

# Data Quality Maturity Index

Published 9 May 2017



**Information and technology**  
**for better health and care**

# Contents

<b>Introduction</b>	<b>4</b>
Data Quality Maturity Index	4
Publication History	5
Associated Tools	6
<b>Key Drivers</b>	<b>7</b>
Health and Social Care Act 2012	7
False or Misleading Information	7
<b>Methodology</b>	<b>8</b>
Core Field Descriptions	8
Data Collection Fields	9
Calculations of the DQMI	11
Contents of spreadsheet	14
Suppression Rules	16
Provider Codes	17
<b>Changes since last publication</b>	<b>17</b>
Changes to the DQMI Calculation	18
Changes to the Data	18
Changes to presentation	18
No changes to presentation have been made in DQMI-5	18
Changes to the Field Validations	18
Changes to Power BI Interactive Report	18
<b>Proposed Changes for Future Publications</b>	<b>19</b>
Future Developments	19
<b>History of changes from Previous Publications</b>	<b>19</b>
<b>SUS Data Quality Dashboards</b>	<b>20</b>
<b>IG Toolkit</b>	<b>20</b>
<b>Featured Topics of Interest</b>	<b>21</b>
<b>Appendix 1: Example of DQMI Calculation</b>	<b>22</b>
<b>Appendix 2: Caveats</b>	<b>23</b>
<b>Appendix 3: Supplementary Information</b>	<b>26</b>
Links to Information and Reports for Individual Datasets	26

Scope of commissioning dataset submissions (CDS) and AE) and Mental Health CDS Activity	26
Scope of Community and Mental Health Submissions (CYPHS, IAPT, MSDS and MHMDS)	27

## Introduction

The [Health and Social Care Act 2012](#)<sup>1</sup> (section 266) states that our statutory data quality role is to assess the extent to which the data it collects meets defined national standards and to publish the results of the assessments. The Data Quality Maturity Index (DQMI) is a quarterly publication intended to raise the profile and significance of data quality in the NHS.

## Data Quality Maturity Index

The Data Quality Maturity Index (DQMI) provides healthcare data submitters with timely and transparent information about their data quality.

The current DQMI (DQMI-5) publication is a refined version of the previous publications and will be further improved based on stakeholders feedback and includes additional datasets and data quality dimensions. DQMI is currently based on the completeness, validity, default values and Coverage<sup>2</sup> of the core data items agreed by the National Information Board (NIB) working group. Fields or data items include NHS number, date of birth, gender, postcode, speciality and consultant. For a full list please refer to the table on the section [Data Collection Fields](#).

DQMI-5 reports from the following datasets for the quarter, October-December 2016:

- Admitted Patient Care
- Outpatient
- Accident and Emergency
- Mental Health Services
- Improving Access to Psychological Therapies
- Diagnostic Imaging
- Maternity Services
- Children and Young People's Health Services

---

<sup>1</sup> <http://www.legislation.gov.uk/ukpga/2012/7/contents>

<sup>2</sup> Coverage is the degree to which data have been received from all expected data suppliers (see [Calculation of Coverage](#)).

## Publication History

DQMI Short code	Description	Publication Date	Reporting Period	DQMI Inclusion
DQMI-1	First publication	24 May 2016	Calendar year, January to December 2015	Completeness and validity
DQMI-2	Second publication	15 August 2016	Quarter of January to March 2016	Completeness and validity
DQMI-3	Third Publication	8 November 2016	Quarter of April to June 2016	Completeness, validity and default values
DQMI-4	Fourth Publication	7 February 2017	Quarter of July to September 2016	Completeness, validity, default values and coverage
<b>DQMI-5</b>	<b>Current Publication</b>	<b>9 May 2017</b>	<b>Quarter of October to December 2016</b>	<b>Completeness, validity, default values and coverage</b>
DQMI-6	Next Publication	August 2017	Quarter of January to March 2017	Completeness, validity, default values, coverage and consistency

## Associated Tools

All these tools can be accessed from the [NHS Digital website data quality page](#)<sup>3</sup>.

### Power BI

An interactive reporting tool, produced in Power BI, provides an interactive distribution of scores, with the option to create groupings by commissioning region and locality. This tool can also be used to create peer comparisons.

