New York and beyond that, it is hoped that the recommended revisions and resulting examples would be accepted by this body. It is also hoped that relevant examples could be sent to the editors and publishers of the various social science journals, who in turn can begin to implement conventions on the use, style and content of bibliographic references.

Tasks of C AG participants: (1) read and review the draft of the critique of the ANSI Standard, providing any comments, suggestions, or additions, etc.; (2) read an abbreviated version of the standard prepared by Ellis Mount and which appeared in the Journal of the American Society for Information Science; (3) prepare a list of the necessary and descriptive bibliographic elements (e.g., title, author, edition, imprint, etc.) for social science data and at the same time indicate those that would be essential, recommended, or optional; and, (4) utilizing the list of rated bibliographic elements, begin to compose bibliographic references of the respective data files.

Results of this task at the Canadian Conference: The first draft of the critique was reviewed and comments accepted. The beginnings of a list of bibliographic elements was compiled. Each element was grouped and rated (i.e., essential, recommended or optional) according to its importance in identifying a data file. Finally, some eleven examples were compiled by the participants, including the samples listed below:

Title- Mexico's naturalized citizens, 1828-1931 [Machine-readable data file].
 first: Harold Sims; Susan Sanderson; Philip Sidel, principal investigators.
 Pittsburgh, PA: University of Pittsburgh, 1975-76. 1 data file (8066 logical records).

Author- Shanas, Ethel. The health of older people [Machine-readable data file]:
 first: A social survey: public attitudes on older people. NORC rev. ed.
 Chicago: National Opinion Research Center, 1957 [producer and distributor]. 1 data file (2567 logical records) and accompanying codebook (166 p.).

Swidzinski, Susan. Syllabication [Machine-readable data file]: a drill and practice lesson. Bloomington, MN: Control Data Corporation, 1976.
 On-line program lesson available only via the Plato System.

Henry, Neil. MAXCLS.BAS [Machine-readable data file]: a program for maximum likelihood estimation of parameters of unrestricted latent class models. Lafayette, IN: Gary Income Maintenance Experiment, 1974;
 Pittsburgh, PA: Social Science Computer Research Institute [distributor]. 1 program file (95 statements, BASIC) and accompanying manual (53 p.).

The work at the Canadian Conference on this task was an important first attempt towards achieving the goal of providing relevant and acceptable examples of bibliographic references of social sciences data files. However, these examples should be viewed as only first attempts and not as the final recommended examples of bibliographic references; and, any and all comments are welcome.

DATA ARCHIVE DEVELOPMENT

Canada- Laine Ruus, Data Library, Computing Centre, University of British Columbia,
2075 Wesbrook Place, Vancouver, British Columbia V6T 1W5

Europe- not activated

United States- Alice Robbin, Data and Program Library Service, 4452 Social Science
Building, University of Wisconsin-Madison Wisconsin 53706

Report of the Joint Canadian-United States Action Groups on
Data Archive Development Meetings in Toronto, May 11-12, 1977 (DAD AG)

Submitted by

Laine Ruus, University of British Columbia

Members Present

Canada

Laine Ruus, Ch, University of British Columbia
Marilyn Berry, University of Victoria
Peter Clinton, Memorial University of Newfoundland
Judy Demaine, University of Guelph
Ed Hanis, University of Western Ontario
Alan Kirby, Queen's University
Elaine Kozak, Data Clearing House for the Social Sciences

United States

John Heddeshheimer, National Archives and Records Service, Washington, D.C.
Alice Robbin, University of Wisconsin-Madison
Judith Rowe, Princeton University
Ed Vickery, Research Triangle Institute, Chapel Hill, NC

Objectives of the meeting: (a) to revise and redefine the outline of the Guide to providing social science data services; (b) to elaborate and expand on the contents of the separate sub-sections of the Guide; (c) to finalize arrangements for contributions from the other action groups to relevant sub-sections of the Guide, and set deadlines for abstracts and full-text contributions; (d) to allocate responsibility for contributing remaining sections and sub-sections of the Guide to persons or groups with relevant expertise.

