



Ministry of Advanced Education Central Data Warehouse

Quality Assurance and the Data Quality Management Plan

May 2016

Table of Contents

HISTORY	2
PURPOSE	3
PLAN OBJECTIVES & LIMITATIONS	3
SECTION 1: QUALITY ASSURANCE (QA1_AND_QA2)	4
SECTION 2: DATA QUALITY METRICS & THRESHOLDS	6
SECTION 3: DATA QUALITY MANAGEMENT & INTERVENTION	8
APPENDIX A: TABLE NAMES AND ABBREVIATIONS	10
APPENDIX B: DATA ELEMENT USES	11

History

The Data Quality Management Plan (DQMP) was introduced in October 2002 and has undergone a number of revisions over the years to add new elements for completeness of threshold testing. This revision, as of February 2013, introduces new elements for testing, removes some elements from testing that are routinely 100% complete, and expands the plan to include other quality assessment processes that have always been in place but not formerly acknowledged as part of the DQMP. As well, this revision removes some of the historic information about completeness thresholds that existed in 2002 but have long been out of use. Some of the wording in section 3 has been changed to improve clarity and better reflect practice. Previous versions of the DQMP are available upon request.

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 those purposes by collecting a standard set of data from public postsecondary institutions, and storing that data in a central database that can be accessed by the Ministry, post secondary institutions and other users under particular approved circumstances.

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 on products developed and published by the Ministry should be directed to the Ministry's CDW Coordinator, AEIT.CDWContact@gov.bc.ca

Plan Objectives & Limitations

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

- *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.
- *Completeness* – refers to the completeness of the entire CDW data set
- *Accuracy* – refers to whether the CDW data is a valid and correct representation of educational activity.
- *Flexibility* – refers to whether the data can be easily used to satisfy a wide variety of purposes and answer specific questions.

The Ministry's ability to assess each component of data quality varies depending on the particular component. 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 the Data Quality Metric Reports) 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 (e.g. the data model, the database management system, and the hardware), the data standards, and the data interrogation tools (e.g. Oracle Discoverer)

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

Section 1: Quality Assurance (QA1_and_QA2)

QA1 report tables count values for all elements in the DDEF database and QA2 report tables test the logical relationship for or between various table elements. The QA1_and_QA2 script was created for institutions to test (count) the table element values prior to CDW submission, to alert institutions to potential data quality issues, and to provide an opportunity to cleanse data errors.

The Ministry requires the QA1 and QA2 reports to be exported as an Oracle export file so that the Ministry can run automated checks against the reports and data. The QA2 reports should also be exported to Excel to facilitate internal institutional data quality review and to assist in the Registrar’s sign-off process. Below is the list of reports that are created from the script and the DDEF2000 table that is being analyzed.

Table 1: Quality Assurance (QA1 Reports)

QA1 Reports	CDW Table
QA1_CAM	Campuses
QA1_CRS	Courses
QA1_CSDM	Course Section Delivery Mode
QA1_CSEC	Course Section
QA1_CSFS	Course Section Funding Source
QA1_PRO	Programs
QA1_SC	Student Credentials
QA1_SCA	Student Course Achievements
QA1_SCR	Student Course Registrations
QA1_SES	Sessions
QA1_SSR	Student Session Registrations
QA1_STU	Students
QA1_YOUNG_REGISTRANTS	Students less than 14 years old

Table 2: Quality Assurance (QA2 Reports)

QA2 Reports	CDW Table	Logical Tests: Not all data returned in the QA2 reports are considered ‘errors’. QA2 is meant to provide the institutions an opportunity to consider data values that may not seem right in relation to other data values.
QA2 COURSES	Courses	<ul style="list-style-type: none"> • MIN Credit Val>Max Credit Val: The minimum credit value is greater than the maximum credit value. • CRS EFF Date >= CRS End Date: The course effective date is greater than, equal to, the course end date.
QA2 CSEC	Course Sections	<ul style="list-style-type: none"> • CRS EFF Date > CSEC Start Date: Course effective date is greater than the course section start date. • CSEC Start Date>CSEC End Date: Course section start date is greater than the course section end date. • Section Lab but Lab Hours NULL: Course section is ‘Lab’ (02) but course lab hours is NULL.
QA2 CSFS SEATS	Course Section Funding Source Seats	<ul style="list-style-type: none"> • Sum_Seats_From_CSFS: Lists course sections where the number of seats does not equal the number of seats in the course section funding source table. Indicates “SUM SEATS OK” if seats are equal. This relationship is critical to FTE counting where sections have at least 1 funding source code.
QA2 PRO	Programs	<ul style="list-style-type: none"> • PRO EFF Date >= PRO End Date: Program effective date is greater than, equal to program end date. • Grad Credits NULL/Cred Unit NN: Program graduation credit is NULL when the program graduation credit unit is NOT NULL.