There have been some significant developments to the user functionality of this interactive report. See [Changes to Power BI Interactive Report](#) section for details.

### Performance Evidence Delivery Framework

NHS Digital recommends the use of a supportive Performance Evidence Framework designed to help data providers to improve their level of data quality by enhancing their own local processes. This framework has been used successfully in an Acute Trust environment, leading to an improved understanding of the importance of data quality alongside an improvement in the quality of data itself.

### CSV File

A CSV file of the raw data on which the DQMI is based is available in machine readable format.

It consists of the following fields:

Field	Description
PERIOD TO	First month included in the reported period.
PERIOD FROM	Last month included in the reported period.
DATASET	The dataset to which the row of data relates.
DATA PROVIDER CODE	The organisation provider code to which the row of data relates. Codes starting with an '8' are displayed at five character level and all other codes are displayed at three character level.
DATA PROVIDER NAME	The full organisation name
DATA ITEM	The specific data item to which the row of data relates.
TOTAL RECORDS	The total number of records submitted.
VALID, COMPLETE RECORDS	The number of records submitted that contained a value that was valid. NHS Data Dictionary default codes are included in this count.
DEFAULT RECORDS	The number of records submitted that contained an NHS Data Dictionary default code.
DEFAULT THRESHOLD	Threshold set for acceptable default codes
DEFAULTS IN EXCESS	The number of default records that are not considered meaningful and are excluded from the data item score, therefore, from the dataset score and DQMI score.
COVERAGE DATASET	Coverage is the degree to which data have been received from all expected data suppliers.
DQMI	Data quality maturity index for that provider
DATASET SCORE	Dataset Score for that provider and dataset
DATA ITEM SCORE	Data Item Score for applicable field for that provider and dataset

<sup>3</sup> <http://content.digital.nhs.uk/dq>

## Key Drivers

### Health and Social Care Act 2012

Section 266 of the [Health and Social Care Act 2012](#)<sup>4</sup> states that NHS Digital's statutory data quality role is to assess the extent to which the data it collects meets applicable published standards and to publish the results of the assessments. In addition, [Section 265](#)<sup>5</sup> of the Act states that NHS Digital may give advice or guidance on data quality relating to the collection, analysis, publication or other dissemination of data and information.

### False or Misleading Information

The [Francis Inquiry](#)<sup>6</sup> found that false or misleading information, such as inaccurate statements about mortality rates, allows poor and dangerous care to continue.

In response, the [Care Act 2014](#)<sup>7</sup> put in place a new criminal offence applicable to care providers who supply, publish or otherwise make available certain types of information that is false or misleading, where that information is required to comply with a statutory or other legal obligation. The offence also applies to the 'controlling minds' of the organisation, where they have consented or connived in an offence committed by a care provider.

This [guidance](#)<sup>8</sup> sets out the context for the offence and explains how the offence is applied.

---

<sup>4</sup> <http://www.legislation.gov.uk/ukpga/2012/7/section/266>

<sup>5</sup> <http://www.legislation.gov.uk/ukpga/2012/7/section/265>

<sup>6</sup> <https://www.gov.uk/government/publications/report-of-the-mid-staffordshire-nhs-foundation-trust-public-inquiry>

<sup>7</sup> <http://www.legislation.gov.uk/ukpga/2014/23/contents/enacted>

<sup>8</sup> [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/403084/FOMI\\_Guidance.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/403084/FOMI_Guidance.pdf)

# Methodology

## Core Field Descriptions

The core fields included in the DQMI along with the description of how the validity and default measures have been calculated is available in the 'Core Field Descriptions' tab of the DQMI publication accessible via the [NHS Digital website data quality page](#)<sup>9</sup>.

