PAPERS PRESENTED AT THE SECOND IASSIST NORTH AMERICAN MEETING, MAY 11 – 12, 1977

[Editor's Note: The following are abstracts of papers presented at the Second IASSIST North American meeting. The full proceedings of the Conference will be available in the near future. All participants of the Conference will receive a full copy of the Proceedings; IASSIST members will receive a copy at a reduced rate; all other individuals should contact Sharon Henry of the Data Clearing House for the cost to non-IASSIST members.]

IASSIST - Progress Towards Solving Problems in Data Archiving

Carolyn Geda
Inter-university Consortium for Political and Social Research

The paper reviews the establishment and development of IASSIST including the rationale for such an Association. Problems encountered during the formation of the International Organization are presented with interim proposals for dealing with them. The governing body, regional structure, and functions of the action groups are discussed. A brief review of Regional activities is given and potential directions for the Association are projected.

Canadian Secretariat Report

Rachel Des Rosiers
Data Clearing House for the Social Sciences

The report summarizes the membership campaign for IASSIST in Canada. It outlines the results as well as expected plans to increase the membership in Canada, e.g., by contacting related associations in the field and publicising IASSIST information in professional journals and at conferences. It also discusses its relationship with the American secretariat, especially in regard to Action Groups' projects and plans. Finally, it invites comments and suggestions from the Canadians present concerning its intended role and functions.

IASSIST Secretariat Report: United States

Judith Rowe
Princeton University

The report of the United States Secretariat covers the current status of membership enrollment as well as proposed membership recruitment activity. Specifically, it addresses the goal of recruiting for Action Group activity all of the professional staff of each data archive and data library. It includes a general summary of both secretariat and Action Group activity, as well as a report on the first North American Action Group Conference in Cocoa Beach. Other topics include an initial report on the plans for next year's conference and some activities and programs planned by other organizations which would be of interest to IASSIST members. Among the latter are the annual conferences of Special Libraries Association (SLA), Association of Public Data Users (APDU) and the American Association for Information Science (ASIS), as well as the Inter-university Consortium for Political and Social Research (ICPSR) workshop for data librarians.

Data Archive Registry-Survey of Past Effort,
Suggestions for the Future

Lisa Lasko
Canadian Consortium for Social Research

The mandate of the IASSIST Action Group for Data Archive Registry is described. The major portion of the paper is devoted to a brief review, description, and evaluation of some of the more important existing directories. These are: Social Science Data Archives in the United States, published by the Council of Social Science Data Archives in 1967; A Directory of Information Resources in the United States: Social Sciences, 2nd edition, published by the Library of Congress in 1973; the second edition of the Encyclopedia of Information Systems and Services, edited by Anthony Kruzas and published in 1974; and the Directory of Data Bases in the Social and Behavioral Sciences, edited by Vivian Sessions and published in 1974.

Recent developments in data archive registries, such as the UNESCO sponsored Directory of Data Services and the Directory of Data Centres, to be published by the Data Clearing House for the Social Science in Canada are also discussed. Essential elements and user requirements of social science data archive directories are dealt with in reference to these directories.

Finally, a summary of the problems facing the Data Archive Registry Group, both abroad and in Canada, is given, along with several proposed options include a discussion of the viability of a Data Archive Registry Group in Canada; unique contributions to be made in this area, one of which might be a compilation of a list of subject headings for social science data archives; and, the production of a directory of data archive personnel.

Problems in Handling Process-Produced Data

John DeVries
Social Science Data Archive
Carleton University

The paper is an attempt to outline the problems one encounters in the handling of process-produced data, starting with their definition through the stages of obtaining information about them and acquiring them, to the phases of transforming and analysing them. In addition to enumerating and discussing the often unique problems with which the user of process-produced data is faced, I will try to mention various projects which have already begun, and indicate future plans by the IASSIST Action Groups on process-produced data.

Definitions. In the last few years, various definitions have been suggested. Although there is considerable overlap between them, the remaining differences make it important that we come to some agreement about the proper "domain" for an Action Group on process-produced data, and thus to a clear definition which unambiguously specifies this domain. I will discuss the definitions of which I am aware, and present some arguments in support of the definition which was developed at the Cocoa Beach conference.

Information and acquisition. Generally, we will find that process-produced data have been generated by governments and other public organizations. Problems in this phase are thus to a large degree subsumed under the general heading: "The relationship between the research community and the government." Specific issues I will discuss include: privacy versus the right to information, i.e., the government as a research resource; and, obtaining information from governmental agencies.

Documentation. In many non-trivial ways, process-produced data differ from surveys. Some of the unique aspects regarding documentation are: definitions of the universe to which the data pertain, description and estimates of the errors of coverage and content, and requirements for a "codebook".

Data transformation. Once a data set has been acquired by a researcher or an archive, there are various difficulties one faces in transforming and analyzing the data. Many data sets are "ragged" rather than rectangular, thus requiring either intricate transformations or specialized statistical packages. Aggregated data require a high level of statistical sophistication on the part of the analyst. Finally, one will frequently wish to link data from various data sets. In doing so, one often runs into problems of comparability: stimuli may not be standardized across data sets; for identical or comparable stimuli, response categories are frequently not identical across data sets; different data sets may not relate to exactly identical universes; finally, where spatial and/or temporal delimiters are involved, they are often not identical across data sets.

