An Assessment of the Integrity of Centre Databases

Evaluation Unit
June 1995

Table of Contents

An Assessment of the Integrity of Centre Databases

1.   Introduction

This evaluation of the quality of some of the data in the Centre databases grew out of the implementation of the new electronic version of the PCR. With the introduction of the PCR database in 1994, the Evaluation Unit hired John Gordon to review the new PCR database and to begin to generate reports based on the data. As the PCR database was designed to draw on RADIUS data, part of John's work necessitated looking at the RADIUS data as well. This was in the period August-October 1994 when the Centre was in mid-stream in modifications to the new RADIUS system. At the time, he found it, not unexpectedly, to be quite incomplete, with significant gaps in the data.

Following the main download of data into the new version of RADIUS (1.4) in May, 1995, the Evaluation Unit decided that it would be important to again test the integrity of the data, in preparation for the generation of PCR reports. This was discussed with the Data Control Committee of PUG (DCC) and with MIS.

John Gordon was engaged for the week of 19-23 June 1995, during which time he looked at 1) **referential integrity;** and 2) the **consistency of data between selected fields.** In his report, summarised below, he found that in the fields important to PCR data analysis, there are still significant inconsistencies. This has implications Centre-wide as it suggests problems with other data of importance to the Centre.

2.   Main Findings of the Report

John Gordon's detailed report is attached. The report both outlines the major problems found in the data and makes initial recommendations on its clean-up. The report only deals with selected tables from the database, in particular those with relevance to PCR analysis. The report does not provide lists of specific activities which contain the inconsistent data, as the specifics should be generated during the clean-up as each field is corrected. Several illustrations of the data inconsistencies are outlined below:

- The comparison of data between the main table, PROJECT, and other tables illustrates a number of activities which are included in other tables but do not show up in the main table (p. 3).

- There are 1322 cases where a comparison of three indices of project completion are inconsistent (p. 6 and Annex D).

- There are over 500 projects listed as active which have former staff members listed as responsible PO (p. 7 and Annex G).

- Over 100 items indicated as projects have no information on total grant level (p. 7).

3.   Conclusions

Based on the findings in John Gordon's report, the following issues are put forward as the most critical for the Centre to address in ensuring data integrity:

3.1. **It is essential to verify the integrity and consistency of Centre data as we move from two data systems to one.** It will not be possible to consider the complete closure of PROMIS until the financial systems are able to link accurately and consistently with RADIUS. The longer we maintain two data systems the more problems there are likely to be.

3.2. **The planned shift to DELPHI in the Centre offers an opportunity to clean up the data, before the inconsistencies are built into the new system.**

3.3. There are two elements to a data system: a) the design of the system which is the responsibility of the MIS group; and b) the data which is entered into the system, which is the responsibility of the users. However, there are a number of important elements of the system which fall between the users and the program designers. As a result, the Centre has no one who has a clear mandate to ensure data integrity. **This mandate should be ensured. "Designer" and "user" mandates and responsibilities should be clarified and made contiguous.**

3.4. The inconsistencies are overlapping and nested, but there are some clear starting points for cleaning up the data. Many of the fields will have to be cleaned up manually, a somewhat tedious and time consuming process. **The clean-up starting points are outlined in the attached report.**

3.5. There are a number of fields in the system containing data which could be generated in the preparation of reports. **Calculated data fields should be eliminated from the database and moved to the reports themselves.** This will help reduce significantly the inconsistencies over time.

3.6. Given the number of inconsistencies, it is likely that many of the reports for which data is being collected, are not often used. The report also notes that a very large amount of data is gathered on small initiatives; this demand on those who have to enter data often results in information not being entered. **Consideration should be given to reducing the amount of data collected.**

4.   Follow-up

John Gordon's report was shared with other stakeholders in IDRC's information systems: MIS, PUG and its Data Control Committee. These three groups' responses are included in Appendices A, B, and C (pp. 21, 23 and 25 respectively).

The response from MIS clarifies some of the points raised by Mr. Gordon, and commits MIS to helping users correct inconsistent data by methods such as generating exception reports or creating programs to do batch data fixes.

PUG proposes to hire a consultant to take stock of what information is currently collected, how it is collected, and what reports are required, in order to determine how Centre data systems might be streamlined. This analysis would be timely, especially now as new hardware and software are being considered for the Centre.

The Data Control Committee's response outlines a strategy of getting users to clean "dirty data" by generating exception reports to validate the completeness of the main projet table and identify inconsistencies within it. Clean-ups will proceed systematically, coordinated by DCC. Moreover, to ensure data quality, DCC proposes appointing a Data Expert for each program unit. This person would advise personnel on RADIUS inputting and proofread selected reports for accuracy and standardization as well as for providing DCC with feedback on data problems.

John Gordon's evaluation of the integrity of a part of the Centre's main database articulates a broad-ranging concern and has already mobilized action among stakeholders across the system. These actions should ensure that data quality problems are reduced and information is better managed in the future.

Evaluation Unit 17
August 1995

CONSULTANCY TO THE EVALUATION UNIT, IDRC--JUNE 19-23, 1995

The consultant carried out the study from June 19 to June 23, 1995 at IDRC headquarters in Ottawa.