The information provided includes the following:

- 'Data Item' – the name of the core field as defined in the [NHS Data Dictionary](#)<sup>10</sup>. This is also a hyperlink and opens onto this field definition in the NHS Data Dictionary.
- 'Plain English Description' – a description of the core field in layman's terms.
- 'Definition of Validity' – a pseudo-code description of the validation rules applied to the core fields in the DQMI.
- 'Threshold for proportion of defaults' – Percentage of defaults accepted as valid and *included* in the Percentage Valid, Complete percentage scores:
  - 100% - all default values are meaningful and considered valid
  - 0% - all defaults are meaningless and considered invalid
  - '-' - default values are not applicable for this data item
  - Any other percentage value - a mix of meaningful and meaningless default values. This gives the percentage of default values considered valid. (See [Calculation of variable thresholds](#))

Further information on the validity definitions used in the DQMI can be obtained by contacting the relevant dataset teams through [enquiries@nhsdigital.nhs.uk](mailto:enquiries@nhsdigital.nhs.uk).

---

<sup>9</sup> <http://content.digital.nhs.uk/dq>

<sup>10</sup> [http://www.datadictionary.nhs.uk/web\\_site\\_content/pages/help\\_pages/data\\_collections\\_help.asp?shownav=1](http://www.datadictionary.nhs.uk/web_site_content/pages/help_pages/data_collections_help.asp?shownav=1)



## Data Collection Fields

Only core fields (or data items) for each of the seven datasets are measured within the DQMI. The core fields are outlined in the table below.

Data Item	Accident and Emergency	Admitted Patient Care	Children and Young People's Health Services (CYPHS)	Diagnostic Imaging	Improving Access to Psychological Therapies	Mental Health Services	Maternity Services <sup>1</sup>	Outpatient
	(AE)	(APC)		(DID)	(IAPT)	(MHSDS)	(MSDS)	(OP)
ACTIVITY TREATMENT FUNCTION CODE	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
ADMINISTRATIVE CATEGORY CODE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
ADMINISTRATIVE CATEGORY CODE (ON ADMISSION)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ADMISSION METHOD (HOSPITAL PROVIDER SPELL)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
CARE PROFESSIONAL MAIN SPECIALTY CODE	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
CONSULTANT CODE	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
DECIDED TO ADMIT DATE	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
DISCHARGE DATE (HOSPITAL PROVIDER SPELL)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
DISCHARGE DESTINATION CODE (HOSPITAL PROVIDER SPELL)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
DISCHARGE METHOD CODE (HOSPITAL PROVIDER SPELL)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ESTIMATED DATE OF DELIVERY METHOD (AGREED)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
ETHNIC CATEGORY	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
ETHNIC CATEGORY (MOTHER)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
GENERAL MEDICAL PRACTICE CODE (PATIENT REGISTRATION (MOTHER))	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Data Item	(AE)	(APC)	(CYPHS)	(DID)	(IAPT)	(MHSDS)	(MSDS)	(OP)
GENERAL MEDICAL PRACTICE CODE (PATIENT REGISTRATION)	✓	✓	✓	✓	✓	✓	☐	✓
NHS NUMBER	✓	✓	✓	✓	✓	✓	☐	✓
NHS NUMBER (MOTHER)	☐	☐	☐	☐	☐	☐	✓	☐
NHS NUMBER STATUS INDICATOR CODE	✓	✓	☐	✓	☐	☐	☐	✓
NHS NUMBER STATUS INDICATOR CODE (MOTHER)	☐	☐	☐	☐	☐	☐	✓	☐
ORGANISATION CODE (CODE OF COMMISSIONER)	✓	✓	☐	☐	✓	✓	☐	✓
ORGANISATION CODE (CODE OF PROVIDER)	☐	✓	☐	☐	☐	☐	☐	☐
PATIENT CLASSIFICATION CODE	☐	✓	☐	☐	☐	☐	☐	☐
PERSON BIRTH DATE	✓	✓	✓	✓	☐	☐	☐	✓
PERSON BIRTH DATE (MOTHER)	☐	☐	☐	☐	☐	☐	✓	☐
PERSON GENDER CODE CURRENT	✓	✓	✓ <sup>2</sup>	✓	✓	✓ <sup>2</sup>	☐	✓
POSTCODE OF USUAL ADDRESS	✓	✓	✓	✓	✓	✓	☐	✓
POSTCODE OF USUAL ADDRESS (MOTHER)	☐	☐	☐	☐	☐	☐	✓	☐
PRIMARY DIAGNOSIS (ICD)	☐	✓	☐	☐	☐	☐	☐	☐
SITE CODE (OF TREATMENT)	☐	✓	☐	☐	☐	☐	☐	✓
SOURCE OF ADMISSION CODE (HOSPITAL PROVIDER SPELL)	☐	✓	☐	☐	☐	☐	☐	☐
SOURCE OF REFERRAL FOR A&E	✓	☐	☐	☐	☐	☐	☐	☐
SOURCE OF REFERRAL FOR COMMUNITY	☐	☐	✓	☐	☐	☐	☐	☐
SOURCE OF REFERRAL FOR MENTAL HEALTH	☐	☐	☐	☐	✓	☐	☐	☐
SOURCE OF REFERRAL FOR OUTPATIENTS	☐	☐	☐	☐	☐	☐	☐	✓

