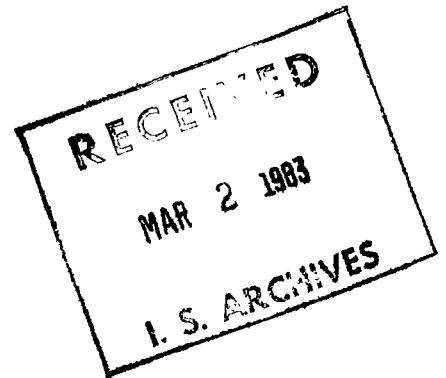51225

IDRC Doc No.: 343

A Co-operative Inter-Agency Research Activity Information System

Feasibility Study

Prepared for the

International Development Research Centre
Ottawa

February 1, 1983

Donald F. Thompson
Consultant
3151 Kingsley St.
Victoria, B.C. V8P 4J5   Canada

RESARC
002 338
T S

This study was carried out on behalf of the following agencies:

BOSTID: Board on Science and Technology for International Development

GATE: German Appropriate Technology Exchange

IDRC: International Development Research Centre

IFS: International Foundation for Science

NUFFIC: Netherlands Universities Foundation for International Cooperation

SAREC: Swedish Agency for Research Cooperation with Developing Countries

1. Introduction

This report examines the feasibility of a computer-based system for sharing information on development research activities funded by the members of an international group of research funding agencies. The report arises out of a need expressed by those agencies to have available a common store of information from which they could determine whether there were duplications in their efforts or possible areas of inattention, and to help them identify sources of expertise and knowledge within the agencies themselves and in the developing world.

The study was undertaken by an IDRC consultant on behalf of all of the agencies. After examining some similar attempts at donor-based cooperative information systems, as well as a number of active recipient-based research registers, a background paper was produced and distributed to the six agencies. Each agency was then visited in order to find out how much information was available, how it was organized within each organization, and how experienced the individual agencies were in the business of information systems. What follows documents those findings, and makes a series of recommendations as to a possible course of action.

2. Methodology

As a starting point for the discussions with each agency, a model was suggested in which each discrete research activity (frequently called a "project") would be described in a "record" composed of a set of carefully defined "data elements" (Appendix A). The records would be stored in a computer where they would be organized so that those dealing with a specific subject area, geographic area, recipient or granting institution, etc. could be called up and displayed at an on-line terminal. Records could also be selected and then printed, complete with indexes, in a form suitable for publication. An added feature would allow each agency to have "private" information stored in the computer in order to fulfil some of its own information-processing requirements. The on-line feature was suggested on the premise that it would provide more selectivity and timeliness; a service relying solely on printed products was felt to be too unresponsive to day-to-day needs.

The general response was one of considerable enthusiasm for such a system. It was recognized as being capable of providing an up-to-date and accurate source of information, albeit limited by the number of participating donor agencies. There was a consensus that funding agencies from outside the present group should be encouraged to join any such system. Although existing recipient-based research registers are not heavily used within the agencies, it was generally felt that this donor-based system, with a searching capability, would provide a much more useable and useful source of information.

At least two methods are available for entering new and retrospective data into a common system. The more traditional conventional method is for each agency to complete its records in typescript on "worksheets" and forward these to a central node of the network where the data would be entered into the database. Another method, made possible by less costly international telecommunications, is to enter the data "on-line" directly from a terminal in each contributing agency. In our discussions there was considerable interest in the latter, since it would provide information on a more timely basis, and reduce dependance on both the central node and the postal system. A further alternative might be local data entry onto diskettes, with these being sent to a central node. The adoption of "on-line" data entry for the group would not, however, preclude any one agency from choosing the "worksheet" route, or any other reasonable alternative.

Within this group of funding agencies there is a wide range of information-handling activities and capabilities. The differences are shaped by different needs, different organizational structures and different attitudes toward the kind and amount of information that should be collected. It manifests itself in a number of different methods for gathering, storing and displaying corporate data. In spite of these differences, there was a common acknowledgement of the utility of shared information. Futhermore, we were able to identify a common core of information about each research activity which could be supplied by each agency. This is documented in Appendix B.

3. Recommendations

Based on the information available at this time, it is reasonable to propose a cooperative information system, in which each member agency would be responsible for providing data about each of its research activities. The exact definition of "research activity" would be left to the individual agency, although guidelines would be set down by an inter-agency committee. This common system would provide the following capabilities:

1) On-line information retrieval from the common data base. This might be provided by a local computer, or via the international telecommunication networks.