The first priority of the action group was to write abstracts of those sections of the Guide to be contributed by other action groups, so as to afford them a chance to consider more fully, during their meetings, their contributions and react to them. On the basis of these reactions, some sub-sections of the outline will be redefined, and some responsibilities reallocated.

Specific responsibilities for remaining sub-sections of the Guide were allocated to such persons of expertise in relevant areas as could be identified by the members of the action group, both within and outside the action group itself. Those who could not be contacted during the course of the conference, or who had not previously been contacted, will be solicited for their collaboration immediately.

A deadline for contribution of abstracts of all sections was set for June 30, 1977, and deadline for full-text for December 31, 1977.

In addition, a system of readers, for the several sections of the Guide, was established to provide a pre-editorial review of input, and to ensure that foci of the sections include minimal as well as maximal levels of service.

It was further decided that preliminary copies of the glossary should be distributed to all contributors and readers for comment and amendment, to ensure that all vocabulary necessary is included, and that definitions in the glossary agree with usage by the contributors.

Based on reaction from members of the Data Organization and Management Action Group, it was found to be necessary to further refine and amplify the definition of the target audience, especially in terms of levels of expertise. A fuller redefinition will be distributed to all collaborators.

The final order of business was a decision on a topic for the DAD-AG symposium at the February 1978 meetings. Various topics were considered, and the final choice was 'networking', as a topic of some immediacy for data archives/libraries. Consideration will be given to a definition of 'networking', and its implications for all machine-readable data service units.

PROCESS-PRODUCED DATA

Canada- John DeVries, Social Science Data Archives, Department of Sociology,
Carleton University, Ottawa, Ontario K1S 5B6

Europe- Paul Müller, Institute for Applied Social Research, University of
Cologne, Greinstrasse 2, 5000-Köln 41, Federal Republic of Germany

United States- Donald Harrison, National Archives (NNR), Washington, D.C.
20408

Report of the Joint Canadian-United States Action Groups on
Process-Produced Data Workshop in Toronto, May 11-12, 1977 (PPD AG)

Submitted by

John DeVries, Carleton University

Members Present

Canada

John DeVries, Ch, Carleton University
Bill Bradley, Health and Welfare Canada, Ottawa
Hy Burshtyn, Carleton University
Tony Falsetto, Public Archives of Canada, Ottawa
Richard Guttormson, Ministry of State for Science and Technology, Ottawa
Pierre Lacasse, Centre de recherche en aménagement régional, Sherbrooke, Québec

United States

Harriet Dhanak, Michigan State University
Shirley Gilbert, Princeton University
Elizabeth Powell, LEAA/NCJISS, Department of Justice, Washington, D.C.
Ed Vickery, Research Triangle Institute, Chapel Hill, NC

We continued our discussions on the Directory of Catalogues which list Machine Readable Data, which we had determined to have the highest priority for our AG. Although we did not meet our self-imposed deadline (which expected us to produce a first draft of this proposed document at this meeting), some progress has been made. We have obtained copies of several existing catalogues and propose to continue this collection phase.

We developed a listing of minimally required elements of information, which we hope each entry in a Catalogue of Data Files would provide. The listing which follows was based on an examination of existing catalogues. The essential items, we feel are:

- a) File name, catalogue number or reference number (if any);
- b) A short description of the data. This description should especially indicate the time-span and the geographical area covered by the data;
- c) Medium on which the data are held (e.g. magnetic tape, cards);
- d) Size of file (number of records);
- e) References to available documentation and other publications pertaining to the file (e.g. a listing of reference numbers for working papers);
- f) The degree of accessibility (we propose: unconditional, conditional, never);
- g) Cost of acquiring the file (if the file is accessible);
- h) Department or agency where the raw data originates;
- i) The position of the person to contact for further information.

On the last-mentioned item, we felt that the name of the person, and the telephone number, were additional items of importance. Given the high degree of turnover of personnel, we felt that the position would generally be more enduring information than the name of the incumbent at a given point in time.

