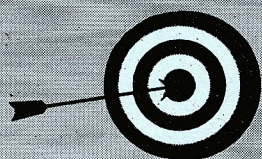




Indian and Northern Affairs Canada
Affaires indiennes et du Nord Canada



Data Quality Review of the 1995-1996 Nominal Roll

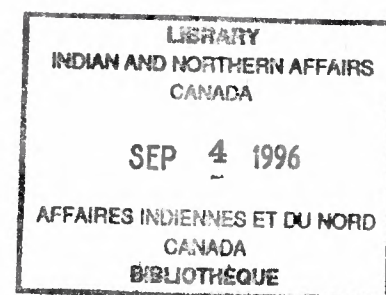
**Information Management Branch
June, 1996**

E96.2
N653
c.1

Canada

TABLE OF CONTENTS

Executive Summary	i
Part I: Introduction	1
Part II: Data Quality Findings of the Nominal Roll	3
Part III: Recommendations	6
Annex I: Methodology for Nominal Roll Data Quality Review	7



Executive Summary

Data Managers in the Department of Indian Affairs and Northern Development (DIAND) are accountable for the quality of the data they collect. To meet this obligation, a great deal of care is taken in the data collection process to ensure that quality data are entered on their databases. Notwithstanding this important work, answers are needed to two basic questions about our data holdings: (1) What is the level of quality; and (2) What level of quality is adequate? With no solid answer to these questions, data quality is a nagging issue, particularly because the department is so dependent on data.

DIAND defines Data Quality as the accuracy, completeness and currency of data. To check the accuracy of student records on the 1995/1996 Nominal Roll database, a sample of over 20,000 student records in six regions were compared to the school register and other source documents. These checks were conducted by regional staff during school visits. Completeness and currency were assessed by the Data Quality Section with the assistance of the headquarters Data Manager.

It is important to note that for the Nominal Roll accuracy, currency and completeness were measured after the numerous checking procedures have been conducted on the data, including the on-site verification of eligibility of students. In other words, data quality results on the Nominal Roll reflect the quality of the Nominal Roll data that are disseminated to stakeholders.

The 1995-1996 Nominal Roll data on students in school is estimated to be 99.4 percent accurate and this meets business requirements.

In the six regions that conducted on-site reviews, a total of 198,279 data fields were checked. Of those, 196,735 corresponded with the original source. As regions undertook to correct the errors that they found, the overall accuracy rate (which factors in the corrections) for the Nominal Roll data on students in school is estimated to be 99.4 percent. This accuracy rate meets business requirements.

**For the 1995/1996 collection year, Nominal Roll data
are 100% complete except for one region.**

Completeness reflects the number of First Nations that provided reports as a proportion of First Nations expected to report. Completeness also reflects the amount of data missing from reports. For the 1995/1996 collection year, all First Nations required to report did report. With the exception of one region, all the reports submitted were 100 percent complete.

In 1995-1996, the Nominal Roll Data were available for Business Needs on time.

Nominal Roll Data are collected yearly on September 30. In 1995/1996, preliminary data were first released to the Finance Branch for resource allocation purposes on time in January, 1996. The official release of finalized data was in April, 1996.

Finally, the Nominal Roll not only contains data regarding students in school, but also on students who left school. The quality of data on students who left school needs improvement.

Data Quality Review of the Nominal Roll

Part I: Introduction

Measuring Data Quality

The Data Quality Section developed a standard process for reviewing data quality, described in the paper Framework for Measuring and Improving Data Quality. That process, approved by all regions and sectors, can be summarized by four milestones. First, as success depends on the participation of both headquarters and regional experts, the first milestone is the assembly of a working group with membership from both headquarters and regions. The working group on the Nominal Roll met from November 1-3, 1994. The product of that meeting was a set of detailed procedures for measuring the quality of data on the Nominal Roll. Those procedures were written by the Data Quality Section (and later incorporated into an On-Site Guide) and sent to regions for approval. By May, 1995 approval-in-principle was received from all regions - the second milestone.

The third milestone is the measurement of data quality. In synchronization with the schedule for data collection, between October and December, 1995, the level of data quality on the Nominal Roll was measured on-site for every authoritative data element. Finally, measures of data quality are only meaningful when compared against standards of data quality. Minimum standards of data quality are currently being set through the department's Information Management Framework process - meeting the fourth milestone of the data quality process.