2) Specialized products with content and format to suit the needs of individual member agencies. These might be in the form of project lists, institution lists, researcher lists, or indexes of subjects, geographic areas, etc., and would be suitable for publication.

3) The ability to add further funding agencies as this becomes desirable.

4) Repatriation of each member's information in a standard data communication format.

If such a system is to operate successfully it will be necessary to carry out the following actions:

1) Guidelines must be established to identify what types of research activities will be reported to the common system.

2) A common set of data elements must be agreed upon. (This should conform as closely as possible with the UNISIST standard: Reference Manual for Machine-Readable Descriptions of Research Projects and Institutions. Paris: Unesco, 1982.)

3) Indexing rules must be established.

4) Each agency must organize itself to produce the agreed-upon information for all discrete research activities.

5) The agencies must collectively decide whether the information contained in the system is to be freely available to any bona fide inquirer, or whether specific restrictions are to be applied. If so, these restrictions need to be defined.

6) Each agency should consider how it will exploit the common system to meet its own needs, and make provision for the training of staff to this end.

7) A computer and appropriate software must be identified to host the common system, permit the management of the records and enable agencies to selectively retrieve them. Telecommunication links must be specified to permit on-line operations.

8) Each agency must provide the necessary equipment to interface with the chosen system.

9) Technical personnel must be identified to co-ordinate the whole of the system, and to help agencies organize their activities in such a way that they mesh with the proposed system.

The amount of information to be collected (see Appendix C) suggests that the implementation of this kind of system need not be an overly cumbersome task. However, because of the wide range of information handling capabilities within the funding agencies surveyed (see Appendix D), it will require a considerable coordination effort in order to bring all of the organizations to that common point necessary for the successful implementation and operation of such a system.

4. Required Resources

Because no decision has been made on the implementation, it is difficult to give exact figures for required resources. However, the following tabulation of the tasks to be carried out, and the resources needed to perform them, gives some idea of what would be required.

Suggested System Implementation Schedule

		Person Weeks	Elapsed Weeks	Total Weeks
1.	Individual agencies agree to participate			0
2.	Technical paper outlining decisions to be made respecting operating environment, data field definitions, general technical matters	3	3	3
3.	Distribution of above for reactions from agencies, preparation for meeting	1	6	9
4.	Meeting to decide on technical matters and determine implementation schedule	1	1	10
5.	Document final system design, write user manuals	2	2	12
6.	Implement system in each agency, which includes obtaining required equipment, developing necessary in-house information flows, and in-house user training	9	13	25
7.	Begin entering current data, collecting and organizing retrospective data for entry	6	9	34
8.	Initial system operation	13	13	47
9.	Interim evaluation	1	1	48
		36	48	48

The personnel requirements are for one or more persons able to advise and act on the following matters:

- database design
- system operation, both hardware and software
- communications systems
- documentation writing

The person-weeks estimate assumes fairly smooth implementation, and that equipment will be available when scheduled. It does not include any overhead or travel time. No time is included for implementation of "private", i.e. agency-specific, databases, as this will vary with user requirements.

The personnel requirements shown above do not include those required within each agency. This will vary greatly, depending on the organization involved, how much information is to be organized, and how much it will have to be re-organized. Appendix C shows estimates of the amount of information being considered. Each agency will also have to provide some equipment, and this will vary according to their needs and present situations.

Costs other than personnel will include travel, computer services, and general overhead. The latter will include communication costs. The exact nature of all of these will not be known until a decision has been taken on an implementation strategy.

5. Conclusions

It is clear from this survey that support systems for a shared information service are readily available. The technology exists to implement the database on a minicomputer or larger machine, and to support international data communications between computers and agencies at a reasonable cost. The methodology for cooperative information systems development has been demonstrated in the United Nations environment (INIS and AGRIS, for example). Data standards exist in the form of the UNISIST standard.

The remaining and essential ingredient for success is the desire to create this system and use the information therein. The participants in this study indicated that such a system would be useful to the agencies themselves, and at least as useful to the developing world. The enthusiasm expressed for the concept must be translated into an institutional commitment on the part of each participating agency, to work as an active member of a cooperative system.

Appendix A: Proposed Data Elements for a Record Describing a
               Research Activity ("Project")

1. Project Title

   A descriptive title, containing unambiguous, informative terms.

