



**Ministry of Advanced Education
Central Data Warehouse**

Data Quality Management Plan

Implemented October 2002

Revised May 2005

Contents

SECTION 1	PURPOSE	4
SECTION 2	PLAN OBJECTIVES & LIMITATIONS	5
SECTION 3	DATA QUALITY METRICS & THRESHOLDS	7
SECTION 4	DATA QUALITY MANAGEMENT & INTERVENTION	12
SECTION 5	DEVELOPMENT AND IMPLEMENTATION SCHEDULE	14
APPENDIX A	TABLE NAMES AND ABBREVIATIONS	15
APPENDIX B	DATA ELEMENT USES	16

Document Revision Summary

This section summarizes the changes that have been made to this document.

Revision Date	Summary
October 15, 2002	Initial draft released.
October 19, 2003	Document Revision Summary added; Contact information for CDW Coordinator updated
February 23, 2004	Contact information for CDW Coordinator updated
April 21, 2004	Four data quality metrics and thresholds added to Table 2 to support the new FTE methodology; metric descriptions revised for Birth Date, PEN and Gender in Table 2
October 8, 2004	Thresholds updated for FTE elements, new element added (FTE Divisor Unit)
May 1, 2005	Thresholds updated for FTE and Credit Attempted elements (4 elements in total)

Section 1 Purpose

The Ministry, post secondary institutions, and others require data for a variety of purposes, including research and analysis, as a basis for decision-making, and to satisfy accountability requirements. The Central Data Warehouse (CDW) was developed to serve many of those purposes, by collecting a standard set of data from each college, university college, institute and agency, and storing that data in a central database that can be accessed by the Ministry, post secondary institutions and other potential users.

The objective of the Data Quality Management Plan (DQMP), which is presented and explained in this document, is to ensure that the CDW data is of sufficient quality to serve its purposes, and that the quality of the data is actively managed for ongoing improvement over time.

Any questions about the DQMP, the CDW, or information products developed and published by the Ministry should be directed to the Ministry's Data Warehouse Coordinator, Hugh Judd (email: hugh.judd@gems3.gov.bc.ca, phone: (250) 387-8873).

Section 2 Plan Objectives & Limitations

The DQMP considers data quality to consist of the following components:

- *Accuracy* – refers to whether the CDW data is a valid and correct representation of educational activity (e.g. all data elements, such as student birth date, gender, Provincial Education Number (PEN), course section end date, student course fee type, etc., accurately represent the underlying facts).
- *Completeness* – refers to the completeness of the entire CDW data set (e.g., all required educational activity/programs are included in the CDW and all required data elements are populated).
- *Flexibility* – refers to whether the data can be easily used to satisfy a wide variety of purposes and specific questions (e.g., the data can provide users with summary level information as well as various levels of detailed information).
- *Timeliness* – refers to whether data is submitted and loaded into the CDW on a timely basis, in accordance with specified submission dates and data-loading schedules.

The Ministry's ability to determine or assess each of these components of data quality is somewhat limited, and varies depending on the particular component. For instance, the Ministry has very little ability to determine the accuracy of CDW data, and is limited to identifying only the most obvious errors (e.g., a student's birth date is identified as a date in the future). Given these limitations, the Ministry relies heavily upon the **Registrar's Sign-off Letter**, which must be submitted along with the CDW data submission, to ensure that the CDW data submission is an accurate and complete reflection of the institution's student registration and achievement activity. While Registrar sign-off procedures may vary from institution to institution, it is expected that all institutions will utilize the available data quality tools (e.g. QA Check, QA1 reports, QA2 reports, and Data Quality Metrics Reports, etc.) prior to submission of their CDW data and their Registrar Sign-off Letter.

Although flexibility is an important component of data quality, the primary factors that determine the degree of flexibility in the CDW is the database itself (i.e., the data model, the database management system, and the hardware), the data standards, and the data interrogation tool (i.e., Oracle Discoverer).