Data Acquisition for Archives

Pierre Lacasse
University of Sherbrooke

The purpose of this paper is to raise a number of issues about data acqui-
sition in order to narrow down the subject to the most important avenues for
further discussion. With so diversified stock of social data being created, we
need to ask "who will keep what, and on what basis"? To clarify this question
we will particularly discuss three major inputs to the subject: (1) The col-
lectors of data: Who are they? What types of data are they collecting? What
data are we interested in? (2) The archives: What are they? What are their
purposes in holding data files; and for what types of secondary users? What
mandate do they have and what is their scope of interest? (3) The users for
whom data is held: Who are they? Where do they stand? On what issues? What
are their needs? How will they use the data held by the archives?

After discussing these questions, we will take a look at what has been
done in the data acquisition action group. In particular we will discuss the
questionnaire prepared by the European section of IASSIST on "Archives data
acquisition policies and problems."

Consider the group's mandate, which requires proposing ways of linking
collectors and archives, with the stress on defining the group objectives and
conditions which must be met in order to close the gap between collectors and
archives. Finally, we will try to suggest some possible courses of action in
the short and long term.

Organization of Data Archives

Laine Ruus
University of British Columbia

The mandate of the IASSIST Data Archive Development Action Group is two-fold:
the creation of a "procedures manual consolidating current archival organization-
al, administrative, and personnel structures, procedures, and policies...to aid
developing archives", and the organization of training workshops and seminars to
aid in personnel training and professional development. The efforts of the Data
Archive Development Action Group have been entirely concentrated on the develop-
ment of A guide to providing social science data services. While response to
past workshops has indicated the need for this guide, plans are for the moment in
abeyance. Problems in data archive development occur at three levels. Intra-ar-
chival problems are primarily administrative: planning and policy making, staff,
acquisition, data management, technical and user services. Intra-institutional
problems, impinging on the former, occur in the areas of planning, inter-depart-
mental cooperation, support services, and funding. Problems at the inter-archi-
val level result primarily from lack of adequate standards and conventions and
from underdeveloped formal channels of information dissemination.

Standards for Data Documentation

Dave Salley
Statistics Canada

In recent years, a number of initiatives related to the concept of data docu-
mentation have been undertaken within the Canadian Federal Government. The motives
for these initiatives have centered around the need to inform users on data avail-
ability, the need to monitor duplication of data collection, the need to control
questionnaire content and the need to assess the costs incurred in the collection,
compilation and dissemination of federal information. While these motives are not
exhaustive, they do represent a significant class of problems for consideration by
the IASSIST Action Group.

The shortcomings of most efforts to date have centered on the particular or
special purpose nature of the systems. In many cases, a particular subject area
or class of users restricted the approach severely. In short, little considera-
tion has been given to the development of standard components of a Data Documen-
tation system with a view to meeting general requirements. The reason for this
problem may well be that information managers in the Federal Government have not
yet developed a real appreciation of the need for standard documentation proce-
dures, both as a tool for managers and planners of information activities and to
assist "end data users". Thus, most initiatives have been quite ad hoc and nar-
row in scope. However, one should not underestimate the useful technical ap-
proaches and methodologies developed to date.

As a departure point for the Action Group, it is suggested that Terms of
Reference be developed which contain a strong "Standards" ingredient. Further-
more, an intensive needs and benefits study should be undertaken with respect to
the whole data documentation issue. Certainly, the group should bring together
current work in a systematic fashion, but the development of standard approaches
in the scope, technique, uses and purposes of data documentation systems will
have far more useful results in the long run.

Cataloguing and Classification of Machine-Readable Data Files;
A Preliminary Report on the US IASSIST Classification Cataloguing Project

Sue Dodd
University of North Carolina

The primary emphasis of the US Classification Action Group of IASSIST has
been on establishing standards and on the study of library information systems
as they may apply to social science data files. Some of the recent developments
within the library system which the Classification Group is examining include:
1) the development of rules and guidelines for cataloguing machine-readable data
files (MRDF); 2) the development toward the acceptance of the MARC (Machine-Read-
able Catalog) record format as a universal standard for the automated biblio-
graphic record; 3) the development of networks and on-line information systems
which allow for multiple input and immediate retrieval of information; 4) the
development of thesauri and controlled vocabularies for social science terms;
and, 5) the development towards future considerations of a national union list
of available MRDF and their location.

Given the current work and interest in cataloguing MRDF, the first task of the IASSIST Classification Group was to participate in an organized project designed to test the feasibility of cataloguing social science data files according to the American Library Association (ALA)'s designated Subcommittee to recommend rules for cataloguing MRDF.

To facilitate the test, which was conducted by mail, a "Working Manual for Cataloguing Machine-Readable Data Files" was compiled based on an interpretation of the ALA Subcommittee's recommendations. The task required that the participants select six data files, either numerical, text or program files, and proceed to catalogue these data using the information and guidance in the manual; apply subject descriptors on the content of the files; and complete an evaluation form. The outline of the project, the actual test, and the preliminary results will be discussed in this paper.

Data Organization and Management
Applications in Data Archiving

Greg Morrison
Social Science Data Archive
Carleton University

The formal mandate of the Data Organization and Management Action Group of IASSIST is used as the starting point for consideration of various kinds of problems with which the Group might concern itself -- problems which most data archivists will confront sooner or later. These include: the transfer of datasets; the cleaning, editing, transforming, merging and sub-setting of data; the organization of data files for the above activities and for statistical analysis; the technical aspects of the documentation of individual datasets by codebooks; and, the technical aspects of the documentation of collections of datasets by catalogues or inventories.

While an obvious organizing focus of the Group is computer software, it will try to define appropriate procedures in these areas, and to encourage the exchange of information about programs and procedures. The activities of the Group to date are reviewed, and some suggestions are made for the future.