The terms of reference of the consultancy were to :

1)   meet with staff from the Evaluation Unit, MIS, and Data Control Committee to discuss issues related to the integrity of the data on the Centre's information systems as related to the Project Completion Report database system, and to discuss issues relating to the conversion of the PCR from DOS to Windows on Delphi software;

2)   look at the integrity of the data on the current PCR database system (following the conversion from PROMIS to RADIUS) and identify any problems and inconsistencies in the data, including their source;

3)   on the basis of the discussions and assessment, prepare a detailed and satisfactory report on the integrity of the data on the PCR system, discuss any problems and inconsistencies including their source, and propose means for dealing with these;

4)   submit the required report of the work accomplished to Terry Smutylo, Director, Evaluation Unit of the Corporate Affairs and Initiatives Division of the Centre by June 30th, 1995.

This report is the report referred to in section 4) of the terms of reference.

This report is based on a number of conversations and meetings with Mr. Fred Carden of the Evaluation Unit, Ms. Bohdana Dutka of the Evaluation Unit; Mr. Richard Albert of the Data Control Committee, Mr. Charles Morin Assistant Treasurer and Mr. Dominique Rivard of MIS. During theses meetings and discussions the specific items to be examined in terms of data integrity were defined. Problems of inconsistent data entry were also discussed. (In the report the term project will refer only to research activities which have been formally defined as projects--when all activities are being discussed the more generic "research activity" will be used.)

Based on these discussions, two areas of potential inconsistency in the RADIUS system were selected--referential integrity and inconsistency of data between selected fields. (Data entry errors will not be specifically searched for, but major errors found will be noted in the report or the annexes.) Non-technical explanations of "referential integrity" and "data consistency" follow:

1) Referential integrity is "computer-speak" and refers to ensuring that when a record is deleted that all of the data dependent on that record in other tables is eliminated as well. Data base developers speak of parent and child relationships--for example you could have a main (parent) table listing all of your friends and their addresses with a separate linked

(child) table on their preference for wines. If you had a fight with your friend and took them off the main (parent) table, referential integrity would ensure that all references were deleted from the linked (child) wine preferences table. To continue the comparison--referential integrity ensures that there are no "orphans". (Referential integrity would allow you to delete the information in the wine table if your friend stopped drinking, without requiring that the information in the main database was deleted.) In applications developed in "old" database software--1993 and prior-- referential integrity had to be hand programmed, a labourious task. In modern software such as Delphi or Access 2.0 the choice of implementing referential integrity is usually not much more difficult than making a menu selection and checking a box.

2) The checking of inconsistency of data between fields either in the same table or in different tables is more intuitively understandable. In the case of IDRC for example, a research activity must be legally closed before it can be administratively closed. Under the IDRC business rules no research activity should be shown as administratively closed if it is not legally closed. The consultancy examined the relationship between a number of selected fields to ensure that some of the main business rules applied to managing research activities have been applied consistently to the data entered.

The size and structure of the RADIUS database are such that it would not be possible to examine more than a modest selection from among the potential problems of referential integrity and data consistency in the time allocated for the study. Fields of particular relevance to the preparation of the PCRs were selected from the following RADIUS tables-- PROJECT, *PROJLIFE, RECIP, FNDSTAGE* and *EMPLOYEE.* The specific fields will be identified below in the context of the analysis.

Baseline data was collected on these relationships and is discussed in the body of the report or shown in annexes. At this stage, data on individual research activities has not been listed in the report, but the problems are discussed in terms of the criteria which would permit lists of non-conforming records to be easily prepared. This baseline data was collected by analysing tables on the active transaction database and slight inconsistencies in research activity totals etc., may be caused by changes in the RADIUS database during the course of the analysis.

Referential Integrity-- From the point of view of the Evaluation Unit, the main (parent) table in the RADIUS system--the one which identifies all of the research activities either completed, ongoing, or in the pipeline-- is the *PROJECT* table. If the standards of referential integrity are imposed on the RADIUS database then there should be no data on any research activity in any of the other tables, if that research activity is not included in the records in the *PROJECT* table. In this context the *PROJLIFE, FNDSTAGE* and *RECIP* tables were compared to the *PROJECT* table with the following results.

The table below shows the number of orphans in the three selected tables.

Table	Unique Project IDs	Project IDs which match *PROJECT* Table	"Orphans"
PROJECT	8.537	8.537	0
PROJLIFE	6,319	6,288	31
FNDSTAGE	7,950	7,911	39
RECIP	6,050	6,032	18

Although the total number of "orphans" appears to be 88, when duplicates were removed the total was 79. That means that in the three child tables, there were 79 research activities that were referred to that are not in the main parent *PROJECT* table and as far as RADIUS is concerned do not exist.

There is another group of research activities which remain undefined. The RADIUS system requires that in the *PROJECT* table the fields containing the Project ID, the research activity type (Project, RSA, etc.) and administrative unit code be filled in before a research activity is considered to be valid.

The Project table contains 8,537 records, distributed by research activity type as shown below:

Research Activity Type	Number of records
Projects	7,299
RSA	878
DTP	133
Undefined	227
TOTAL	8,537

As shown above there are 227 records without the research activity type defined.

In addition there are 387 records in the *PROJECT* table which do not have a responsible administrative unit defined. In all there are 398 records which are not totally defined, of which 216 records have neither the research activity type nor the administrative unit defined. A complete breakdown showing research activity type by administrative unit is given in Annex A.

While these research activities are undefined in the main parent database, some of them in fact also occur in the "child" databases as shown in the table below.

