NHS

National Treatment Agency
for Substance Misuse

National Drug Treatment Monitoring System (NDTMS)

NDTMS Data Set

CSV File Format Specification

Author:	J. Jaswani
Approver:	M. Roxburgh
Approval Date	16/11/09
Version	7.01

Document Information and revision history

File Name	NDTMS - CSV Input File F#2B.doc		
Original Author(s)	**Graham Scott**		

Version	Date	Author(s)	Revision Notes
3.0	29/05/2007	G Scott	Updated for Core Data Set E
5.3.0	02/07/2008	G Scott	Update external references
6.0.0	11/11/2008	G Scott	Updated Extract Criteria
6.0.1	29/04/2009	A Juna	Updated NDTMS Current Year Extract Criteria
7.01	11/11/2009	M Roxburgh	Changed extract period
7.01	12/02/2010	J Jaswani	Added statement around TOP Records being expected in the submission file. Expanded on example in Section 3.5 around client data order.

Sign Off Authorised By:			
Name	**Responsibility**	**Signature**	**Date**
Malcolm Roxburgh	**Information Manger**		**12/02/2010**

CONTENT

1   Introduction

This document defines the format of the Comma Separated Variable (CSV) file used as the primary means of inputting the Core Data Set data items into the National Drug Treatment Monitoring System (NDTMS).

2   Treatment Provider Data Export

The NDTMS systems (DAMS) do not require a particular methodology or application to be used to create the CSV input file. As a *technical* minimum the file should contain all new records and all changes to existing records since the previous submission, but doing so is not recommended. The recommended minimum submission period for activity data is the current reporting period (se section 3.4) As National Statistics and performance figures are published monthly from the data, it is assumed that a CSV file will be submitted monthly.

Input files can contain any number of client record updates and could represent data from any or all of the following:

All client data from a particular Treatment Provider, or a collection of Treatment Providers.

A single client record from a particular Treatment Provider.

Selected client records from a particular Treatment Provider.

A mixture of any of the above.

3   Input File

3.1 File Structure

The CSV Input file will be structured according to common CSV file format standards. Each line, including the last, will be terminated by a Carriage Return (CR)

3.2 File Naming

The CSV Input file filename must contain the following items of information:

Originating Agency code (where the file contains data from more than one Treatment Provider, the code of the primary Treatment Provider should be used).

Dates relating to the time span of the extracted client data

File structure suffix: .CSV

Filenames will take one of the following forms:

AAAAA-YYYYMMDD-yyyymmdd-TREAT-IN-X.CSV

AAAAAA-YYYYMMDD-yyyymmdd-TREAT-IN-X.CSV

The first naming standard is intended for use by third-parties that are providing data, the second naming standard is for use of the NDTMS Core Date Set - Data Collection spreadsheet.

AAAAA	Originating Treatment Provider code (padded with any necessary leading zeros) – The regional NDTMS will be responsible for specifying the Treatment Provider code, typically this will be a 5 character code (only in exceptional cases will it be a 6 character codes).
YYYYMMDD & yyyymmdd	Dates defining the time span that the file relates to: YYYYMMDD = From date, yyyymmdd = to date
TREAT-IN	Literal to denote the type of data contained in the file
X	The version of the Core Data set that was used to produce the extract –see NDTMS Core Data Set - Technical Definition.
CSV	File Structure indicator

The following are example filenames:

L0105-20031101-20031130-TREAT-IN-A.CSV

This file would be a CSV input file from Treatment Provider 'L0105' containing all client data relavent to November 2003, working to version "A" of the core data set.

3.3 File Content

A CSV Input file will consist of multiple rows/ lines of data.

The first row will be the header record

Subsequent rows will contain a row for each instance of a treatment modality or Treatment Outcomes Profile (TOP) record for a client.

The minimum file size would be a single row (i.e. a header row, with zero client rows).

The header row is used to define which of the Core Data Set fields are represented within the CSV file, and the order in which they are provided. The header row must be constructed as a list of all the field name of the items contains in the file followed by a comma.

The field names are specified in the NDTMS Core Data Set - Technical Definition. Given that this is a CSV file, the CSV file header values must be used.

The data rows must be in the field order given in the header and a comma must follow each field. A comma must follow null or empty fields (all rows in the file must contain all fields seperated by a comma – even if the field is null).

An example of a CSV Input file is given overleaf

Please note that this is an example only, for full details of the required fields and header names, please refer to the NDTMS Core Data Set - Technical Definition.

FINITIAL,SINITIAL,DOB,SEX,ETHNIC,PC,………..,MODAL,MODST,MODEND
M,M,1984-10-04,M,,CM14 4,………. 4,2004-01-18,04,2004-01-19
D,D,1957-03-12,M,B,NW7 3,………..,,,

The above example also demonstrated the handling of null values

If a field contains a comma, then the field must be wrapped with a single set of quotation marks.

