

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- David Nasatir, Behavioral Sciences Graduate Program, California State College, Dominguez Hills, California 90747

UNESCO Directory of Data Services

The UNESCO Social Sciences Sector has passed a contract with the International Committee for Social Sciences Information and Documentation (ICSSD, 27 rue Saint-Guillaume, 75007 Paris) to establish a first Directory of Data Services. The ICSSD has sent out a complex questionnaire which some of our readers may already have seen. Philippe Laurent and Stein Rokkan discussed the project with Jean Meyriat of the ICSSD in November and tried to work out an arrangement under which the IASSIST Action Group for Data Archive Registry could be associated with this work, but this proved difficult under the contract established with UNESCO. It was agreed that the current project should take its course and be looked upon as a pilot phase. The further work of systematization, computerization and updating would be undertaken by BASS in co-operation with the Action Group for Data Archive Registry. A detailed plan for this follow-up project will be prepared during 1977 and submitted to UNESCO and other funding agencies.

In North America, efforts continue to survey existing directories of data libraries, archives, and data information services in order to determine the type of directory which will best meet both data information and general reference community needs and to determine how effective are existing directories, who uses them and with what success.

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 CO4 3SO
- United States-

As the amount of data available in machine-readable form increases, it becomes even more important for data services to address on local, national, and international levels questions of what should be preserved, by whom, and in what form. In North America, this group will study the needs for data services' collection statements similar to those produced by paper archives and traditional archives, appraisal of machine-readable data files, and de-acquisition.

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

In North America, this action group will begin to address minimum standards for the documentation of social science MRDF. IASSIST has received requests from both funding agencies and data information services and individuals to provide some formal means by which documentation may be evaluated. Some work has been done in this area and it should be possible at least to produce a check list in the near future. The special documentation requirements for process-produced data will be a prime concern of that AG.

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, Institute for Research in Social Sciences, Manning Hall, University of North Carolina, Chapel Hill, North Carolina 27514

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

Submitted by

Sue Dodd, University of North Carolina
Mohan Sharma, University of Alberta

Members Present

United States

Sue Dodd, Ch, University of North Carolina
Nancy Carmichael, Center for Social Indicators, Social Science Research Center,
Washington, D.C.
Debra Powell, DUALabs, Washington, D.C.

Canada

Mohan Sharma, Ch, University of Alberta
Elliot M. Avedon, University of Waterloo
Sharon Chapple Henry, Data Clearing House for the Social Sciences

Agenda topics for the C AG included (1) discussion of useful areas of concern and future coordination of tasks between US and Canadian AG members; (2) review of cataloguing efforts to date and discussion of any or all problems concerned with cataloguing tasks, including a review and discussion of the Working Manual for Cataloguing Machine-Readable Data Files [MRDF] compiled by Sue Dodd; (3) discussion of the ramifications of MRDF catalogue records, such as the national union list of social science data; shared cataloguing; cataloguing-in-production; the MARC II record as a standard format for storing automated bibliographic records of data files and its flexibility for an expanded record of information which would resemble a data abstract or study description; network interactive systems among data centers and libraries; and, an on-line search and retrieval system for the expanded MARC II-type record; (4) practical exercise in applying subject headings and descriptors for several large and uniquely held data sets, with a view towards compiling the beginning of the thesaurus or authority list of social science terms for data files; (5) discussion of existing printed and other available thesauri or authority lists in the social sciences, and a review of them in terms of future applications to data files; (6) discussion of the lack of adequate subject headings and sub-headings currently provided by the Library of Congress for social science data files, with a view toward providing constructive recommendations.

Given these broad topics of discussion, the classification AG decided on the following projects and sets of recommendations:

1. There are mutual areas of concern and projects on which the US and Canadian representatives could work together. There would continue to be an active exchange of information, with the next meetings scheduled for May 1977 in Toronto.