QA2 Reports	CDW Table	Logical Tests: Not all data returned in the QA2 reports are considered 'errors'. QA2 is meant to provide the institutions an opportunity to consider data values that may not seem right in relation to other data values.
		<ul style="list-style-type: none"> • Grad Cred Unit NULL/Credits NN: Program graduation credit unit is NULL when the program graduation credit is NOT NULL.
QA2 SC	Student Credentials	<ul style="list-style-type: none"> • PRO EFF Date>Cred Achiev Date: Program effective date is greater than the credential achievement date. • SC CTYP > PRO CTYP: Student credential type is greater than (hierarchy rank) that the program credential type.
QA2 SCA	Student Course Achievements	<ul style="list-style-type: none"> • AchStat PCWI Pass Fail NULL: Achievement status is PLA, Complete, Withdraw, Incomplete but the Pass/Fail indicator is NULL • Grade NN But Pass Fail NULL: There is a grade but the Pass/Fail indicator is NULL • Ach Stat Date<CSEC Start Date: Achievement status date is less than the course section start date. • Grade But No Grade Type: A course grade was provided with no grade type. • PLA Type NULL but PLA: Prior learning assessment type is NULL when the achievement status is 'prior learning assessment'. • AchStat Null but Grade Type: Achievement Status is NULL but there is no Grade type. • Grade Type is NN but No Grade: Grade type is Not NULL while there is no grade.
QA2 SCR	Student Course Registrations	<ul style="list-style-type: none"> • StuEndDate>CSECEnd: Student end date is greater than the course section end date. • RegStatDateNOTinCourseDates: Registration status date is not between the continuous enrolment start date or course section start date and the student end date. • STUinOneSec>1: Students in one course section more than once.
QA2 SES	Sessions	<ul style="list-style-type: none"> • Start Date >= End Date: Session start date is greater than the session end date. • Start Date > Withdraw Date: Session start date is greater than the session withdraw date.
QA2 SSR	Student Session Registration	<ul style="list-style-type: none"> • CITZ Code CA Immi Stat Non-Can: Citizenship code is Canadian but the immigration status is non-Canadian. • Imm Stat Canadian Citz Not CA: Immigration status is Canadian but the citizenship is not Canadian. • CITZ Code CA SCFT 003: Citizenship code is Canadian but the student course fee type is 003 (International). • IMM_STAT 00 SCFT 003: Immigration status is Canadian citizen but the student course fee type is 003 (international).
QA2 STU	Students	<ul style="list-style-type: none"> • GradDate NN but GradStat<>G E: High school grad date is provided but the high school grad status is 'Did not graduate from high school' or 'Unknown' • BirthDate<APR 1902: Birth date is earlier than April, 1902.

Section 2: 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 3); and
- Those that measure the quality of **specific** data elements within the CDW (Table 4).

Table 3: Data quality metrics and thresholds for the **entire** CDW data set.

Quality Component	Metric Description	Acceptable Component Source / Threshold
Accuracy	Data submissions from institutions to the CDW are correct and valid representations of educational activity (see also Table 4 below).	Registrar's Sign-off Letter and Quality Analysis reports detailing data accuracy.
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 4 below)	Registrar's Sign-off Letter and Quality Analysis reports detailing data accuracy.
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	Final data submissions from institutions to the CDW are submitted on or before the regular submission due dates (Oct 31 st and May 31 st).	Submission received on or before regular submission due date.

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 the 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.

Thresholds have not been established for every data element in the CDW. While institutions are expected to submit data for all applicable data elements, emphasis has been placed on those elements that the Ministry considers to be a priority. 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 4: Data quality metrics and thresholds for specific data elements within the CDW data set.