Data Quality: Definitions and Methodology

As discussed in the Framework for Measuring and Improving Data Quality, DIAND's definition of data quality¹ is the level of accuracy, completeness and currency of each data element relative to the data source (what is supposed to be counted). The main source for data on the Nominal Roll is the school register. Accordingly, accuracy on the Nominal Roll means the degree to which the data on the Nominal Roll matches the data on the school register (and other sources). Therefore, measuring the accuracy of data was done during visits to a sample of schools.

It is important to note that for the Nominal Roll, accuracy, currency and completeness are measured after the numerous checking procedures have been conducted on the data, including the verification of eligibility on-site. In other words, data quality results on the Nominal Roll reflect the quality of the Nominal Roll data that are disseminated to stakeholders.

¹ Readers are reminded that data quality should not be confused with eligibility, although regions confirmed eligibility separately during the on-site reviews.

For the Nominal Roll, completeness reflects the number of First Nations that report data as a proportion of First Nations expected to report. Completeness also reflects the amount of data missing from reports.

Currency is a measure of the age of the data. For Nominal Roll data, the main issue with currency is whether or not data are available to business managers on time.

Whereas regions assessed accuracy mainly during school visits, completeness and currency were assessed by the Data Quality Section in headquarters through an internal assessment of Nominal Roll data.

The focus of all data quality reviews is on the authoritative data elements.² A list of the authoritative data elements are detailed in the box below (see Annex I for a list of all of the data elements).

Authoritative Data Elements for the Nominal Roll Database

- | | | |
|-----------------|-----------------------------|----------------------|
| ■ Status Code | ■ Special Education | ■ School |
| ■ Grade | ■ Transportation (Regular) | ■ Leaver Reason |
| ■ Residence | ■ Transportation (Special) | ■ Leaver Destination |
| ■ Accommodation | ■ Extent of Indian Language | |

School visits for six regions were done between October and December, prior to submitting their data to headquarters. Errors found in the sample were corrected immediately on the Nominal Roll System by regions. At the same time, the errors found in the sample served as an estimate for the number of errors expected in the schools which were not part of the sample. Given this is a first-time review, one region restricted their exercise to federal and band schools. As well, two regions were not able to participate in the on-site exercise and therefore, accuracy results are not available for them. However, a full data quality review for these two regions is planned for the 1996/1997 collection cycle. Completeness and currency results are available for all regions.

² A single data element can appear in many databases but only one database is the authoritative source. For example, the Indian registration number is shown for each student on the Nominal Roll database, but the authoritative source for the Indian registration number is the Indian Register, not the Nominal Roll.

Part II: Data Quality Findings of Nominal Roll Data

Finding #1

For the 1995-1996 Nominal Roll data on students in school, the overall accuracy rate is estimated to be 99.4 percent and this meets business requirements.

In the six regions that conducted on-site reviews, a total of 198,279 data entries were checked. Of those, 196,735 corresponded to the source documentation. As regions undertook to correct the errors that they found, the overall accuracy rate (which factors in the corrections) for the Nominal Roll data on students in school is estimated to be 99.4 percent. The impacts of the 99.4 percent accuracy rate are insignificant and therefore 99.4 percent meets business requirements.

The table below summarizes in more detail the findings on data quality.

Data Elements (Authoritative)	Estimated Regional Accuracy Rates (%)						Total
	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	
Status Code	99.6%	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%
Grade	97.8%	99.0%	98.0%	98.6%	99.0%	100.0%	98.5%
Residence	98.2%	99.7%	100.0%	100.0%	99.8%	96.3%	99.8%
Accommodation	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Transport regular	96.0%	100.0%	99.8%	100.0%	100.0%	100.0%	99.7%
Transport special	93.4%	100.0%	99.9%	100.0%	100.0%	100.0%	99.5%
Special Education	99.6%	99.9%	99.9%	100.0%	100.0%	100.0%	99.9%
Extent Indian Language	85.2%	93.1%	100.0%	99.5%	100.0%	100.0%	97.9%
School	98.7%	99.9%	99.9%	99.7%	100.0%	97.0%	99.8%
Overall Accuracy	96.5%	99.1%	99.7%	99.7%	99.9%	99.3%	99.4%

It is notable that every region but one had a few errors associated with the grade data element. The accuracy rate ranges from 100.0 percent to 97.8 percent. The errors did not apply to any specific grade level, but the majority (just over 70 percent) of the grade levels were inflated. The main reason for this is probably the automatic promotion of students by the database system each year prior to the September 30 count date - apparently the automatic promotion introduces these errors. However, in most cases, these errors are corrected.