In view of the above, the balance of this DQMP document is primarily focused on the completeness and timeliness components of CDW data quality. The balance of the DQMP document is presented in two sections, as follows:

- **Section 3 Data Quality Metrics & Thresholds.** This section of the DQMP outlines the quantifiable metrics (i.e., standardized measurements) that the Ministry will use to ensure CDW data is of sufficient quality to meet the needs of the Ministry, post-secondary institutions and other potential users. This section also establishes the minimum threshold levels required for each metric.

- **Section 4 Data Quality Management & Intervention:** This section of the DQMP outlines the monitoring and intervention procedures that will be used to ensure that established data quality thresholds are met.

Section 3 Data Quality Metrics & Thresholds

This section of the DQMP outlines the quantifiable metrics (i.e., standardized measurements) that the Ministry will use to ensure CDW data is of sufficient quality to meet the needs of the Ministry, post secondary institutions and other potential users. This section also establishes the minimum threshold levels required for each metric.

The quantifiable metrics can be categorized into two groups:

- those that measure the quality of the **entire** CDW data set (Table 1); and
- those that measure the quality of **specific** data elements within the CDW (Table 2).

Table 1: Data quality metrics and thresholds for the **entire** CDW data set.

Quality Component	Metric Description	Acceptable Threshold
Accuracy	Data submissions from institutions to the CDW are correct and valid representations of educational activity (see also Table 2 below).	Registrar’s Sign-off Letter; Quality Analysis reports.
Completeness	All required educational activity/ programs included in the data submissions from institutions to the CDW, and all required data elements are populated (see also Table 2 below).	Registrar’s Sign-off Letter; Quality Analysis reports.
Flexibility	Data submissions from institutions to the CDW are in accordance with current CDW data definitions and standards	Data submission successfully loaded into the CDW
Timeliness	Data submissions from institutions to the CDW are submitted on or before the regular submission due dates: <ul style="list-style-type: none"> • November 30th, and • June 30th. 	Submission received on or before regular submission due date
Timeliness	All data submissions received from institutions are reviewed, tested for data integrity and loaded into the CDW on a timely basis.	Loaded into CDW within 14 calendar days of being received in required format.

In establishing the thresholds for **specific** data elements in the CDW, the following factors were taken into consideration:

- The perceived importance of the specific data element in meeting the needs of CDW users;

- The originating source of the specific data element (e.g., course section end date or credential achievement date is generated by the post secondary institution; student birth date is obtained from students), recognizing that institutions have more control over the quality of data elements they generate themselves than over the quality of data elements they obtain from students and other sources;
- The complexity in producing or extracting the specific data element.

For the November 2002 submission to the CDW, thresholds ranged from 80% to 90%. This reflected the notion that thresholds below 80% and 90% would limit the usefulness of the CDW data, and recognized that the ultimate threshold level of 99% for the full complement of original data elements would not realistically be achieved until the November, 2003 submission to the CDW.

Thresholds have not been established for every data element in the CDW. While institutions are required to submit data for all applicable data elements, emphasis has been placed on those elements that the Ministry considers to be a priority, as detailed in *Appendix B, Data Element Uses*. As well, elements that comprise a primary key do not require thresholds, since the entity and referential integrity rules require these elements to be 100% complete.

Table 2: Data quality metrics and thresholds for specific data elements within the CDW data set.

