

Standard for the file naming convention of batch processing data files submitted to or extracted from non-legacy Department of Higher Education and Training Management Information Systems

Contents

1	Purpose of the standard.....	2
2	Scope and applicability of the standard.....	2
3	Technical specification of the file naming convention.....	2
3.1	Batch processing data file formats.....	2
3.2	File naming convention.....	2
3.2.1	Source.....	2
3.2.2	Recipient.....	3
3.2.3	File.....	3
3.2.4	Version.....	3
3.2.5	Date.....	3
3.2.6	File extension .dat.....	3
4	Unique identifiers.....	3
4.1	ETCBIS Identifier.....	3
4.2	ETQE Identifier.....	4
4.3	Employer identifier.....	4
4.4	Training Provider identifier.....	4
4.5	File identifier.....	5
	Annexure A – Illustrative Example.....	6

Acronyms

DHET	Department of Higher Education and Training
DHET MIS	Department of Higher Education and Training Management Information System developed after 2012
ETCBIS	Education and training related computer-based information systems
ETQE	Education and Training Quality Entity
HETMIS	Higher Education and Training Management Information System
NCAP	National Career Advice Portal
NLRD	National Learners' Records Database
SAQA	South African Qualifications Authority
SDL No	Skills Development Levy reference number

Definitions

- “batch processing data files” - input data collected into batches of files and processed in batches
- “computer-based information system” – a system which uses computer technology to perform some or all of its tasks.
- “Education and Training Quality Entity” -
- “HETIS Officer” - an official charged with certain responsibilities regarding post-school education and training information in terms of the Higher Education and Training Information policy

1 Purpose of the standard

The purpose of this standard is to provide a framework for the naming convention for batch processing data files received by Department of Higher Education and Training Management Information Systems (DHET MISs) developed after 2012 that form part of the overall implementation of the Higher Education and Training Management Information System (HETMIS). It is however envisaged that the standard could be implemented by other education and training stakeholders in the implementation of data exchanges between disparate data sources within the education and training sector of South Africa.

The standard is intended to provide a common set of rules to apply to the naming of batch processing data files. Naming files in a logical and predictable manner will distinguish similar files from each other at a glance and by doing so will facilitate the storage and retrieval of such files.

2 Scope and applicability of the standard

This standard applies to all DHET MIS that form part of the overall implementation of HETMIS.

3 Technical specification of the file naming convention

3.1 Batch processing data file formats

The file format for all batch processing data files is fixed length record ASCII files. In other words each record in a file takes up the same amount of space as all of the other records in the same file, regardless of how many characters are in each field. Each record must be terminated by a Microsoft compatible end of line character (CR+LF).

3.2 File naming convention

The file naming convention for batch processing data file submitted to or extracted from DHET MISs must have the following naming pattern:

[Source]_[Recipient]_[File]_[Version]_[Date].dat*
Example: CHED_0001_101_v002_20130714.dat
Indicates a file extracted from the information system at the Council on Higher Education (CHED) submitted to the National Learners' Records Database (0001) which contains provider data records (101) in the format defined on version 002 of the Specifications for Load Files for the National Learners' Records Database (v002) as extracted on 14 July 2013.

A description of the acceptable values that may be provided for each part of the file naming convention follows:

3.2.1 Source

This value indicates the source of the data file. The source identifier may be an ETCBIS, ETQE identifier, Employer or Training Provider identifier (refer to Section 4.1, 4.2, 4.3 or 4.4 for further

* The names of the type of value required have been delimited with square brackets [] for clarity in order to denote that the string is replaced by a value

information in regard to what these identifiers represent, how they are constructed and the custodians of these values).

3.2.2 Recipient

This value indicates the recipient of the data file. The recipient identifier may be an ETCBIS or ETQE identifier (refer to Section 4.1 and 4.2 for further information in regard to what these identifiers represent, how they are constructed and the custodians of these values).

3.2.3 File

This value indicates the type of data contained within the data file. A file identifier may only be a valid file identifier as registered by the HETIS officer of the DHET (refer to Section 4.5 for more information in this regard).

3.2.4 Version

This value indicates the version of the file that is being submitted. Each data load specification, as published by the recipient of the batch processing data file, has a version identifier which has a length of exactly 4, starts with the letter v (lower case) followed by three digits.

Convention: vnnn

Examples: v001 indicates version one (1) of a data load file specification

3.2.5 Date

This value indicates the date on which the data file was extracted from the source system. Any data files that belong to the same data submission must share the same date value. This value has a length of exactly 8 digit characters and must represent a valid date.