Also notable is that the "Transportation" and "Extent of Language" data elements were quite accurate, however, in two regions, the accuracy rates were lower with the majority of the errors isolated to a few schools. Indeed, in those schools, the data for most of the students were incorrect (i.e. the errors were global). The errors indicate that "transportation" and "language" coding was, at least in a couple of schools, unchallenged.

The accuracy rates are virtually the same regardless of school type (provincial/private, band/federal) - they differ by only two tenths of one percent, with the accuracy rates for band/federal schools being slightly higher than those for the provincial/private schools.

School Type	Estimated Regional Accuracy Rates (%)						Total
	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	
Provincial/Private	97.3%	---	99.6%	99.7%	99.9%	99.5%	99.4%
Band/Federal	95.6%	99.1%	99.8%	99.8%	99.8%	100.0%	99.5%

Similarly, for most regions, the accuracy rates do not differ when they are broken down by funding type. The overall accuracy rate for AFA funding types is 98.6 percent which is about one percentage point lower than the accuracy rate for CFA funding types (99.7 percent).

Funding Type	Estimated Regional Accuracy Rates (%)						Total
	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	
AFA	96.6%	98.4%	99.7%	99.9%	99.7%	---	98.6%
CFA*	96.7%	99.6%	99.7%	99.8%	99.9%	99.3%	99.7%
* The CFA accuracy rates for two regions (3&5) include some student records under Contribution, and Financial Transfer Agreements							

Finding #2

For the 1995/1996 collection year, Nominal Roll data are 100% complete except for one region.

In the 1995/1996 collection year, completeness is outstanding. All First Nations expected to report did report. With the exception of one region, all the reports submitted were 100 percent complete. Unfortunately, due to complications with a new Nominal Roll System that was implemented in the 1995/1996 year, no data was provided on "Extent of Indian Language" for one region (however, this data element was 100 percent complete in previous years).

Finding #3

In 1995-1996, the Nominal Roll Data were available to Business Managers on time.

Nominal Roll Data are collected yearly on September 30. In 1995/1996, preliminary data were first released to the Finance Branch for resource allocation purposes on time in January, 1996. The official release of finalized data was in April, 1996. The table below shows the approximate schedule of Nominal Roll data submission for the 1995/1996 collection year.

<u>Approximate 1995/1996 Nominal Roll Data Collection Schedule</u>				
Collection/Census Date ⇔	First Nations to Regions ⇔	Regions to Headquarters ⇔	Preliminary Data ⇔	Finalized Data
September 30	November 15	December 15	January 15	April 15

Finding #4

Nominal Roll data on students who left school needs improvement.

In addition to the detailed analysis of data on students in school, a less rigorous analysis of data on students who left school was conducted in headquarters and the findings suggest that the data need improvement. In particular, there appears to be inconsistent reporting between regions and between years. For example, in 1995/1996, one region showed over 1,000 students "transferred to another school", while another region showed only 21 and four other regions show zero. Between 1994/1995 and 1995/1996 the total number of "transferred to another school" records fell by more than half.

Part III: Recommendations

Recommendation #1

The Nominal Roll data on students who left school needs improvement.

- (a) The Headquarters Nominal Roll Data Manager should clarify the current leaver category definitions. The current definitions should be refined with the objective to make them clear, exhaustive, and mutually exclusive. This should be done in consultation with the appropriate business managers. As well, the Nominal Roll Data Manager should confirm that the changes to the definitions (and source documents) are consistent with the Information Management Framework.
- (b) The Headquarters Nominal Roll Data Manager should inform regions of the changes, ensure that the changes are implemented in time for the 1996/1997 collection year and ensure that regions understand the importance of collecting quality leaver data. It is suggested that the regions clarify the leaver data definitions in a letter to First Nations.