2. The AG's cataloguing project can be viewed as a tremendous success and as an example of a cooperative effort to evaluate both the ALA recommended rules and the Working Manual for cataloguing MRDF. (A total of 40 examples were submitted by 8 individuals from Control Data Corporation, Rutgers University, the National Archives, DUALabs, Yale University, University of Pittsburgh, National Opinion Research Center, and Drexel University.)

Each cataloguing effort will be individually identified and circulated among the classification group. Representative samples will be selected for inclusion in the Manual (with permission of the cataloguer).

The C AG will outline procedures for shared cataloguing of MRDF to avoid duplication of work and to maintain a high quality of effort. The cataloguing project will be extended to those now cataloguing data files or those wishing to be included in the project. The final report will follow the completion of the initial cataloguing effort and the final revision of the Manual.

3. The Working Manual for MRDF will be revised based on numerous suggestions by members of the C AG and by participants of the Working Conference. These recommendations along with the evaluation forms, will be circulated among the C AG members. The draft of the revised Manual should be ready for review by the C AG meetings in Toronto. [A further measure of success of this project is the possibility that the Manual may be accepted for publication by one of the library-affiliated associations.]

4. The C AG recommended that procedures for describing a data file, such as study description, cataloguing, and descriptors, (i.e., "cataloguing in production") be built into the daily routine of classification of any data file, and indeed, that these procedures be implemented at the early stage of data file creation.

In addition, the AG recommended procedures that could be directed toward the major funding organizations (e.g., NSF) which would be built into the contractual provisions: (a) data be documented according to some type of recommended standard; (b) information be provided for cataloguing; (c) data descriptors be applied; and, (d) data be designated to the appropriate archive within a reasonable amount of time.

5. The C AG will prepare two position papers reflecting the pros and cons of establishing a national union list of MRDF.

6. A bibliography of existing thesauri or authority lists for social science terms will be collected and reviewed in view of their adaptability to social science data files. In addition, representatives of the C AG will be in close contact with existing work in this area, such as the Committee on Conceptual and Terminological Analysis (COCTA). Any professional groups currently involved in compiling thesauri or conceptual lists of social science terms will be contacted to coordinate and share information. As a voluntary and non-funded group without expertise in this area, the C AG is not equipped to assume the task of creating a thesaurus of social science data files. However, it will attempt, in so far as possible, to provide guidance in this area and to share experiences gained from C AG associated projects to be carried out by members within their respective institutions. Two such projects are the compiling of an authority list for census and census-related data files [DUALabs] and an authority list of terms representing question/item level descriptors for Harris Public Opinion Polls [North Carolina Social Science Data Library].

7. A Standards and Quality-Control Review Board (SOQCRB) will be established within the North American Joint Classification AG to evaluate mutual projects, recommend procedures, establish standards, and work with other established organizations committed to or involved with the bibliographic control of MRDF within the social sciences.

One of the first tasks of this Board will be to provide guidance and examples of bibliographic references for social science data files. At the present time, there are no rules or a standard format for citing data in the published literature. The quality or amount of information varies among authors, editors, and students; often, the information is insufficient to enable a researcher to replicate the data file for the type of secondary analysis so important to the discipline. (For example, it is often very difficult to identify a data file, or its source, or data elements on which the published analysis has relied.) The standardization of bibliographic references for MRDF would pave the way for their inclusion into such reference work as the Social Science Citation Index.

Upon completion of this task, it was decided that a letter would be sent to the Standards Committee Z39 of the American National Standards Institute (ANSI) recommending an amendment to their forthcoming publication, The American National Standards for Bibliographic References. The amendment would be applied to this committee's method of citing "data files," which now provides examples for only one type of data (bibliographic data files) and which was created without any coordination with representatives of the library community who are establishing standards for cataloguing MRDF. The C AG feels very strongly that any standard citing data files should be compatible with the rules set forth in the Anglo-American Cataloguing Rules II [forthcoming].