Undefined research activity linkages

Table Name	Number of records
PROJECT	397
PROJLIFE	1
FNDSTAGE	16
RECIP	22

The table shows a total of 39 references to research activities in child tables which are not properly defined in the parent *PROJECT* table. Once duplicates have been eliminated the total is 37.

While the overall numbers do not appear large, it must be remembered that this analysis has only been applied to the parent table And three of the child tables--there are about 47 tables in the RADIUS system where the project number field appears either as a key field or a foreign key field.

RECOMMENDATION-- The discrepancies between research activities identified in the parent database and those in the child databases raise important questions about the completeness of the main *PROJECT* table. Each of the "orphans" in the child databases should be examined to see if it is a "real" record and if so, a corresponding entry should be made in the *PROJECT* table. If the orphan is not a "real" entry it should be deleted from the child database. With respect to the incompletely defined research activities in the *PROJECT* table, each one will have to be verified to see if it is "real" and if it is linked to records in any of the child databases. If these records are not real they should be deleted as well as any linked records in the child database. This is particularly important as IDRC plans to implement referential integrity in the development of new RADIUS modules in Delphi.

Data integrity-- Data integrity refers to the consistency of data in related fields in the same table or in different tables. The analysis not only looks at actual data entered but at fields which are blank but which would normally be expected to contain data. In this context, given the Evaluation Unit's specific interest in projects, only data referring specifically to research activities defined as "projects" was examined. Based on discussions the following specific areas were considered:

1) The consistency of selected indicators of a project's completion status. Indicators of administrative completion, legal completion and completion of research activities by recipients were examined.

2) Staffing--The database was examined to determine if all active projects had project officers identified and if those project officers were still in fact working for IDRC. The employee database was also examined to determine if data on staff who had left the organization was handled in a consistent manner.

3) Funding-- Data on total project funding was examined to determine if data was entered for all active projects. Comparisons were also made between data in the RADIUS system and data for selected projects down loaded from the HI-FI system.

Projects-- Unless otherwise stated, all of the analysis which follows concerns research activities which have been defined as "Projects" in the *PROJECT* table and for which data has been entered on the unit which is administratively responsible. There are 7,137 projects in the *PROJECT* table meeting this definition.

<u>Project Completion Status</u>--One of the critical areas analysed was the consistency of indicators on project status--ie. whether a project was administratively or legally closed. There are a number of indicators in the system--some of them equivalent (for example the indicator of administrative status should be compatible with the project administrative completion date actual) which bear on this issue. A great deal of inconsistency was found between theses indicators. A number of fields were tested--for example, in the *PROJECT* table there are fields which indicate whether the project is administratively closed and whether the project is legally closed. If the IDRC business rules are applied, a project cannot be administratively closed, yet legally open. The summary table showing administrative and legal status in the *PROJECT* table is in Annex B. This analysis shows that according to the *PROJECT* table 78 projects are shown as being legally active, but administratively closed--a clear contradiction.

In addition, in the *PROJECT* table there are 1,215 records identified as projects where both the administrative and legal status fields are blank. This table is shown by administrative unit in Annex C. When these projects are analysed in terms of the pipeline *(FNDSTAGE* table) 134 are shown as approved and should consequently presumably be shown in the *PROJECT* table as "Active/actif "

There is a further indicator of project completion in the *PROJLIFE* table. This table has a field for the project administrative completion date actual (six of the dates in this field are larger than June 30, 1995 and some of them are in 1998). When the two fields previously analysed from the *PROJECT* table are analysed in the context of the completion date, a large number of inconsistencies become apparent. The analysis identified 1,322 projects where the three status indicators examined give conflicting results. The full table is given in Annex D, while a summary of inconsistencies is given below.

Projects with conflicting indexes of Administrative Completion

Project Admin Status Code	Project Legal Status Code	Actual Admin Comp Date Entered	Number
Active	Active	Yes	520
Active	Closed	Yes	406
Closed	Active	Yes	75
Closed	Active	No	4
Closed	Closed	No	317
TOTAL			1,322

A further comparison was carried out between the project completion status code in the *PROJECT* table and the recipient research status code in the *RECIP* table. In principle, if research was ongoing, the project status code should also read active--however in 35 cases the project administrative status code was blank and in a further 48 cases the project administrative status code indicated that the project was closed. The table showing the summary data of this analysis is in Annex E.

The conflicting data on actual legal and administrative status of projects makes planning extremely difficult. In addition, much of the analysis of other discrepancies depends on the administrative status of the project. For example, if a project is administratively closed it doesn't matter if the project officer responsible is no longer an IDRC staff member, however if the project is still administratively open, the project officer should be an existing staff member.

RECOMMENDATION--The indicators of administrative and legal status are among the key indicators for planning. It is essential that they be verified as one of the first steps in cleaning up the database. Unfortunately the number of discrepancies is large. It is possible that errors occurred in the download from PROMIS, and if this is the case correction may not be too difficult. However, if the same inconsistencies are found in PROMIS then the checking will have to be done on a project by project basis--this will be time consuming, but without these corrections RADIUS is of only limited use to the Evaluation Unit as a planning tool.

Staffing-- Most of the projects which were shown as indicators in the administrative and legal status fields had project officers assigned--for only 97 of these projects was the project officer field blank--see Annex F. However, 591 projects for which the Administrative completion status indicator was "active/actif" had former staff members listed as their project officer (many of these

were projects administered by regional offices who are apparently not yet updating RADIUS). The table below shows this data in summary form. Annex G shows the status of staff on administratively active projects by unit and by unit to which the staff now belongs.