**Notes:**

<sup>1</sup> In the case of Maternity these fields are completed in respect of the mother.

<sup>2</sup> In the case of Children and Young People's Health Services (CYPHS) & Mental Health Services Dataset (MHSDS) this field is PERSON STATED GENDER CODE

## Calculations of the DQMI

### Calculation of Coverage

The quarterly<sup>11</sup> coverage is calculated as:

$$C = \frac{\text{No. of periods for which a provider submitted data}}{\text{No. of periods a provider was expected to submit data}}$$

Examples:

For a quarter that includes October, November and December. “✓” Indicates data submitted, “✗” indicates data expected but not submitted and “-” indicates where data is not expected.

Oct	Nov	Dec	Coverage Factor	C	Plain English
✗	✗	✗	0	0.00	Data expected for all months, but no submissions made
✗	✗	✓	1/3	0.33	Data expected for all months, but only 1 month submitted
✓	✓	✗	2/3	0.66	Data expected for all months, but only 2 month submitted
✓	✓	✓	3/3	1.00	Data expected for all months, and all data submitted
✓	-	-	1/1	1.00	Data expected for 1 month, and all data submitted
-	✗	✓	1/2	0.50	Data expected for 2 months, but only 1 month submitted

If a Provider has closed, then it is not expected to submit data after the month of closure. If a Provider is new then it is expected to submit data after the month opened, despite having no previous submission history.

### Coverage Expectancy

#### APC, OP and AE

Expected coverage is checked against [Monthly Hospital Activity Unify Return](#)<sup>12</sup> reports published by NHS England which give activity by provider for these datasets.

The reports used are:

- ‘Attendances and Emergency Admissions’ for AE
- ‘Monthly Hospital Activity’ for APC
- ‘Quarterly Hospital Activity’ for OP activity.

<sup>11</sup> A three-month period that acts as a basis for the reporting. The four quarters are: Jan-Mar; Apr-Jun; Jul-Sep; and Oct-Dec.

<sup>12</sup> <https://www.england.nhs.uk/statistics/statistical-work-areas/hospital-activity/monthly-hospital-activity/>

A provider is expected to submit when the provider has Unify activity (see above) or has submitted CDS data to SUS for the relevant dataset (APC, OP, AE) in the relevant month.

### CYPHS, DID, MHMDS and MSDS

Expected coverage uses a rolling six-month window, as there is not a master list of MHMDS providers. For example, a provider is expected to submit in December 2016 if the provider has submitted at least one month during the period June 2016 to November 2016.

### IAPT

Expected coverage looks back over previous 2 months (3 periods including current), as there is not a master list of IAPT providers. For example, a provider is expected to submit in December 2016 if the provider has submitted at least one month during the period October 2016 to December 2016

### Methodology for Excluding Default Values from Percentage Valid, Complete

Default values for each data item are identified and defined on NHS Data Dictionary. Loshin (2011)<sup>13</sup> defines that there are two kinds of default values: meaningless and meaningful. A meaningless default is equivalent to a true “null” value that represents the absence of a value. A meaningful default value is the one used to represent some concept without specifying a value. For the purposes of the DQMI, we have made a distinction between these as:

**Meaningless:** When the actual value is not available or is defined as ‘Not Known’. These values are eliminated from the valid attribute during the calculation of the DQMI score. *For example, Ethnic Category ‘99’ (Not Known), as there is a separate (valid) code of ‘Not Stated’ for these cases.*