The second task of the Review Board will be to compile a list of recommended subject headings and sub-division headings for social science data files. These recommendations would be forwarded to the head of the Subject Cataloguing Division of the Library of Congress. The work of the C AG and any resulting standards or reports should be given the widest possible coverage among the existing scholarly and professional journals. Someone will be appointed to write brief abstracts and notices of the work and will be responsible for circulating it in the available press.

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 United States-Canadian Action Groups on
Data Archive Development (DAD AG)

Submitted by

Alice Robbin, University of Wisconsin-Madison
Laine Ruus, University of British Columbia

Members Present

United States

Alice Robbin, Ch, University of Wisconsin-Madison
LaVerne D. Knezek, Texas Christian University [also DOM AG]
Judith Rowe, Princeton University
Peter Tolousis, Temple University

Canada

Laine Ruus, University of British Columbia
Edward Hanis, University of Western Ontario

Objectives of the Meeting were to (1) develop a conceptual framework for a publication, the working title of which is, A Guide to Providing Social Science Data Services, (2) revise, modify and elaborate a preliminary outline for this publication, (3) identify areas of responsibilities to be assumed by other Action Groups, (4) establish deadlines of completion for various sections of the Guide by contributing AGs.

The DAD AG defined the target audience for this publication as those individuals already providing or intending to provide data services for research, policy, and planning purposes. These individuals are typically in academic, research, and governmental organizations. The nature of these data is non-bibliographic, primarily quantitative, including micro- and macro-level of aggregation, but may also be textual.

The following chapters of the Guide were agreed upon: (1) Preface describing the target audience, purpose for the book, etc., (2) Introduction providing the historical context for the development of social science data services, (3) functional overview defining a social science data services system and its components, (4) detailed functional description of the system which includes management or technical processing and user services aspects, (5) organizational models and services provided, (6) organizational development including resource requirements and scenarios for the development of services, (7) glossary of terms, (8) extended bibliographic references tied to each of the sections, (9) detailed index, (10) appendix of procedural, management, and administrative forms, with introductory remarks on the limits and usefulness of recordkeeping.

The preliminary outline prepared for the meeting was revised and will be circulated to all AG members attending the joint conference. Comments will be forwarded to AG coordinators, in time for the Toronto meetings. Further modifications based on recommendations made during the Toronto meetings will result in a detailed outline whose projected completion date is June 30, 1977. At that time, individuals in the various AGs will assume responsibility for the substantive topics/sections within each chapter. Alice Robbin will act as general editor for chapters (1) through (6); Laine Ruus, for chapters (7)

through (10). The DAD AG plans to have a first draft completed by August, 1978, and a second draft by December, 1978. Discussions are now underway with a potential publishing house who has indicated an enthusiastic interest in the project and has already offered useful advice.

A second project [discussed in greater detail by K. Heim in the "Book Notices" section of this Newsletter] is now fully underway. This is the production of an annotated bibliography, whose provisional title is Perspectives in Social Science Data Services: An Annotated Bibliography. Thus far the DAD AG has gathered almost 200 articles relating to this topic. A letter was sent to all data archives last fall asking that they participate in this report. The DAD AG noted the importance of a collaborative effort, and therefore requests assistance: we would like every IASSIST Newsletter reader to search his/her files for all pieces of literature (articles, unpublished papers, conference addresses, parts of books) concerning data archiving (retrospective and current) and to send citations to the IASSIST Newsletter editor. All respondents will be acknowledged in the bibliography.

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 (NRR), Washington, D.C.
20408

Report of the Joint United States-Canadian Action Groups on Process-Produced Data (PPD AG)

Submitted by

John DeVries, Carleton University
Donald F. Harrison, National Archives, Washington, D.C.

Members Present

United States

Donald F. Harrison, Ch, National Archives, Washington, D.C.
Charlotte Boschan, National Bureau of Economic Research, New York
Harriet Dhanak, Michigan State University
Shirley Gilbert, Princeton University [also DOM AG]
Elizabeth Powell, LEAA/NCJISS, Department of Justice, Washington, D.C.