Table	Data Element	Metric Description	Threshold %
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 course with a BILLING_COURSE_INDICATOR of 'Y' or course sections established to record transfer or PLA credits.	99
CSEC	EXPECTED COMPLETION WEEKS	Percentage of course sections with an expected completion weeks value 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.	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.	99
CSEC	CAMPUS 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.	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.	99
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.	99
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 date submission.	99
SC	CREDENTIAL ISSUED INDICATOR	The percentage of credentials with indicator values other than NULL.	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.	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')	99
SCR	CREDIT ATTEMPTED	The percentage of registrations with Credit Attempted not NULL where Registration Status Date is on or after April 1, 2003.	99
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.	99
STU*	BIRTH DATE	The percentage of students with valid birth dates. For the annual October CDW submission, this measure applies to all students taking courses that began between April 1, 2004 and September 15 of the submission year. For the annual May CDW submission, this threshold applies to all students taking courses that began between April 1, 2004 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.	99
STU*	PEN (Personal Education Number)	The percentage of students with valid (not NULL, containing no letters and not less than or greater than 9 digits) PENs. For the annual October CDW submission, this measure applies to all students taking courses that began between April 1, 2004 and September 15 of the submission year. For the annual May CDW submission, this threshold applies to all students taking courses that began between April 1, 2004 and March 31 of the submission year.	99
STU*	GENDER	The percentage of students with a known gender (i.e. male or female, not NULL or "unknown"). For the annual Oct CDW submission, this measure applies to all students taking courses that began between Apr 1, 2004 and Sept 15 of the submission year. For the annual May CDW submission, this threshold applies to all students taking courses that began between April 1, 2004 and March 31 of the submission year.	99
SSR	CITIZENSHIP CODE	The percentage of Student Session Registrations with valid citizenship codes. For the annual October CDW submission, this measure applies to all student taking courses that began between April 1, 2012 and September 15 of the submission year. For the annual May CDW submission, this threshold applies to all students taking courses that began between April 1, 2012 and March 31 of the submission year.	99
SSR	IMMIGRATION STATUS	The percentage of Student Session Registrations with valid immigration statuses. For the annual October CDW submission, this measure applies to all student taking courses that began between April 1, 2012 and September 15 of the submission year. For the annual May CDW submission, this threshold applies to all students taking courses that began between April 1, 2012 and March 31 of the submission year.	99
CRS	DISCIPLINE CODE	The percentage of Courses with valid discipline codes. For the annual October CDW submission, this measure applies to all students taking courses that began between April 1, 2009 and September 15 of the submission year. For the annual May CDW submission, this threshold applies to all students taking courses that began between April 1, 2009 and March 31 of the submission year.	99
<p>*Thresholds for elements in the Students Table 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" (001, 101, 002, 005, 006, 007, 015, 020)). See Data Definitions and Standards document for details.</p>			

Section 3: Data Quality Management & Intervention

The DQMP is overseen by the Ministry's CDW Coordinator. The 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 (*Submission date is determined when the first successful data submission is loaded with all constraints enabled*)

Level 1: A successfully loaded CDW Data Submission and / or 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 or Registrar sign-off letter have not been received, and requesting that they be received immediately.

Level 2: CDW Data submission and or 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 or Registrar sign-off letter have not been received and requesting that they be received immediately. A supplementary message will accompany the Ministry's sign-off letter detailing the institution's late submission.

Level 3: CDW Data submission and Registrar sign-off letter not received within 14 calendar days following 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 agree to extend the period beyond 14 calendar days depending on circumstances following discussions with the Registrar.

Completeness Thresholds Not Met

Level 1: Minimum completeness threshold level for at least one data element is not met and the Registrar's office provides a data quality explanation. The Ministry accepts the institution's CDW data submission and includes the data in the CDW. The submitted data will be reflected in CDW standard reports and will be used for analysis and decision making.

- In the form of the Ministry's acceptance letter, the Ministry communicates directly with both the Registrar and President, advising him/her of the specific data elements that did not meet the minimum completeness threshold level, and requesting that the institution improve data quality to meet or exceed minimum completeness threshold levels for all data elements by the following CDW submission.

Level 2: Minimum completeness threshold level for at least one data element is not met and the Registrar's office does not provide an explanation, or the Registrar's office does provide an explanation and it is not accepted by the Ministry.

- The Ministry communicates immediately and directly with the Registrar, 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). **The recorded submission**

date will be reset to the new revised and successfully loaded CDW data submission date.

- 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 made available for analysis and decision-making.

The Ministry reserves the right to exclude an institution from CDW standard reports and analysis if a submission is late to the point that system reporting deadlines are jeopardized or if the data are not of sufficient quality.

Appendix A: Table Names and Abbreviations

Following is a list of DDEF version 2000 table names and abbreviations used in this document.

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

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 2, Table 4.

Use Description	Elements Needed	
	Table	Element
To determine education activity over a period of time. (e.g. AY, FY, Nov 1 st by Headcount, FTE).	CSEC CSEC CSEC CSEC SCR SCR SCR	START_DATE END_DATE EXPECTED_COMPLETION_WEEKS COURSE_HOUR_EQUIVALENT CONTINUOUS_ENROLMENT_START_DATE REGISTRATION_STATUS_DATE STUDENT_END_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 Domestic/International).	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) and accountability reporting.	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 identify duplicate students (e.g. unduplicated headcount of students in the system) and to establish data linkages to student mobility	STU	PEN
To determine number and type of credentials.	SC SC SC	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
To analyze course achievements	SCA SCA	ACHIEVEMENT_STATUS PASS_FAIL_INDICATOR
To analyze educational activity based on student characteristics (e.g. Canada – BC Immigration Agreement).	SSR SSR	CITZ_CODE IMMIGRATION_STATUS
To categorize discipline of courses (e.g. group course activity by discipline area).	CRS	DIS_CODE

Additional elements may be added as standard reports are defined and methodology developed.