Status of Staff on Administratively Active Projects by Unit

Unit	Staff	Former Staff	Total
ASRO/BRASI	234	223	457
CAI/IAI	139	24	163
CGT	1	0	1
COMM	3	3	6
COOP	15	5	20
EAROBRACO	171	115	286
ENRIERN	395	43	438
FAD/DB	13	3	16
HS/SSA	230	28	258
ISS/SSI	132	1	133
LACRO/BRALA	264	79	343
MERO/BREMO	27	1	28
ROSABRAFS	8	0	8
SARO/BRASU	52	7	59
SEC	9	0	9
SS/SSO	228	12	240
WARO/BRACO	186	47	233
Total	2,107	591	

During the analysis inconsistencies in the *EMPLOYEE* table were noted. Annex H shows the status of staffing in each unit as represented in the *EMPLOYEE* table. For some reason the status code for three of the staff is "odd". In addition the procedure for handling former staff is not standard. In most cases the information on unit for former staff is removed, however for 25 cases, the old unit information still exits.

RECOMMENDATION--The employee table is independent of other tables and it could be cleaned up independently.

Grant Amount-- Of the 5,908 records identified as projects in the project database 133 had no information on the total grant level. In a separate exercise information form the HI-FI financial database was compared to data in the project total grant data field. A number of discrepancies were noted--but more important, because of inconsistencies in the way that projects are numbered, it was very difficult to compare the 1995 projects. MIS provided tables which mapped

the Project number and the FIS for 1993 and 1994, but it would be very useful to have a table which mapped all project and FIS numbers.

RECOMMENDATION--As a first step in making data from different systems comparable, a map of project IDs, FIS numbers and possibly activity numbers should be created for all data in the *PROJECT* table. This should not be a major undertaking, as much of the data already exists in different tables, but it would be very convenient to have one correct table to work with.

PCRs-- There does not appear to be an indicator which indicates whether a PCR has been prepared or where it is located.

RECOMMENDATION--A table of PCRs completed should be included in the PCR system. Once the table is completed it should be fairly easy to update from the PCR system itself.

CONCLUSIONS

There are a great number of logical inconsistencies in the data in the RADIUS database as well as a significant number of cases of missing data, misspelled indicators and fields which have not had data entered into them. The analysis carried out for this report has concentrated on those fields which are of special interest to the Evaluation Unit and has examined only a small part of the total RADIUS database, but enough to indicate the likely magnitude of the problem. The cleaning up of the data in RADIUS will be a long and tedious exercise. However, it is crucial that the tables be cleaned up before the MIS unit begins rewriting the RADIUS application in Delphi and particularly before any effort is made to convert it to a client server environment. One of the conclusions that has to be considered is that much of the data is not needed--if the data were being used regularly in reporting, it is not likely that it would have been allowed to get into such a sorry state. The overall data requirements of the organization should be examined before a major effort is made to rewrite the application. This is not to say that efforts should not begin to clean up the areas which have raised issues, and which in any event will be important in the future--such issues would include information on the administrative status of the projects.

The cleaning up will also have to be carefully planned. Some of the tables like *EMPLOYEE* stand alone and correction in them will not have a major impact on other fields, however changes in the fields indicating administrative or legal status in the *PROJECT* table or the administrative completion date in the *PROJLIFE* table can produce a ripple effect if they change the overall administrative status of the project. For this reason it is important to plan the sequence in which corrections are made.

The easiest errors to correct are those which involve errors in the spelling of a status code or the use of the wrong status code-- closed/ferm_ or CLOSED for closed/ferme in the *PROJECT* table. These corrections are easily made and while they may not change the logical consistency make it easier to analyse and report on. Similarly, the updating of fields which have been left empty also helps clean up the database without impacting on the logical relationships. Both of these activities

could be undertaken immediately by summer students, and would provide an important first step in making RADIUS more useful.

The most important aspect of RADIUS to rationalize from the point of view of the Evaluation Unit is the correction of the indicators which determine whether a project is administratively or legally closed. These indicators should be modified before other areas are changed because they can have a influence on whether other indicators are considered correct or not--ie. whether or not it matters that the Project Officer is no longer on staff. There are more than 1,000 anomalies in this aspect of the data and unless it can be determined that the problems were caused by an improper download from PROMIS, the inconsistent data will have to be verified and changed manually.

It will also be important to identify "orphans"--references to projects that are not in the main *PROJECT* table--in the subsidiary tables. Once these orphans are identified it will be necessary to determine whether they are really orphans or whether data is missing from the *PROJECT* table. Incomplete records in the *PROJECT* table should also be verified to see if they are linked to other portions of the database and if so the records should be properly updated.

In the context of managing the cleaning up of the tables, it would be useful to have a set of status reports which can be run regularly and which would give management a good idea of the progress being made on key changes. Most of these reports could be written in Crystal Reports.

An important issue, which was not included explicitly as part of the study requirements, is the division of responsibility within IDRC for the data in the RADIUS database and specifically for finding and correcting errors in the existing database. Many of the existing inconsistencies are obviously a joint responsibility of the users and MIS. The users are responsible for actually entering inconsistent data, but MIS is responsible for having developed systems which allow business rules to be broken in data entry and which allow non-standard indicators to be entered into some fields. In defence of MIS, the PROMIS system from which the RADIUS database was down loaded, is in an old technology and the tools for imposing referential integrity and consistent data entry when the system was written were much less effective than the modern tools--if they existed at all. (MIS also appears to be understaffed given what is expected of it.)