**Meaningful:** When the actual value offers some valuable information. *For example, using a default to specify non-consultant led activity such as nurses or midwives on the Consultant Code field.*

### Calculation of defaults in excess

Defaults in Excess include all meaningless default values and also a proportion of records with a mix of meaningful and meaningless defaults. These values are eliminated from the valid attribute during the calculation of the DQMI score. *For example, Administrative Category Code is a mixed default as ‘98’ (Not Applicable) is considered meaningful, and ‘99’ (Not Known) is considered meaningless. This field is given a threshold of 4% for Outpatients and 1% for Admitted Patient Care; see below for the calculation of thresholds.*

Defaults in excess =  $Total\ Defaults - (Threshold * Number\ of\ Records)$

If less than zero then count as 0

The validity of the field will be affected depending on the types of default a data item has:

<sup>13</sup> Reference: Loshin, D. *The Practitioner’s Guide to Data Quality Improvement*. 2011. USA. Morgan Kaufman.

Default type defined for the data item	Threshold	Notes
Only meaningful defaults	100%	All records with default values will be treated as valid
Only meaningless defaults	0%	All records with default values will be treated as invalid ( <i>defaults in excess</i> )
A mix of meaningful and meaningless defaults	Variable, see details below this table	A proportion of records with default values will be treated as invalid ( <i>defaults in excess</i> )
No defaults defined	-	No records with default values will be present on this data item

### Calculation of variable thresholds

1. Obtain the proportion of defaults per data item per dataset for each provider from DQMI-4 data.

$$\text{proportion of defaults} = \frac{\text{Number of Default}}{\text{Number Valid, Complete}}$$

2. Identify outliers (biggest offenders).
3. The upper limit (threshold) is calculated excluding outliers as:

$$UL = \text{average}(\text{proportion of defaults}) + 2(\text{stdev}(\text{proportion of defaults}))$$

The average of proportion of defaults for the data item plus two times the standard deviation of proportion of defaults

The calculated threshold is presented as percentages by multiplying the proportion by 100 and adding the percentage (%) symbol. It is available in the 'Core Field Descriptions' tab of the DQMI publication.

### Data Item Score

Data Item Score (e.g. NHS Number) = Percentage Valid, Complete for the applicable data item

The Data Item Score may be formally expressed as:

$$\text{Data Item Score} = \left( \frac{\text{Number of valid and complete records} - \text{defaults in excess}}{\text{Number of records}} \right) * 100$$

For the applicable data item.

See [Calculation of defaults in excess](#) for an explanation.

See [Appendix 1](#) for a worked example of a DQMI calculation.

### Dataset Score

Dataset Score (e.g. APC Score) = Mean of all the Data Item Scores for Percentage Valid & Complete for the applicable dataset multiplied by the coverage score for the applicable dataset.

The Dataset Score may be formally expressed as:

$$\text{Dataset Score} = \frac{1}{n} \sum_{i=1}^n \left( \frac{\text{Number of valid and complete records} - \text{defaults in excess}}{\text{Number of records}} \right)_i * C * 100$$

Where:

$n$  is the number of fields for which data was submitted for the applicable dataset and  
 $i$  is the index number of each of those fields.

$C$  is coverage calculated as  $\frac{\text{No. of periods a provider submitted data}}{\text{No. of periods a provider was expected to submit data}}$

See [Calculation of defaults in excess](#) for an explanation.

## DQMI

The DQMI is an overall score calculated for each provider; it is defined as the average of the percentage of valid and complete entries<sup>14</sup> in each field of each dataset and is proportional to the coverage<sup>15</sup>. The [excessive use of default values](#) is penalised from the valid values.

Where no data items have been submitted for a field, the percentage is treated as a null value and is not included in the calculation of the mean.

DQMI = Mean of all the Data Item Scores (for Percentage Valid & Complete) multiplied by the coverage score.