We envision the eventual document--still planned to appear in February 1978--to be an annotated bibliography of Catalogues, where the annotations are based on our proposed set of minimal information requirements. The bibliography would also state these requirements, and would, finally, contain a listing of governmental departments and agencies which, to our knowledge, have not (yet) produced a catalogue of their machine-readable data files.

We are proposing to circulate our initial findings to the members of our AG before August 31, 1977. At this point each AG member will be asked to produce the annotations for a specified set of catalogues. We expect to review the eventual set of annotations at the projected IASSIST meeting for February 1978.

With regards to our second priority: the Inventory of Existing Guidelines in Use by the Originating Institution or Archival Institution, we decided that little could be done at this stage, until we had an opportunity to examine the Catalogues of Data Bases. In addition, Tony Falsetto will mail out (again) a package containing various documents related to the documentation and preservation of MRDF by the Machine-Readable Archives Division of the Public Archives of Canada.

We discussed the issue of communications between national governments and the community of researchers using process-produced data. We agreed that there is a great need for improvement and propose the following strategy:

(1) Comments, suggestions, queries and complaints about process-produced data would be sent to an IASSIST member who is at the same institution as the person requiring the information (or in a nearby institution). In short, we would propose that IASSIST endorse a network of "Official Representatives".

(2) Such questions, etc. would then be passed on to the appropriate AG coordinator (i.e., the U.S. Coordinator if one were dealing with U.S. PPD, the Canadian coordinator in the case of Canadian PPD).

(3) The AG coordinator would convey the question to the appropriate government agency and would be responsible for channeling the response back to the user (or, if the information were of wider relevance, to submit an item to the IASSIST Newsletter).

The AG also discussed ways in which other organizations working on related problems could be contacted to prevent any duplication of effort regarding process-produced data. We agreed that the two coordinators would get in touch with the chairman of such groups as the Association of Public Data Users, the Committee on Population Statistics of the Population Association of America, and so on.

DATA ORGANIZATION AND MANAGEMENT

- Canada- Greg Morrison, Social Science Data Archive, Department of Sociology,
Carleton University, Ottawa, Ontario K1S 5E6
- Europe- Eric Tannenbaum, Social Science Research Council Survey Archive, Uni-
versity of Essex, Wivenhoe Park, P.S. Box 23, Colchester, Essex,
England CO4 3SD
- United States- William Gammell, Social Science Data Center, University of
Connecticut, Storrs, Connecticut 06268

Report of the Joint Canadian-United States Action Groups on
Data Organization and Management Meetings in Toronto, May 11-12, 1977 (DOM AG)

Submitted by

Greg Morrison, Carleton University

Members Present

Canada

Greg Morrison, Ch, Carleton University
Clement K. M. Chan, McMaster University
Janet Chan, University of Toronto
Rachel Des Rosiers, Data Clearing House for the Social Sciences
Robert Logan, University of Guelph
Paula Mitchell, University of Western Ontario
Richard L. Schnaar, Public Archives of Canada, Ottawa
Terry Stewart, University of Waterloo
Richard Wolfe, Ontario Institute for Studies in Education

United States

Bill Gammell, Ch, University of Connecticut
Barbara Aldrich, University of Wisconsin
Gary M. Grandon, University of Connecticut
Pnina Grindberg, Columbia University
Sheldon Laube, C. M. Leinword Associates
Barbara Noble, University of Illinois
Sharon Poss, Duke University
Richard Roistacher, University of Illinois
Phil Sidel, University of Pittsburgh

During the Toronto sessions, members of the Data Organization and Management Group engaged in a reassessment of how various sorts of activities might best be carried out. It was felt that some kinds of projects will have to be done essentially by individuals or small sub-groups, in accordance with their particular areas of interest and expertise. Draft documents, papers, proposals and the like will then be sent to the co-ordinators of the Group, who will circulate them to Group members for reactions and comments. After a review of possibilities, various activities were taken on by individual members.

(a) The most important activity of the Group during the next year will be the writing of sections of the Guide to Providing Social Science Data Services. In consultation with a member of the Data Archive Development Action Group, eleven people agreed to draft material, and another two volunteered as sub-editors.