Table	Data Element	Metric Description	Thresholds (percent)					
			Nov 02	Jun 03	Nov 03	Jun 04	Nov 04	Jun 05
CSEC	END_DATE	Percentage of course sections with valid end dates. Invalid dates include null values and end dates before start dates. Does not apply to course sections that are related to courses with a BILLING_COURSE_INDICATOR of 'Y' or course sections established to record transfer or PLA credits.	90	95	99	99	99	99
CSEC	EXPECTED_COMPLETION_WEEKS	Percentage of course sections with valid expected completion weeks. Does not apply to course sections that are related to courses with a BILLING_COURSE_INDICATOR of 'Y' or course sections established to record transfer or PLA credits.	85	90	99	99	99	99
CSEC	COURSE_HOUR_EQUIVALENT	Percentage of all course sections with values other than null. Does not apply to course sections that are related to courses with a BILLING_COURSE_INDICATOR of 'Y' or course sections established to record transfer or PLA credits.	80	90	99	99	99	99
CSEC	CAM_CODE	Percentage of all course sections with campus code values other than null. Does not apply to course sections that are related to courses with a BILLING_COURSE_INDICATOR of 'Y' or course sections established to record transfer or PLA credits.	90	95	99	99	99	99
PRO	FUNS_CODE	The percentage of programs with values other than null. Does not apply to 'inactive' programs (i.e. programs that have no registrations since January 1, 2000).	90	95	99	99	99	99
PRO	CTYP_CODE	The percentage of programs with credential type codes other than null. Does not apply to 'inactive' programs (i.e., programs that have no registrations since January 1, 2000).	90	95	99	99	99	99
PRO	CIP_CODE	The percentage of programs with CIP codes other than null. Does not apply to 'inactive' programs (i.e. programs that have no registrations since January 1, 2000).	90	95	99	99	99	99
PRO	FTE_DIVISOR	The percentage of programs with FTE Divisor not NULL. Only applies to programs with registrations with a Registration Status Date on or after April 1, 2003. Added to support FTE counting.	n/a	n/a	n/a	85	90	95

Data Quality Management Plan

Table	Data Element	Metric Description	Thresholds (percent)					
			Nov 02	Jun 03	Nov 03	Jun 04	Nov 04	Jun 05
PRO	FTE_DIVISOR_UNIT	The percentage of programs with FTE Divisor Unit not NULL. Only applies to programs with registrations with a Registration Status Date on or after April 1, 2003. Added to support FTE counting. Note that Program Graduation Credit Unit was a proxy for this element in June 04, but has now been replaced.	n/a	n/a	n/a	85	90	95
SC	CREDENTIAL_ACHIEVEMENT_DATE	The percentage of credentials with 'valid' dates. To be 'valid', the credential achievement date must be prior to, or the same as, the date of the CDW data submission.	90	95	99	99	99	99
SC	CREDENTIAL_ISSUED_INDICATOR	The percentage of credentials with indicator values other than null.	90	95	99	99	99	99
SCR	STUDENT_COURSE_FEE_TYPE	The percentage of student course registrations with 'valid' student course fee type. To be 'valid', student course fee type must not be null.	90	95	99	99	99	99
SCR	REGISTRATION_STATUS_DATE	The percentage of course registrations with 'valid' registration status date. To be 'valid', registration status date must not be null, and must be between the course start date and the course end date (unless registration status is 'withdraw').	90	95	99	99	99	99
SCR	CREDIT_ATTEMPTED	The percentage of registrations with Credit Attempted not NULL where Registration Status Date is on or after April 1, 2003. Added to support FTE counting	n/a	n/a	n/a	85	90	95
SCR	CREDIT_ATTEMPTED_UNIT	The percentage of registrations with Credit Attempted Unit not NULL where Registration Status Date is on or after April 1, 2003. Added to support FTE counting.	n/a	n/a	n/a	85	90	95
STU	<i>Until advised further, thresholds for the following data elements in the Students Table (STU) apply only to students taking base-funded educational activity. (Note: Within the CDW, "base-funded educational activity" is defined according to the "student course fee type"; Data Definitions and Standards document for details)</i>							
STU	BIRTH_DATE	The percentage of students with valid birth dates. For the annual November CDW submission, this measure applies to all students taking courses that began between April 1, 2001 and October 15 of the submission year. For the annual June CDW submission, this threshold applies to all students taking courses that began between April 1, 2001, and March 31 of the submission year. Invalid birth dates include null values and birth dates that demonstrate an age younger than 5 yrs or older than 100 yrs as at the CDW submission date.	80	90	99	99	99	99