Canada

John DeVries, Ch, Carleton University
Tony Falsetto, Public Archives of Canada, Ottawa

Ex Officio

Carolyn Geda, ICPSR, University of Michigan

Definitional problems were first addressed by the members. The following Definition of Process-Produced Data was agreed upon.

Processed-produced data are data which are products of routine, administrative activities of private or public institutions or persons. They can be explained with respect to the functions of their origin: either for administrative or policy uses; can be merged, linked, abstracted, or summarized; and, may be used for research and/or statistical analysis.

The Paul Müller paper [see edited version in the Newsletter issue] was reviewed in order to arrive at the Mandate of the Action Group.

Generally, we have asked ourselves, "How can we facilitate the movement of such data from the originating institution or person to the secondary user?" To do this, the AG explored areas of concern, some of which were rejected, and some where it was felt that the AG might logically make a contribution.

We rejected the following:

1. Laws regarding confidentiality: This subject logically should be a matter of concern for this AG. But, because there are so many other institutions/professional groups working on the same subject, the AG prefers to await their reports before proceeding in order to avoid duplication of work. The AG members are particularly anxious to await the final report of the project of Ed Hanis and Dave Flaherty at the University of Western Ontario, which deals with this issue. [Hanis and Flaherty are examining privacy/confidentiality problems as they are dealt with by Central Statistics Bureaus in West Germany, Sweden, Great Britain, Canada, and the United States. A final report is due at the end of this year.]

2. Linkage techniques: A subject as complex as this requires so much technical and concentrated study, that the AG considers this an area of concern for another or a new AG.

3. Uses made of machine-readable files: the AG rejected this subject as having no useful purpose.

The following projects were accepted as within the mandate of the AG:

1. A Directory of Catalogues which List Data Bases (Inventory of Inventories) First priority. The AG anticipates presenting IASSIST with a first draft in Toronto in May, 1977. By February, 1978, it is expected that the Directory will be presentable for publication in the IASSIST Newsletter or in some other suitable journal, after proper IASSIST backing.

2. An Inventory of Existing Guidelines in Use by the Originating Institution or Archival Institution for: (a) documentation and (b) preservation (not to include acquisition activities). Second priority. The AG will begin to collect and disseminate the information within the JAG immediately, and plans to present an outline of the method for compilation by May 1977 in Toronto. A first draft should be available for presentation at Uppsala in August, 1978.

3. An Inventory of Procedures in Use by Originating and Archival Institutions for (a) documentation on processes; and (b) management decisions. Second priority. Like Project #2, the AG will begin to collect and disseminate almost immediately those procedures. Plans are to present an outline by May, 1977 in Toronto. If progress is made, a first draft should be completed in time for the Uppsala meetings.

4. An IASSIST Publication of Desirable Documentation Components Necessary to Service, Retrieve and Otherwise Manipulate a Process-Produced Data Base. Third priority. Although extremely important, it logically follows that such a publication cannot be produced before first surveying the field and consulting practitioners. Plans are to present an outline for consideration at the Uppsala meetings.

These projects have been arranged in priority, not by their relative importance, but in consideration of their reasonable and logical sequential order for work. All projects except the last one will be prepared separately by national action groups, each consulting with institutions in their parent country (i.e., the USAG will prepare lists of inventories published or originated in the United States, and so forth).