Normally the users would be responsible for:

      1) Identifying which data should be included in the database.

      2) Identifying and spelling out the business rules which apply.

      3) Entering the data and monitoring it to ensure that it is correct.

      4) Identifying and in many cases designing management reports.

Normally MIS would be responsible for:
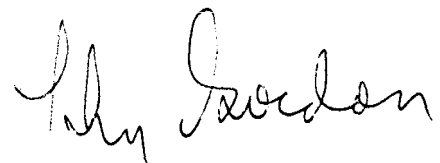
> 1) Developing the application and ensuring that the application automatically applies the identified business rules and to the extent possible protects against the entering of inaccurate data.

IDRC by setting up the Data Control Committee has made a positive step to solving some of the problems. However, IDRC still lacks a single focal point where the "buck stops" and where funds are available to finance the analysis necessary to design the database systems properly and to make necessary corrections.

For example, a brief overview of the database suggests a number of issues that should be addressed before any work is done on converting RADIUS to another software product.

> 1) There seems to be an excess of information being gathered on individual research activities given the relatively small budget of each of these activities. Requesting too much information is a serious problem which usually results in a lot of information not being entered at all because the whole process is too labourious and time consuming. The large number of errors found in the database suggests that the data is not being used by management and consequently errors are not found and not corrected--another symptom of a system which tries to do too much.

> 2) There is a lot of redundant information--for example when the administrative completion date (actual) is entered for a research activity, there is no need for a separate indicator of the activity's administrative status.

> 3) A third issue is the proliferation of "unique" activity identifiers--the HI-FI systems has an activity identification code and there are at least three different ones--project ID, activity number and FIS number in RADIUS. Developing a convention with a single code for each unique research activity is important.

In conclusion, the inconsistencies, data deficiencies and irregularities in the RADIUS system should be corrected before any work is done on converting the application to new software. The time is also appropriate to examine the data actually being entered into the system and the business rules which are being applied to see if these can be rationalized before a major investment is made on additional development.

John Gordon
Grand River Informatics Inc.
26 June 1995

Annex A

Adm_unit	RSA/ASR	Project/Projet	DTP/ODP	Null	TOTAL
AFNS/SAAN	0	427	0	0	427
ASRO/BRASI	57	1090	0	0	1147
CAI/IAI	300	273	133	8	714
CGT	0	3	0	0	3
COMM	0	31	0	2	33
COOP	0	43	0	0	43
EARO/BRAFO	42	468	0	0	510
EES/STG	0	4	0	0	4
ENR/ERN	110	969	0	1	1080
EXE	0	1	0	0	1
FA	1	5	0	0	6
FAD/DB	0	70	0	0	70
HS/SSA	125	825	0	0	950
ISS/SSI	104	571	0	0	675
LACRO/BRALA	39	737	0	0	776
LIB/BIBLIO	0	5	0	0	5
MERO/BREMO	0	33	0	0	33
MIS/SIG	2	25	0	0	27
Not a vali	0	6	0	0	6
Null	9	162	0	216	387
OPE	0	1	0	0	1
OT/BT	0	7	0	0	7
PRES	17	15	0	0	32
ROSA/BRAFS	0	10	0	0	10
SARO/BRASU	0	98	0	0	98
SEC	0	20	0	0	20
SS/SSO	64	1076	0	0	1140
WARO/BRACO	8	324	0	0	332
	======	======	======	======	======
TOTAL	878	7299	133	227	8537

Note that "Null" has been used to identify records for which there is no
data in the specified field.

Page 11--26 June 1995

Annex B

```
                                    Closed /Active Projects
95.06.22                           Project Legal Status Code
```

Project Admin. Status Code	Active /Actif	Closed /ferm,	Closed /ferm_	Null	TOTAL
active/actif	1774	989	5	0	2768
CANCEL	0	2	0	0	2
CLOSED	1	0	0	0	1
closed/ferm,	78	3058	0	0	3136
NOTREC	0	1	0	0	1
Null	14	0	0	1215	1229
TOTAL	1867	4050	5	1215	7137

Note the use of "CLOSED", "NOTREC", etc.

Annex C

```
95.06.22                  Projects without status indicators
Admin Unit                Number
```

Admin Unit	Number
ASRO/BRASI	170
CAI/IAI	83
EARO/BRAFO	48
ENR/ERN	414
FA	5
HS/SSA	181
ISS/SSI	86
LACRO/BRALA	77
LIB/BIBLIO	5
MERO/BREMO	1
MIS/SIG	25
Not a vali	6
OT/BT	7
PRES	6
ROSA/BRAFS	1
SS/SSO	78
WARO/BRACO	22
TOTAL	1215

Annex D

95.06.22   Comparison three status indicators

		Administrative Completion Date		
Pro j_adm_stat_code	RA_legl_stat_code	Yes	No	TOTAL
active/actif	active/actif	520	1254	1774
active/actif	closed/ferm,	402	587	989
active/actif	closed/ferm_	4	1	5
CANCEL	closed/ferm,	1	1	2
CLOSED	active/actif	0	1	1
closed/ferm,	active/acti	75	3	78
closed/ferm,	closed/ferm	2741	317	3058
NOTREC	closed/ferm,	0	1	1
Null	active/actif	0	14	14
Null	Null	0	1215	1215
	TOTAL	3743	3394	7137