Data Quality Management Plan

Table	Data Element	Metric Description	Thresholds (percent)					
			Nov 02	Jun 03	Nov 03	Jun 04	Nov 04	Jun 05
STU	PEN	The percentage of students with valid (not NULL) PENS. For the annual November CDW submission, this measure applies to all students taking courses that began between April 1, 2001 and October 15 of the submission year. For the annual June CDW submission, this threshold applies to all students taking courses that began between April 1, 2001, and March 31 of the submission year.	80	90	99	99	99	99
STU	HS_CODE, HIGH_SCHOOL_GRAD_DATE, and HIGH_SCHOOL_GRAD_STATUS	No Thresholds set for institutions at this time. These elements are currently supplied by the Ministry of Education in order to improve completeness.	n/a	n/a	n/a	n/a	n/a	n/a
STU	GENDER	The percentage of students with a known gender (i.e., male or female, not null or "unknown"). For the annual November CDW submission, this measure applies to all students taking courses that began between April 1, 2001 and October 15 of the submission year. For the annual June CDW submission, this threshold applies to all students taking courses that began between April 1, 2001, and March 31 of the submission year.	80	90	99	99	99	99

* New elements and revised metrics have been shaded. Thresholds for revised elements will be 99% for the November 2005 submission.

Section 4 Data Quality Management & Intervention

The DQMP is overseen by the Ministry's CDW Coordinator. The CDW Coordinator is responsible for monitoring compliance with data quality thresholds and for identifying when intervention action is required.

The nature and extent of intervention action to be taken will depend on the nature and extent to which data quality thresholds are not being achieved. The levels of intervention action are outlined below.

Late Submissions

Level T1: CDW Data submission and Registrar sign-off letter not received by regular submission due date.

- Ministry communicates directly and immediately with the Registrar, advising him/her that the CDW data submission and Registrar sign-off letter have not been received, and requesting that they be received within 7 calendar days following regular submission due date.

Level T2: CDW Data submission and Registrar sign-off letter not received within 7 calendar days following regular submission due date.

- Ministry communicates directly with the President of the post secondary institution, advising him/her that the CDW data submission and Registrar sign-off letter have not been received, and requesting that that they be received within 14 calendar days following regular submission due date.

Level T3: CDW Data submission and Registrar sign-off letter not received within 14 calendar days following regular submission due date.

- In the event that the CDW data submission and Registrar sign-off letter are not received within 14 calendar days following the regular submission due date, the Ministry may initiate a review of the institution's data management systems and procedures, with a view to assisting the institution to revise those systems and procedures and submit the overdue CDW data submission and Registrar's sign-off letter as soon as possible. The Ministry may also consider other alternative measures for ensuring that the CDW data submission and Registrar sign-off letter are received.

Completeness Thresholds Not Met

Level C1: Minimum completeness threshold level for at least one data element is not met, but all data elements are within 5% of minimum completeness threshold level.

- Ministry accepts the institution's CDW data submission and includes the data in CDW. The submitted data will be reflected in CDW standard reports and will be used for analysis and decision-making.
- Ministry communicates directly with the Registrar, advising him/her of the specific data elements that did not meet minimum completeness threshold level, and requests that the institution improve data quality to meet or exceed minimum completeness threshold levels for all data elements by the following CDW submission.

Level C2: Minimum completeness threshold level for at least one data element is not met, but at least one data element is not within 5% of minimum completeness threshold level.

- Ministry communicates directly with the President of the post secondary institution, advising him/her of the data quality problems and requesting that the institution make a revised CDW data submission (with revised Registrar sign-off letter) within 14 calendar days of the regular submission due date. (The Ministry may agree to extend the period for revised submission beyond 14 calendar days, depending on circumstances, following discussion with the Registrar.) In the revised CDW data submission, all data elements are required to meet or exceed minimum completeness threshold levels. The revised CDW submission must be within 5% of minimum completeness threshold level for all data elements.
- Until the revised CDW data submission and revised Registrar sign-off letter are received, the Ministry will determine the extent to which the existing data submission will be included in CDW standard reports and used for analysis and decision-making.
- In the event that the revised CDW data submission and revised Registrar sign-off letter are not received within the prescribed time period, the Ministry may initiate a review of the institution's data management systems and procedures, with a view to assisting the institution to revise those systems and procedures and submit the revised CDW data submission and revised Registrar's sign-off letter as soon as possible. The Ministry may also consider other alternative measures for ensuring that the revised CDW data submission and revised Registrar sign-off letter are received as soon as possible.