Project #/ Priorities	May 1977	February 1978	August 1978
1/1	First Draft	Finished Product	
2/2	Outline	Outline	First Draft
3/2	Outline	Outline	First Draft
4/3			Outline

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, University
of Essex, Wivenhoe Park, P.O. Box 23, Colchester, Essex, England C04 3SU

United States- William Gammell, Social Science Data Center, University of Connecticut,
Storrs, Connecticut 06268

Report of the Joint United States-Canadian Action Groups on
Data Organization and Management (DOM AG)

Submitted by

Bill Gammell, University of Connecticut
Greg Morrison, Carleton University

Members Present

United States

William Gammell, Ch, University of Connecticut
Peggy Cahn, Federal Reserve Bank of New York
Shirley Gilbert, Princeton University
Pnina Grinberg, Columbia University
Gary Klass, State University of New York at Binghamton
LaVerne D. Knezek, Texas Christian University [also DAD AG]
Sheldon Laube, C.M. Leinwood Associates, Newton, Massachusetts
Barbara Noble, Center for Advanced Computation, University of Illinois
Pat Peterson, Agency for International Development (AID)
Elizabeth Powell, LEAA/NCJISS, Department of Justice, Washington, D.C. [also
PP AG]
Richard C. Roistacher, Center for Advanced Computation, University of Illinois

Canada

Greg Morrison, Ch, Carleton University, Ottawa, Ontario
Rachel Des Rosiers, Canadian Data Clearing House for the Social Sciences

Ex Officio

Carolyn Geda, ICPSR, University of Michigan

During the conference, the following activities were carried out:

1. Began compiling an Inventory of Software Relating to Data Organization and Management. This effort is viewed as an on-going project. The coordinators will complete a first draft to be distributed and approved during the IASSIST May meetings in Toronto. The Inventory will be published in the Newsletter. Contributions to this list are encouraged.

2. Discussed advanced developments in data management. Richard Roistacher reported on the development of the Data Interchange File concept and on the use of computerized document processors for the creation and management of machine-readable codebooks. Sheldon Laube reviewed the capabilities of existing data cleaning software and described the design and development work now being done on a generalized data cleaning system. The AG will be active in providing input to these projects.

3. Reviewed with representatives of the Data Archive Development AG the outline for the Guide [see description under DAD AG report] and discussed possible contributions to this project. Copies of the revised DAD AG outline will be circulated to group members in time for extensive consideration at the Toronto meetings.

4. Planned several projects: (a) Development of a list of recommendations addressed to potential researchers on study design as it relates to data management (the "do's and don'ts" list). Group members are to send in their suggestions to the DOM AG coordinators. A consolidated list will be produced in Toronto for publication in the Newsletter and elsewhere. (b) Collation of information on relevant monographs, technical reports, program writeups, etc., which the AG members would like to share. The intent of the collection is to publish this information in a "What's New In Data Management" section of the Newsletter.

5. Prepared an agenda for the Toronto meetings; items for that agenda have been referred to above.

The DOM AG coordinators would like to suggest the following revised mandate for future discussion at international IASSIST meetings:

"This group addresses the problems of data organization and management confronting those archiving or using social science data. The AG will investigate and evaluate software and procedures for data and documentation preparation and management; recommend guidelines for preparation procedures and software development; and, sponsor workshops and seminars for professional training and the exchange of information in these areas."

[The original mandate includes "hardware" as an area to be addressed by this AG; see Newsletter Volume 1, Number 1 for the full text of the mandate.]

AN OVERVIEW OF PROBLEMS ASSOCIATED WITH PROCESS-PRODUCED DATA / paul müller

For the August 1976 IASSIST meetings, Paul Müller prepared a report which provided a broad overview of problems associated with process-produced data. What follows is an edited version of this report. [Future issues of the Newsletter will contain additional Action Group reports. The membership is encouraged to begin a dialogue on this and other issues of concern to the data archive community.]

"Administrative Bookkeeping as a Social Science Data Base"

by
Paul Müller
Institute for Applied Social Research
University of Cologne

1.0 [...] I will try to give a rather broad overview of problems associated with the "production, acquisition, preservation, processing, distribution, and utilization of machine-readable" process-produced data. This report is necessarily biased by my own viewpoints and experiences; other experiences may well be fundamentally different. But, the function of this paper is to initiate discussion and later, joint actions.