Convention: YYYYMMDD

Examples: 20130217 indicates that the submission was extracted on the 17th of February 2013.

3.2.6 File extension .dat

As indicated in section 3.1 all batch processing data file will have a fixed length record ASCII format. As a result the file must have a .dat extension to denote a data file with a specific format.

4 Unique identifiers

4.1 ETCBIS Identifier

This category of systems includes education and training related computer-based information systems (ETCBIS) that cannot be identified by means of any of the other identifiers listed below. Each ETCBIS has a unique identifier that is comprised of exactly 4 digits.

A register of these identifiers will be maintained by the HETIS Officer of the DHET and can be downloaded from <http://www.dhet.gov.za/HETMIS/>. The implementation of new systems on this register will be done in consultation with other relevant stakeholders.

Examples of ETCBISs include the National Learners' Records Database (NLRD) and the Higher Education Quality Committee Information System (HEQCIS).

Convention:	nnnn	
Examples:	0001	indicates the NLRD
	0002	indicates the HEQCIS

4.2 ETQE Identifier

An Education and Training Quality Entity (ETQE) has a unique mnemonic identifier that is comprised of exactly 4 letter characters. ETQE identifiers are issued by the South African Qualifications Authority (SAQA). A listing of these unique identifiers is documented in the "Specifications for Load Files for the National Learners' Records Database" which can be downloaded from the "NLRD Load Specifications" link at <http://www.saqa.org.za/nlrinfo.asp> (refer to the Mnemonic values in Appendix B: Unique Identifiers for Data Suppliers).

Convention:	cccc	
Examples:	CHED	indicates the Council on Higher Education and Training
	SERV	indicates the Services SETA

4.3 Employer identifier

Skills Development Levy registered employers are uniquely identified by the Skills Development Levy reference number (SDL No) as issued by SARS. SDL reference numbers have a length of exactly 10 characters, start with the letter 'L', followed by nine digits.

Convention:	Lnnnnnnnnn	
Examples:	L123456789	indicates a SDL registered employer

Employers that are not registered for the Skills Development Levy must be recorded in the recipient information system as a non-levy paying employer and a unique identifier must be issued to the employer for the purposes of submitting data to the recipient system. Non levy paying employer unique identifiers must have a length of exactly 14 characters, start with the letter 'N', followed by nine digits (unique to the recipient system) and the ETQE or ETCBIS identifier.

Convention:	Nnnnnnnnncccc	
Examples:	N123456789SERV	indicates a non SDL register employer recorded on the Services SETA information system

4.4 Training Provider identifier

Training providers are issued a unique identifier by the primary ETQE of the training provider. This unique identifier is referred to as the Provider Code. ETQE systems however are not integrated and as a result the possibility exists that two ETQE's may issue the same provider code. A provider code is as a result unique only when considered in combination with the identifier of the primary ETQE of the provider.

A provider code may have a length of no greater than 20 characters and comprises of either letter and/or digit characters. The ETQE mnemonic identifier has a length of exactly 4 and is

comprised of letter characters. The unique combination of provider code and ETQE mnemonic identifier is delimited with an underscore.

A listing of the unique identifiers for training providers is documented in the “Provider List on the NLRD” which can be downloaded from the “Provider List on the NLRD” link at <http://www.saga.org.za/nlrinfo.asp>.

Convention: [Provider Code]_[ETQE Identifier]
Examples: H14_CHED indicates the University of South Africa

4.5 File identifier

File identifiers have a length of exactly 3 digit characters which represents the content of the data file. The content of data files fall into one of 6 different file categories that describe the overall focus of the data. The file categories with their numbering bands are as follows:

Miscellaneous	001 – 099
Provider	100 – 199
Employer	200 – 299
Learning Unit	300 – 399
Person and person designation	400 – 499
Enrolment/Achievement	500 – 599

The format of a file may be specific within the context of the recipient of the data file. As an example file 501, which focuses on qualification enrolments, when submitted to the Further Education and Training Management Information System (FETMIS) has a slightly different format to file 501 as submitted to the National Learners’ Records Database (NLRD). However file 400, which focuses on person details, when submitted to FETMIS has exactly the same format as when submitted to the NLRD. It should be noted however that files that have a completely different format and/or focus will not share the same file identifier across recipients. In all instances it is important that the load file specifications of the recipient system are thoroughly reviewed when extracting data.

A register of file identifiers will be maintained by the HETIS Officer of the DHET and can be downloaded from <http://www.dhet.gov.za/HETMIS/>. The implementation of new files on this register will be done in consultation with other relevant stakeholders.

Annexure A – Illustrative Example

