



***National Treatment Agency  
for Substance Misuse***

**National Drug Treatment Monitoring System  
(NDTMS)**

NDTMS Data Set

CSV File Format

Author:	G. Scott
Approver:	M.Roxburgh
Approval Date	11 Nov 2008
Version	6.0.0

**Document Information and revision history**

<b>File Name</b>		NDTMS - CSV Input File F#2B.doc	
<b>Original Author(s)</b>		Graham Scott	
<b>Version</b>	<b>Date</b>	<b>Author(s)</b>	<b>Revision Notes</b>
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

<b>Sign Off Authorised By:</b>			
<b>Name</b>	<b>Responsibility</b>	<b>Signature</b>	<b>Date</b>
Malcolm Roxburgh	Information Manger		9 Nov 2005

CONTENT

1 INTRODUCTION.....4

2 AGENCY DATA EXPORT.....4

3 INPUT FILE.....4

3.1 FILE STRUCTURE.....4

<b>Supporting documentation:</b>	<b>Ver. No</b>
NDTMS Core Data Set – Technical Definition	6.0.0

<b>Distribution List</b>			
<b>Name</b>	<b>Department</b>	<b>Telephone No.</b>	<b>Email Address</b>

3.2 FILE NAMING .....5

3.3 FILE CONTENT .....6

3.4 EXTRACTION CRITERIA .....7

3.5 CLIENT DATA ORDER.....8

## 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 does not prescribe or require a particular methodology or application to be used when creating the CSV input file content. What is expected is that the contents of the fields that are used to update the NDTMS database will contain all changes within an Treatment Provider since the creation of the last input file. It is currently assumed that this process would be undertaken on a monthly basis.

- 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-yyyyymmdd-TREAT-IN-X.CSV

AAAAAA-YYYYMMDD-yyyyymmdd-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 characters code (only in exceptional cases will it be a 6 character codes).
YYYYMMDD & yyyyymmdd	Dates defining the time span that the file relates to:  YYYYMMDD = From date, yyyyymmdd = 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 <a href="#">NDTMS Core Data Set - Technical Definition</a> .
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 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.

### 3.4 Extraction Criteria

The data to be extracted should conform to one of the following standards

- **Full Extract:** All data regardless of the discharge date.
- **NDTMS Reporting Extract:** All open episodes and those episodes that have a discharge date of on or after 1<sup>st</sup> April 2005.

This extract is in line with the NDTMS Reporting Minimum Data level<sup>1</sup>

- **NDTMS Current Year Extract:** All open episodes and those episode that have a discharge date of on or after 1<sup>st</sup> January of the start of the financial year. E.g.

For submission of data relating to April 2008 – March 2009 – date would be 1<sup>st</sup> January 2008

For submission of data relating to April 2009 – March 2010 – date would be 1<sup>st</sup> January 2009

**Note:** This extract should be suitable for most normal monthly submissions, however where data quality exercises have been undertaken data which lies outside this period, an alternative extract **MUST** be used to ensure that all the changes have been passed to NDTMS

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

The Extract <from date> is the date that has been used to determine from when the data has been 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 the above extract criteria.

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

Records relating to Tier2 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.

---

<sup>1</sup> The NDTMS Reporting Minimum Data Level is the minimum amount of data that should be held by a system to support the NDTMS reporting requirements and to ensure that data quality exercises can be undertaken with a Treatment Provider using the data held in their system.

### 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 referral / treatment modality

Consider the example where client A has had 2 modalities as part of one referral episode and four modalities as part of another referral episode. The CSV Input file will contains 6 rows of data for that client:

Row N	Client A details	Referral 1 details	Modality 1 details
Row N+1	Client A details	Referral 1 details	Modality 2 details
Row N+2	Client A details	Referral 2 details	Modality 1 details
Row N+3	Client A details	Referral 2 details	Modality 2 details
Row N+4	Client A details	Referral 2 details	Modality 3 details
Row N+5	Client A details	Referral 2 details	Modality 4 details

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)

**This might give a sort order within the file as follows;**

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



## Appendix A – Abbreviations

**CDS**            **Core Data Set**

**CSV**            **Comma Separated Variable**

**NDTMS**        **National Drug Treatment Monitoring System**