(b) The development of a list of recommendations, addressed to potential researchers and data collectors, concerning study design as it relates to data management (the "Do's and Don'ts List"), first considered during the Group's sessions in Florida, has now been assigned to Rachel Des Rosiers and Terry Stewart.

(c) The proposal for an Inventory of Software relating to data management, also initiated in Florida, has been rather drastically revised, in view of the existence of other groups outside IASSIST who are reviewing social science software, the need for clearer definition of what kinds of functions and programs ought to be included, and the formidable amounts of time, energy and other resources required for a comprehensive job. It seemed wiser to adopt an incremental approach and start with two clearly delineated projects: a survey of the programs used at one institution (Columbia University), to be undertaken by Prina Grinberg as a pilot study; and an investigation of methods and available programs for cleaning multipunched data, to be done by Bill Gammell, Gary Grandon and Greg Morrison. Other such narrowly-focused efforts may be undertaken in the future.

(d) The Group felt it would be useful to gather information about the areas of technical expertise of the members of IASSIST, and to organize and publish the results. The aim would be to facilitate the use of the IASSIST membership as resource people or consultants for each other on technical problems. As a by-product, a profile of the members could be produced. The best procedure for obtaining the required information is under discussion. Gary Grandon and Sheldon Laube will be carrying out this activity.

(e) Another area of interest is the holding of workshops. Sheldon Laube will consider possibilities for organizing workshops at future IASSIST meetings, while Bill Gammell will investigate possibilities for regional workshops.

(f) Consideration of the Inventory brought home to the Group the importance of establishing communication with other organizations working on related problems. The co-ordinators, Bill Gammell and Greg Morrison, will conduct these external relations.

In addition to the above projects, for which individuals have taken on responsibility, two other issues were discussed. The first was the problem of information dissemination and exchange, crucial in the area of data organization and management, where relevant literature appears in a very wide range of publication from many different disciplines. We feel that this problem is particularly amenable to action by all IASSIST members, whether or not they are affiliated with the AG. We urge IASSIST members to send the co-ordinators of the Data Organization and Management Action Group and/or the Editor of the IASSIST Newsletter relevant bibliographical references to any material which relates to data handling, broadly defined. These references could be to published articles or unpublished papers, to books, technical reports, computer programs, items in newsletters, and so on. The reference should include a standard bibliographical citation, information on where to obtain the item if not found in readily available sources, and preferably an annotation or abstract. This information will then appear in the Book Notices section of the Newsletter. Also welcome would be short technical notes, resumes of procedures, documents, etc.; these ought to be sent to the co-ordinators for scrutiny and possible publication in the Newsletter.

The second general issue discussed was that of IASSIST endorsements. While the Group could see the value of IASSIST lending its support to projects in progress, it was not clear whether or not this should come as an official action and if so what process should be followed in determining what to endorse. A distinction was drawn between endorsements of good practices and endorsements of particular products; the latter possibility raises questions which might be considered by other Action Groups or by the Steering Committee. Short, however, of formal endorsements, members can provide feedback for the originators of projects, and engage in lobbying activities on an individual basis on its behalf.

DATA ACQUISITION

Canada- Pierre Lacasse, Centre de recherches en aménagement régional, Université de Sherbrooke, Sherbrooke, Québec

Europe- Marcia Taylor, Social Science Research Council Survey Archive, University of Essex, Wivenhoe Park, P.O. Box 23, Colchester, Essex, England C04 3SU

United States-

[No report submitted.]

DATA DOCUMENTATION

Canada- Dave L. Salley, Management and Central Services Group, Standards Division, Statistics Canada, Tunney's Pasture, Ottawa, Ontario K1A 0T6

Europe- Cees Middendorp, Steinmetzarchief, Kleine-Gartmanplantsoen 10, Amsterdam-C., Netherlands

United States- John Grasso, Office of Research and Development, Center for Appalachian Studies and Development, West Virginia University, Morgantown, West Virginia 26506

[No report submitted.]