2. Granting Institution

   A uniform code or name for the institution granting the funds, and
generating the project record. If more than one member contributes to
the same project, a record will be generated for each member.

3. File Number

   The file number used by the granting institution.

4. Geographic Area Under Study

   The country or region under study. Country names will be drawn from
some authority, for example the OECD Macrothesaurus for Information
Processing in the Field of Economic and Social Development.

5. Subject Area

   A term or terms describing the subject area under study, e.g.:
biology, population, etc. Terms may be drawn from some sort of
authority (Macrothesaurus, Broad System of Ordering, etc.), or
generated as needed by each institution. Some organizations are
already using an authority or variant thereof. If different members
use different authorities, searches will have to take this into
account, possibly chosing different terms for different institutions.
A single authority is strongly recommended.

6. Type of Activity

   A term or terms describing the type of activity being funded. These
can be drawn from some authority.

7. Location of Research

   The geographic location(s) at which the research will be carried
out. An authority list of countries can be used to choose these
terms.

8. Abstract

   A narrative describing the project, its aims, objectives and
background.

9. Contact Person

An individual or office within the granting organization, to whom requests for further information can be directed.

10. Funding

The total funding provided by the institution generating the record. If this is a co-operative project, this field will show only the amount contributed by the donor supplying the record. If there is more than one grantee involved in this project, this will be the total amount provided for all grantees by this donor. This amount should be provided in some uniform currency, or a field should be provided showing currency. It may be desirable to break this amount down into different classifications, e.g.: training, consultants, overhead, etc. It may be necessary and desireable to provide a field showing the funding breakdown over time.

11. Duration

The duration of the grant, in months.

12. Commencement Date

The commencement date of the project, in the form YYYYMMDD, where YYYY is the year, MM is the month of the year, and DD is the day of the month. The use of this format allows for arithmetic comparisons of dates.

13. Notes

Free text notes of possible interest, as defined by the agency supplying the data. Eg.: early termination, extension, etc.

14. Other Grantors

The names and level of support of other institutions contributing to this project. A standard form must be adopted, to avoid problems with acronyms and varying useage.

15. Related Donor Record

A unique identifier of any other record in the system describing a project supporting the same research activity. This can be generated by searching the available data at the time the record is entered into the system.

16. Project Publications/Products

A free text description of any of the products of the research, e.g.: patents, reports, monographs, etc.

17. Project Appraisal

A free text appraisal of the success, failure, or otherwise, of the project.

18. Status

A term indicating the status of this project: PROPOSED, ACTIVE, or COMPLETED.

The following fields are to be repeated for each grantee:

19. Institutional Grantee Name

The name of the institution at which the research will be carried out. This should be from an authority file where possible, and some choice will have to be made as to which language will be used in the name.

20. Individual Grantee Name

The name of the principal researcher carrying out the work.

21. Grantee Funding

The level of funding for this grantee, from this funding agency, expressed in some uniform currency.

Appendix B: Availability of Data by Funding Agency

This table shows the availablility of individual data items within each agency. The data items referred to are those suggested in the Inter-Agency Project Information Network Discussion Paper. They are described in detail in Appendix A.

Data Item Name	BOSTID	IDRC	IFS	SAREC A	SAREC B	GATE	NUFFIC
1 Title	Y	Y	Y	Y	Y	Y	Y
2 Agency	Y	Y	Y	Y	Y	Y	Y
3 File Number	Y	Y	Y	Y	Y	Y	Y
4 Geog. Area	Y	Y	Y	Y	Y	Y	Y
5 Subject	Y	Y	Y	Y2	Y2	? 2	? 2
6 Activity	Y	Y	Y	Y	Y	?	Y
7 Res. Location	Y	Y	Y	Y	N?	Y	Y
8 Abstract	Y	Y	Y	Y	Y	Y	?
9 Contact Person	Y 4	Y 4	Y	?	Y	Y	?
10 Funding (Project)	Y	Y	Y	Y	Y	Y	Y
11 Duration	Y	Y	N/A	Y	Y1	Y	Y
12 Start Date	Y	Y	Y	Y	Y	Y	Y
13 Notes (5)	-----	-----	-----	-----	-----	-----	-----
14 Other Grantees	Y	Y	Y	Y?	N/A	?	?
15 Related Record	Y	Y	Y	Y	N	Y	?
16 Documents/Product	Y	Y	Y 3	Y	Y	N	Y ?
17 Appraisal	?	N	N	N	N	N	N
18 Status	Y	Y	?	Y	Y	?	?
19 Grantee	Y	Y	Y	Y	Y	Y	Y
20 Researcher	Y	Y	Y	Y	Y	N	?
21 Funding (Grantee)	Y	Y	Y	Y	?	Y	Y