Annex E

95.06.22   Project Administrative Completion Status   Page   1

Rec_res_stat_code	Active/actif	Null	CLOSED	Closed/ferm,	CANCEL	NOTREC
CLOSED	1	0	0	0	0	0
active/actif	1813	35	0	48	0	0
CANCEL	6	0	0	3	1	0
CLOSED	72	0	1	897	0	0
closed/ferm,	818	0	0	2216	1	0
INCOMP	36	0	0	45	0	0
NOTREC	0	0	0	0	0	1
Null	59	193	0	6	0	0

  Note the use of inconsistent indicators

Page 13--26 June 1995

.Annex F

		Prog. Off.	Indicator	
Proj_adm_stat_code	RA_legl_stat_code	Yes	No	TOTAL
active/actif	active/actif	1737	37	1774
active/actif	closed/ferme	979	10	989
active/actif	closed/ferm	5	0	5
CANCEL	closed/ferme	2	0	2
CLOSED	active/actif	1	0	1
closed/ferme	active/actif	78	0	78
closed/ferme	closed/ferme	3008	50	3058
NOTREC	closed/ferme	1	0	1
		=====	======	======
TOTAL		5811	97	5908

Annex G

```
         IDRC_Org_Unit         Staff    ExStaff      TOTAL

         Emplyee file)
-------------------


Admin Unit (Project file)--ASRO/BRASI
         HR/RH                   0        23          23
         XSTAFF                  0       200         200
         CAI/IAI                 1         0           1
         ISS/SSI                 1         0           1
         ENR/ERN                 3         0           3
         SS/SSO                  5         0           5
         PO/BP                  12         0          12
         SARO                   27         0          27
         ASRO                  185         0         185
                             =======   =====      ======
    Total ASRO/BRASI          234       223         457


Admin Unit (Project file)--CAI/IAI
         XSTAFF                  0         7           7
         PO/BP                   1         0           1
         FA                      1         0           1
         SS/SSO                  1         0           1
         WARO                    1         0           1
         ASRO                    1         0           1
                                 1         0           1
         CA-PIP                  2         0           2
         ENR/ERN                 4         0           4
         PL                      6         0           6
         EARO                    7         0           7
         LARO                    9         0           9
         CAI/IAI               105        17         122
                              =====     =====       =====
    Total                     139        24         163
    Total CAI/IAI
Admin Unit (Project file)--CGT
         FA                      1         0           1
                              =====     =====       =====
   Total CGT                    1         0           1


Admin Unit  (Project file)--COMM
         XSTAFF                  0         3           3
         CA-PIP                  3         0           3
                              =====     =====       =====
   Total COMM                   3         3           6
```

Page 15--26 June 1995

IDRC_Org_Unit (Emplyee file)	Staff	ExStaff	TOTAL
------------------	------	--------	------
Admin Unit (Project file)--COOP			
XSTAFF	0	5	5
CAI/IAI	1	0	1
FA	3	0	3
SARO	5	0	5
ENR/ERN	6	0	6
	======	=====	=====
Total COOP	15	5	20
Total			
Admin Unit (Project file)--EARO/BRAFO			
CAI/IAI	0	1	1
XSTAFF	0	114	114
ISS/SSI	1	0	1
WARO	1	0	1
PO/BP	1	0	1
ASRO	1	0	1
FA	2	0	2
HS/SSA	7	0	7
SS/SSO	10	0	10
EARO	148	0	148
	=====	=====	=====
Total EARO/BRAFO	171	115	286
Total			
Admin Unit (Project file)--ENR/ERN			
XSTAFF	0	43	43
LARO	1	0	1
ISS/SSI	1	0	1
SS/SSO	1	0	1
EARO	1	0	1
SARO	10	0	10
ASRO	11	0	11
MERO	18	0	18
PO/BP	27	0	27
FA	57	0	57
ENR/ERN	268	0	268
	======	=====	======
Total ENR/ERN	395	43	438

IDRC_Org_Unit (Emplyee file)	Staff	ExStaff	TOTAL

Admin Unit (Project file)--FAD/DB

	Staff	ExStaff	TOTAL
XSTAFF	0	3	3
SS/SSO	1	0	1
PL	2	0	2
WARO	5	0	5
FA	5	0	5
	=======	======	======
Total FAD/DB	13	3	16

Admin Unit (Project file)--HS/SSA

	Staff	ExStaff	TOTAL
XSTAFF	0	21	21
SS/SSO	1	0	1
LARO	2	0	2
ASRO	2	0	2
CAT/TAT	2	6	8
PO/BP	7	0	7
WARO	22	0	22
HS/SSA	194	1	195
	======	=====	======
Total HS/SSA	230	28	258