Headquarters Nominal Roll Data Manager Response:

The Headquarters Data Manager will review and clarify the data definition for the reason field in the leavers file. The revised definitions, once approved by the business areas, will be communicated to regions in August 1996 for implementation in the 1996/97 data collection process. Regions will be informed, in writing, of the revised reason field definitions in our call letter to the Data Managers in September 1996.

Recommendation #2

Improvements to the completeness of the NR data in one region are required.

- (a) The Headquarters Nominal Roll Data Manager should consult with the appropriate regional Data Manager to better understand why in some cases the data were not complete.
- (b) The Headquarters Nominal Roll Data Manager should implement preventative measures to ensure that all required NR data are received for the next collection cycle (1996/1997).

Headquarters Nominal Roll Data Manager Response:

The Headquarters Nominal Roll Data Manager will consult with the one region where completeness was not 100 percent to better understand why in some cases the data were not complete and prevent future occurrences.

Annex I

Methodology for Nominal Roll Data Quality Review

Methodology: On and Off-site Reviews

Details regarding the procedures for measuring the accuracy of Nominal Roll Data can be found in the "Regional Guide for On and Off-Site Reviews." Below are additional notes regarding methodology for the 1995/1996 data quality review exercise.

Nominal Roll Files and Data: The Nominal Roll database consists of two separate files on student records - the "master" file and the "leaver" file. The master file contains a list of DIAND funded students attending elementary-secondary school for any given year. The leaver file contains a list of DIAND funded students who have been previously funded but taken off for various reasons (e.g., moved off reserve, graduated, etc.) New students that have been deleted are not included in the leaver file.

Each student record in the master and leaver files is made up of twenty data elements. Of these, eleven are authoritative (see the box below for the list of all of the data elements for the Nominal Roll).

Data Elements for the Nominal Roll Database

- | | | |
|---|-------------------------------------|---------------------------------|
| ■ Date of Birth | ■ Dist. of Financial Responsibility | ■ Extent of Indian Language* |
| ■ Status Code* | ■ Transportation (Regular)* | ■ % of Indian Language |
| ■ Sex | ■ Transportation (Special)* | ■ Band Financial Responsibility |
| ■ Grade* | ■ Special Education* | ■ Band of Residence |
| ■ Residence* | ■ Language (On entry)** | ■ Reserve of Residence |
| ■ Accommodation* | ■ Language (of Instruction)** | ■ School* |
| ■ Leaver Reason* (Leaver File Only) | | |
| ■ Leaver Destination * (Leaver File Only) | | |

* Authoritative Data Elements

** Optional Data Elements

On-Site Review Samples:

The table below shows the percent of students and schools reviewed in each region. All but two regions completed their on-site reviews and submitted the Data Quality results with their December 15, 1995 submission.

Region	No. of Students in Sample	No. of Students in Region	Student Sample Size	No. of Schools in Sample	No. of Schools in Region	School Sample Size
Region 1	1,378	4,444	31.0%	61	166	36.7%
Region 2	3,675	20,573	17.9%	29	452	6.4%
Region 3	6,820	18,777	36.3%	51	152	33.6%
Region 4	4,455	16,593	26.8%	44	213	20.7%
Region 5	5,646	17,116	33.0%	88	315	27.9%
Region 6	57	99	57.6%	4	5	80.0%
Total	22,031	77,602	28.4%	277	1,303	21.3%

In total, approximately 28 percent of students in six regions were included in the sample. This equates to approximately 22,031 student records. (About 20 percent nationally.) Most of the regions met or exceeded their respective commitments to review 30 percent of all student records.

For two regions, the percentages of schools and students reviewed is actually higher than what is presented in the table above. For these two regions, on-site reviews were carried out at 100 percent of band schools in addition to the provincial and private schools sampled.

The school samples were essentially randomly chosen. Random sampling techniques were used to select the basis for the sample, but in some regions other sampling criteria were taken into consideration. For example, in one region, schools which were identified as "high growth" (over four percent annually) were added to the sample. On the other hand, in another region while random sampling techniques were used, not all school types were included in their sample.