Section 5 Development and Implementation Schedule

The development and implementation schedule for the DQMP is as follows:

2002:

- January – April 2002 – A draft DQMP is presented to the Student Standards Committee for review and comment.
- May – June 2002 - A revised draft DQMP is distributed to all post secondary institutions for review and comment.
- July – August 2002 - June 30 CDW data submissions are tested against tentative thresholds for information only; feedback is provided to, and received from, post secondary institutions.
- October 2002 - DQMP is finalized and distributed to all post secondary institutions.
- November 2002 - Full implementation of the DQMP for the November 2002, CDW submission.

2004:

- June 2004 – Four data quality metrics and thresholds added to support the new FTE methodology
- November 2004 and June 2005 – Thresholds increased for new data quality metrics added in June 2004

Annual Activity for 2003 and Beyond

- January – March - Annual revisions made to DQMP based on emerging priorities, feedback and suggestions from post secondary institutions, and internal/external environmental changes.
- June, November - Ongoing implementation of the DQMP for the regularly scheduled CDW submissions on June 30 and November 30 each year.

Data Quality Management Plan

Appendix A Table Names and Abbreviations

Following is a list of DDEF version 2004 table names and abbreviations used in this document.

<u>Table Name</u>	<u>Table Name Abbreviation</u>
COURSE SECTION	CSEC
PROGRAM	PRO
STUDENT	STU
STUDENT COURSE ACHIEVEMENT	SCA
STUDENT COURSE REGISTRATION	SCR
STUDENT CREDENTIAL	SC

Appendix B Data Element Uses

The following table identifies some of the anticipated purposes for which the data in the CDW may be used, along with the specific data elements required for each purpose. For the specific completeness thresholds established for each data element (excluding those elements which are primary keys in their respective tables), see Section 3, Table 2.

Use Description	Elements Needed	
	Table	Element
To determine educational activity over a period of time (e.g., headcount over a specified time period)	CSEC CSEC CSEC CSEC SCR SCR	START_DATE END_DATE EXPECTED_COMPLETION_WEEKS COURSE_HOUR_EQUIVALENT CONTINUOUS_ENROLMENT_START_DATE REGISTRATION_STATUS_DATE
To determine the registration status at a particular time (e.g., headcount reporting)	SCR	REGISTRATION_STATUS_DATE
To identify base funded activity (e.g., system headcount)	SCR	STUDENT_COURSE_FEE_TYPE
To identify the credit value a student attempts to earn through a student course registration. Added to DQMP to support FTE counting.	SCR	CREDIT_ATTEMPTED
To identify the unit describing the CREDIT_ATTEMPTED for each student course registration. Added to support FTE counting	SCR	CREDIT_ATTEMPTED_UNIT
To categorize program funding (e.g., group program activity by funding code)	PRO	FUNS_CODE
To categorize discipline of programs (e.g., group program activity by discipline area)	PRO	CIP_CODE
To identify the divisor used to calculate enrolments. Added to support FTE counting	PRO	FTE_DIVISOR
To identify the unit describing the FTE Divisor. Added to support FTE counting;	PRO	FTE_DIVISOR_UNIT
To eliminate duplicate students (e.g., unduplicated headcount of students in the system)	STU	PEN
To determine number and type of credentials issued	SC SC PRO	CREDENTIAL_ACHIEVEMENT_DATE CREDENTIAL_ISSUED_INDICATOR CTYP_CODE
To analyze educational activity based on student characteristics (e.g., gender or age participation in particular program areas)	STU STU STU STU STU	BIRTHDATE HS_CODE HIGH_SCHOOL_GRAD_DATE HIGH_SCHOOL_GRAD_STATUS GENDER
To identify the campus location of educational delivery	CSEC CSEC	CAM_INS_CODE CAM_CODE

Use Description	Elements Needed	
	Table	Element
To analyze course achievements	SCA SCA	ACHIEVEMENT_STATUS PASS_FAIL_INDICATOR

Note that additional elements may be added as standard reports are defined and methodology developed.