Admin Unit (Project file)--ISS/SSI

	Staff	ExStaff	TOTAL
HR/RH	0	1	1
MERO	3	0	3
CAI/IAI	4	0	4
WARO	4	0	4
PO/BP	16	0	16
ISS/SSI	105	0	105
	======	=====	=====
Total ISS/SSI	132	1	133

```
                    IDRCOrg_Unit     Staff ExStaff     TOTAL (Emplyee file)
                    -------------------


  Admin Unit (Project file)--LACRO/BRALA
            XSTAFF              0       78             78
            ASRO               1        0              1
            HS/SSA             2        0              2
            ISS/SSI            2        0              2
            PL                 4        0              4
            MERO               5        0              5
            SS/SSO             8        0              8
            PO/BP             10        0             10
            CAI/IAI           18        1             19
            ENR/ERN           56        0             56
            LARO             158        0            158

                             =======    ====       ======

Total LACRO/BRALA             264       79            343


  Admin Unit (Project file)--MERO/BREMO
            XSTAFF              0        1              1
            CAI/IAI            5        0              5
            MERO             22        0             22
                            ======  =======       ======
     Total MERO/BREMO        27        1             28


  Admin Unit (Project file)--ROSA/BRAFS
            EARO               1        0              1
            ROSA               7        0              7
                            =======  =======      =======
     Total ROSA/BRAFS         8        0              8


  Admin Unit (Project file)--SARO/BRASU
            XSTAFF              0        7              7
            ASRO               1        0              1
            SS/SSO             6        0              6
            SARO              45        0             45
                                                          ======
                            ======  ======
     Total SARO/BRASU        52        7             59
```

Page 18--26 June 1995

IDRC_Org_Unit (Emplyee file)	Staff	ExStaff	TOTAL
------------------	---------	------	------
Admin Unit (Project file)--SEC			
PO/BP	9	0	9
	=====	=====	=====
Total SEC	9	0	9
Admin Unit (Project file)--SS/SSO			
XSTAFF	0	12	12
MERO	1	0	1
HS/SSA	1	0	1
	2	0	2
WARO	3	0	3
LARO	3	0	3
CAI/IAI	5	0	5
ASRO	6	0	6
ENR/ERN	11	0	11
EARO	13	0	13
SS/SSO	183	0	183
	======	=====	======
	228	12	240
Total SS/SSO			
Admin Unit (Project file)--WARO/BRACO			
XSTAFF	0	47	47
ASRO	2	0	2
CAI/IAI	2	0	2
SS/SSO	2	0	2
	10	0	10
EARO	15	0	15
ENR/ERN	23	0	23
WARO	132	0	132
	=====	=====	=====
Total WARO/BRACO	186	47	233
GRAND TOTAL	2107	591	2698

Page 19--26 June 1995

Annex H

IDRC_Org_Unit	Former Staff	Staff	Odd	TOTAL
	438	15	0	453
ADMIN	0	39	0	39
ASRO	0	11	0	11
CA-PIP	0	27	0	27
CAI/IAI	2	54	0	56
COMM	1	0	0	1
EARO	0	12	0	12
EARO/BRAFO	0	0	1	1
ENR/ERN	0	42	0	42
EO/BD	1	0	0	1
FA	2	47	0	49
HGKG	0	1	0	1
HR/RH	2	21	0	23
HS/SSA	1	24	0	25
ISS/SSI	1	40	0	41
LARO	5	8	0	13
NERO	0	4	0	4
MIS/SIG	0	19	0	19
OT/BT	5	28	0	33
PL	2	4	0	6
PO/BP	0	61	1	62
ROSA	1	5	0	6
SARO	1	7	0	8
SARO/BRASU	0	0	1	1
SS/SSO	1	22	0	23
WARO	0	8	0	8
	-----	-----		
	463	499	3	965

Page 20--26 June 1995

Appendix A

MIS response to *An Assessment of the Integrity of Centre Databases*

In regards to the Evaluation Unit's summary, and in particular its conclusions:

3.2   The planned shift to DELPHI in the Centre offers an opportunity to clean up the data, before the inconsistencies are built into the new system.

The inconsistencies in the data are a legacy of the old IDRC databases. The shift to DELPHI will help ensure that the integrity issues are resolved at the source but the existing data will have to be dealt with separately.

3.3   "Designer" and "user" mandates and responsibilities should be clarified and made contiguous.

The Data Control Committee is "responsible for all issues related to data...including user compliance to data standards." The mandate is there. But the group is still in the process of defining mechanisms and tools for ensuring data quality. And, while it has the authority, it has very limited resources and no budget.

3.5   Calculated data fields should be eliminated from the database and moved to the reports themselves.

Agreed in theory; however, in practice some field calculations involve major gymnastics from the database engine and the user creating the reports. We feel that if the field is used regularly, it is worth the overhead of storing it in the database. This applies to total grant amount fields as well as status fields whose values are triggered by date fields.

3.6   Consideration should be given to reducing the amount of data collected.

The information requirements for RADIUS were defined by users in Working Groups established by PUG, which has been entrusted the responsibility by Centre management for information systems content and design. As PUG is largely comprised of staff from the operational units, this possibly explains the "excess" of information being gathered.

In regard to John Gordon's report itself and with respect to the cited subsections:

Referential integrity

The "orphans" have already been resolved by MIS as of 28 June 1995.

The activities without type or other critical data such as title, PO, or administering unit, have been identified using exception reports created by MIS.

Data Integrity

Staffing: The reason that there are three records in the employee table with the value "ODD" is to facilitate the requirements of the application SPECTRUM (formerly CENTRA) to register information on consultants. The RADIUS employee table is a direct download from the HURMIS database, which records only employees hired from HQ. It was agreed by HR that when a new human resources system is implemented, consideration would be given to include people working for the Centre with status other than "employee". In the meantime, it was agreed to add these few records directly into the RADIUS database and ensuring that these records would be accessed only by SPECTRUM.

PCRs: It is possible to identify projects for which a PCR exists by selecting records with a PCR indicator of "Y" and an actual administrative completion date. The physical location of the PCR report was not identified as required information when the original specifications were defined.