Notes:

Y   Yes, information can be supplied
N   No, information is not available
?   Don't know if it can be supplied, or doubtful that it can be

A   SAREC Funded Research In Sweden
B   SAREC's Bilateral Funding programme; some multilateral projects may be reported with the same data availability

(1) may be difficult
(2) subject to in-house capabilities, if this is thesaurus indexing
(3) general availability is possible
(4) not very useful because of staff changes
(5) this field contains user-defined data

Appendix C: Data Records by Institution

Approximate Number of Records in an Inter-Agency Database

	Retrospective	Annual
BOSTID	150	40
GATE	150 (2)	20
IDRC	1 400	250
IFS	600	80
NUFFIC	50 (1)	10
SAREC	500 (3)	300 (4)
Total	2 850	700

(1) estimate only, for International Education Projects
(2) current plans call for collection of this limited number of
     retrospective records at this time. Period covered is 1980
     to present.
(3) this is the number of projects documented in the 1981 edition
     of Development Research in Sweden. A 1982 edition is in
     progress.
(4) SAREC funded research in Sweden (approximately 250) plus a
     rough estimate of bilateral funded programs.

Appendix D: Agency Profiles

1. Existing Project Reporting

   BOSTID - no existing project reporting system for any of the
            established programs; one could easily be implemented for
            the newly established Research Grants Program. Content of
            existing information easily conforms to suggested data
            elements.

   GATE -   project register for current research projects in
            progress, across all of GTZ, carried out by consultant.
            No regular data collection program exists, although plans
            are being formulated to collect data twice each year.
            Data content conforms to suggested data elements in all
            significant areas.

   IDRC -   project information collected and stored in computerized
            database for all projects funded by Centre. Content
            conforms to suggested data elements, although formats
            differ for some items.

   IFS -   project information collected and stored manually,
            published as IFS Work, and in somewhat briefer form as the
            IFS Directory. Latter produced by computerized service
            bureau.

   NUFFIC - project information on International Education program
            collected for publication in annual report. Most required
            data elements are present or readily available.

   SAREC - information on SAREC funded research in Sweden now
            collected and published annually. No information on
            research funded under the bilateral program is now
            collected, as the funding decisions are made by national
            bodies withing the cooperating country.

2. Language of Project Descriptions

   BOSTID - English

   GATE -   German, English

   IDRC -   English, French

   IFS -   English

   NUFFIC - English

   SAREC - English

3. Local Needs possibly serviced by specialized outputs from common system

    BOSTID - making data available for processing on in-house IBM Personal Computer; production of listings for projects list

    GATE - production of projects list in German and English

    IDRC - own system now produces necessary products

    IFS - production of both Directory and IFS Work

    NUFFIC - none perceived

    SAREC - production of SAREC Funded Research in Sweden, and possibly project list, if decision was taken to document individual activities funded by cooperating countries.

4. Information Available in Machine-readable Form

    BOSTID - none

    GATE - none

    IDRC - all, various formats

    IFS - IFS Directory on tape, format unknown

    NUFFIC - none

    SAREC - none

5. Thesaurus Indexing Experience

    BOSTID - considerable, using OECD Macrothesaurus

    GATE - some, using OECD Macrothesaurus, considerable with DSE Makrothesaurus (in GTZ library), and SATIS descriptors (in GATE library).

    IDRC - considerable, using OECD Macrothesaurus

    IFS - none

    NUFFIC - none

    SAREC - some, using OECD Macrothesaurus

6. In-house On-line Systems Experience

   BOSTID - some, using commercial services

   GATE -   none

   IDRC -   considerable, using in-house MINISIS system, and
            commercial services

   IFS -   none

   NUFFIC - some, using commercial services

   SAREC - none

7. In-house Terminal available

   BOSTID - yes, in NAS library. Also IBM Personal Computer which
            could be used as terminal.

   GATE -   no

   IDRC -   yes

   IFS -   not now, but proposed Phillips word-processing equipment
            could be upgraded for use as terminal

   NUFFIC - yes

   SAREC - not now, but Phillips word-processing equipment could be
            upgraded for use as terminal.