COMMERCIAL IN CONFIDENCE

3.4 Extraction Criteria

The data to be extracted should conform to one of the following standards and should contain all data relating to a treatment episode including all associated treatment modality and TOP records.

Full Extract: All relevant activity data on the treatment provider database, regardless of the discharge date. Where practical, this should be the default option. The advantage of doing so will be to ensure that the NDTMS and the treatment provder database are fully synchronised, and that performance figures produced from the data are an accurate representation of activity recodrded at the provider.

NDTMS Reporting Extract: This extract covers the period over which changes to activity data may have an impact on published performance figures. The period required for the reporting extract is the 1st January from the previous financial year, with the financial year changing after the July upload. E.g.

For submission of data on 01/10/2009 the extract should contain all records relating to treatment episodes that are not discharged, or who were discharged after 01/01/08.

For submission of data on 01/07/2010 the extract should contain all relating to treatment episodes that are not discharged, or who were discharged after 01/01/08.

For submission of data on 01/08/2010 the extract should contain all records relating to treatment episodes that are not discharged, or who were discharged after 01/01/09.

This level of extract is the recommended minimum, as changes to data from previous financial years that may affect current performance figures will be updated by the submission,

NDTMS Current Year Extract: All records relating to treatment episodes that have a discharge date of on or after 1st January of the start of the financial year or that are currently open. For Healthcare Commission reporting requirements this extract criterion will have to be extended to July of the following year. For example:

 For submission of data on August 2009 – July 2010 – start from date would be 1st January 2009.

 For submission of data on August 2010 – July 2011 – start from date would be 1st January 2010.

Note: This level of extract is technically acceptable, but is not recommended as it cannot be used to ammend submitted data from previous financial yhears data which affects current performance figures.

Note: An open episode means an episode without a Discharge Date

The Extract <from date> is the date that has has been used to determine from when the data has be extracted (e.g 20040401 or 20050101 etc) and the Extract <to date> is the date when the extract was run.

E.g.

A **Full Extract** file will have the <from date> as the earliest discharge date in the file

An **NDTMS Reporting Extract** file will have the <from date> as 20050401

An **NDTMS Current Year Extract** file will have the <from date> as

20080101 – for submissions of data relating to April 2008 – March 2009

20090101 – for submissions of data relating to April 2009 – March 2010

Software suppliers should ensure that their system can support at least all of the above extract criteria.

Treatment Providers should take guidance from their Regioanal NDTMS database manager as to what extract criteria should be used for their monthly submission.

Records relating to Tier2 treatment modalities in YP services and carers, relatives and concerned others should be excluded from the extract – for further clarification please refer to the NDTMS Core Data Set – Business Definitions for Young People's Treatment providers.

3.5 Client Data Order

Due to the representation of client data in the CSV Input file, it is likely that there will be multiple rows of data per client. There must be one row per Treatment Provider episode referral / treatment modality and one row per episode referral /TOP record.

Consider the example where client A has had 2 treatment modalities and two TOP records which are part of one treatment episode and four treatment modalities as part of another treatment episode. The CSV Input file will contain 8 rows of data for that client:

Row N	Client A details	Treatment Episode 1 Details	Modality details	1
Row N+1	Client A details	Treatment Episode 1 Details	Modality details	2
Row N+2	Client A details	Treatment Episode 1 Details	TOP Details	1
Row N+3	Client A details	Treatment Episode 1 Details	TOP Details	1
Row N+4	Client A details	Treatment Episode 2 Details	Modality details	1
Row N+5	Client A details	Treatment Episode 2 Details	Modality details	2
Row N+6	Client A details	Treatment Episode 2 Details	Modality details	3
Row N+7	Client A details	Treatment Episode 2 Details	Modality details	4

Note: TOP data and treatment modality data CANNOT be in the same row within the CSV Input File.

COMMERCIAL IN CONFIDENCE

The NDTMS does not expect client records to be sorted in any particular order. It may be helpful however for the data to be sorted as described below to aid the manual review / checking of data.

By client (ascending Initials, Date of Birth, Gender)

By Referral Date (in chronological order)

By treatment Modality Start date (in chronological order, with blank treatments last)

By TOP date (in chronological order)

This might give a sort order within the file as follows:

FINITAL	SINITIAL	DOB	SEX		REFLD		MODST
G	S	1969-01-15	F		2004-04-12		2004-04-16
G	S	1969-01-15	F		2004-04-12		2004-05-10
J	A	1963-10-19	M		2004-04-04		2004-04-04
J	A	1963-10-19	M		2004-04-10		2004-04-26
M	M	1987-10-04	M		2003-03-31		2004-05-01

Appendix A – Abbreviations

CDS   **Core Data Set**

CSV   **Comma Separated Variable**

NDTMS   **National Drug Treatment Monitoring System**