General comments

"..it is crucial that the tables of the data be cleaned up before the MIS unit begins rewriting the RADIUS applications in Delphi and particularly before any effort is made to convert it to a client server environment." (page 8)

Here, I think it is important that we realize that the data consistency is not the responsibility of MIS. It is understood that MIS must do whatever is necessary in order to help the users clean up the data, be that creating exception reports, writing programs to do batch data fixes or assisting the different user groups in identifying what needs to be done. MIS is committed to helping in this major endeavour. However I do not believe that it is necessary and strategically wise to halt development of new technologies - which will insure that in the future more data inconsistency does not occur - while the data is being cleaned up.

"The overall data requirements of the organization should be examined before a major effort is made to re-write the applications." (page 8)

Although it is very possible that some of the data gathered by the application may not be required, each data field is the result of information requirements which were defined and validated by a working group of users set up for this purpose. From our point of view in MIS, a new platform has been chosen to implement RADIUS for a number of reasons which are entirely separate from the underlying data structure.

Appendix B

PUG response to *An Assessment of the Integrity of Centre Databases:* Hiring a consultant

PUG believes that many of the problems raised in the preceding report may be due to the complexity of the system and the heavy demand on resources to maintain it. The complexity may, in turn, be due to a "missing link": While MIS is designing an information system in response to user requirements, those requirements are often based on assumptions formed by past practices, one-of-a-kind questions posed by senior management or outsiders, or pressures to track events in the life of projects down to the last detail. The missing link has been sufficient knowledge about both sides, which would enable the user *to weigh the value of the information desired against the costs of providing it,* in terms of user input, system capacity, and MIS resources.

PUG wishes to hire a consultant to provide an objective, comprehensive overview of the Centre's present information system, in order to:
- determine the essential information required for effective project/program management;
- determine the system's capacity to deliver it; and
- make recommendations in view of the resources available and the need for flexibility.

This exercise is expected to result in streamlining the collection of data and reporting, and a less complex system.

The following terms of reference have been drawn up to guide these actions:

Background

The transition from the PROMIS to the RADIUS data systems at the Centre has pointed out problems in the data systems. These problems relate both to the quality of the data in some parts of the system and to users' frustrations with the system itself. The central problem would appear to be that there is a gap between the designers and users of the system, so that the Centre has developed a highly complex information system based on the demands of the users but without sufficient knowledge on the part of the users of the potentials and limitations of the data systems.

Since the acquisition of more effective and efficient hardware and software is being considered, it appears essential to consider the transition to a more effective and efficient information system. We need to pause and take stock of what reports are needed throughout the system, what information needs to be collected, and how to produce those reports. It is anticipated that a number of reports could be discarded and some data collection could be dropped without compromising the accountability of the Centre or its ability to respond to the

demands and expectations of its constituencies, notably the Parliament and the Office of the Auditor General.

The recent and forthcoming changes in Centre structure also highlight the need for a more open and flexible information system, which is not tied so tightly to structure and within which users and designers can effectively design reports around whatever needs emerge over time.

In order for the users to make an effective contribution to the modifications, it is essential that there be a comprehensive overview of the real needs of the different components of the user community, the potentials with new data systems, and some scenarios on the cost effectiveness of various approaches.

Terms of Reference

In order to address this problem, the consultant will:

1. work with Centre staff (Management Information Services, Policy & Planning Group, Program Officers, Research Officers, Finance & Administration, Library, Internal Audit) based in Ottawa and in the Centre's regional offices (using electronic communication) to:

   a) define essential information required to manage resource allocations, monitor projects and programs supported by IDRC, and meet reporting requirements; and,

   b) assess the capacity of the current and planned information systems to provide the information defined as essential; and,

2. present alternative approaches to project/program information management in IDRC, giving consideration to cost-effectiveness (budget, learning curve, input labour, etc.) and the implications of their implementation.

Appendix C

The Data Control Committee's response to *An Assessment of the Integrity of Centre Databases*

Dirty Data

PUG's Data Control Committee (DCC) discussed at length the "dirty data" problem raised in the John Gordon report. It was agreed that there where many causes: incorrect data loaded from PROMIS; corrupted data resulting from system problems; and the evolving nature of program activities (e.g. Secretariats) which were not part of original RADIUS structure.

Members agreed that cleanup should be user-driven as owners of the data˙ and experts on the business it represents.

A small Ad Hoc cleanup committee identified a list of thirty exception reports that would be needed to clean up RADIUS data. These reports will help validate the download from PROMIS and identify data inconsistencies.

The exception reports have been divided into two groups: group one are reports which will help us validate the completeness of the main project table, and group two is a set of reports that will identify data inconsistencies.

The DCC will coordinate the cleanup by selecting the sequence in which corrections are made. Cleanup will be done systematically one exception report at a time starting with group one. Exception reports will be examined by PSU Managers and Regional Controllers to determine why the problem exists and identify who should clean the dirty data.

Cleanup is a priority and is currently under way.

Ensuring Data Quality

The DCC is now in the process of defining mechanisms and tools for ensuring data quality. A Data Expert function will be created in each program unit. Data Experts will be responsible for providing on-site consultation and advice regarding the entry of new data into RADIUS, proofreading selected reports to ensure data is entered consistently and according to standards and to provide feedback to DCC regarding data problems that require action. Data Experts will be a mix of PSU members and those using RADIUS as a source of other than administrative data. In the long run, having data experts will help prevent dirty data at the source.

Other tools (i.e. exception reports) will be required to ensure data quality. This issue will be discussed at a future meeting.