The DQMI may be formally expressed as:

$$\text{DQMI} = \frac{1}{n} \sum_{i=1}^n \left( \frac{\text{Number of valid and complete records} - \text{defaults in excess}}{\text{Number of records}} \right)_i * \frac{1}{n} \sum_{i=1}^n (C)_i * 100$$

Where:

$n$  is the number of fields for which data was submitted

$i$  is the index number of each of those fields

$C$  is coverage calculated as  $\frac{\text{No. of periods a provider submitted data}}{\text{No. of periods a provider was expected to submit data}}$

See [Calculation of defaults in excess](#)

See [Appendix 1](#) for a worked example of a DQMI calculation.

**Please note:** The DQMI is NOT equal to the mean of all the Dataset Scores. The DQMI gives equal weighting to each data item, whereas calculating the Mean of all the Dataset Scores would give a lower weighting to fields within datasets reporting on a higher number of fields, than those reporting on a lower number of fields.

## Contents of spreadsheet

<sup>14</sup> 'Number Valid, Complete' – a count of valid and complete records for that data item (e.g. ethnic category) for each Provider. This count includes default values.

<sup>15</sup> Coverage is the degree to which data have been received from all expected data suppliers (see [Calculation of Coverage](#)).

### **‘Title Sheet’ tab**

The ‘Title Sheet’ includes an Introduction and list of linkable contents to navigate to individual tabs in the DQMI publication around the workbook, including the ‘Provider DQMI Values’, ‘dataset’, ‘Core Field Values’ and ‘Core Field Description’. Those tabs, in turn, contain links to return to the ‘Title Sheet’ and ‘Provider DQMI Values’ tabs.

### **‘Provider DQMI Values’ tab**

The ‘Provider DQMI Values’ tab contains the following information for each organisation that submitted data:

- The current DQMI score (DQMI-5) for quarter October – December 2016
- The previous DQMI score (DQMI-4) for the quarter July – September 2016
- The previous DQMI score (DQMI-3) for the quarter April – June 2016
- The previous DQMI score (DQMI-2) for the quarter January – March 2016
- The initial DQMI score (DQMI-1) for the annual period of January– December 2015 (note this is not, therefore, directly comparable with subsequent publications).
- The Dataset Score for the current quarter for each dataset. This information is presented over several columns.

### **‘Dataset’ tabs**

The Dataset Score are reported in individual tabs for each dataset (E.g. MHSDS Score, APC Score) in Column F to aid navigation of the data by reducing the number of columns through which a user needs to scroll. Data is presented with one row per provider, the columns being grouped together by DQMI data item.

- ‘COVERAGE’ (Columns C to E): Flags where the provider is submitting each activity month where expected:
  - ✓ Submitting current month
  - ✗ Not submitting current month
  - Not expected to submit current month

For each data item in the dataset tab, the following columns are given:

- Dataset Score ‘Percentage Valid, Complete’ (Column F): Mean of all the Data Item Score (Percentage Valid & Complete) for the applicable dataset multiplied by the coverage value.
- Data Item Score ‘Percentage Valid, Complete’ (From Column G): the ‘Number Valid, Complete’ - ‘defaults in excess’ divided by the ‘Number of Records’. Expressed as a percentage).
- ‘Number of Records’ – a count of records submitted by each Provider for the relevant quarter.
- ‘Number Valid, Complete’ – a count of valid and complete records for that data item (e.g. ethnic category) for each Provider. This count includes default values.

- ‘Defaults’ – a count of records containing default values for that data item for each Provider.
- ‘Defaults in Excess’ – a count of records containing default values which are treated as invalid, for that data item for each Provider, and subtracted from the Percentage Valid, Complete when calculating the DQMI. (See [Calculation of defaults in excess](#))

### ‘Core Field Values’ tab

The data in this tab is an aggregation of the provider-level data in the ‘Dataset’ tabs - by provider. The following fields are presented per dataset and data Item:

- ‘Number of Records’ – a count.
- ‘Number Valid & Complete’ – a count.
- ‘Number Default’ – a count.
- ‘Defaults in Excess’ – a count.
- ‘Percentage Valid & Complete’ – a percentage

### ‘Core Field Descriptions’ tab

The data in this tab gives the core fields included in the DQMI and the explanation how the validity and default have been calculated. The following fields are presented per dataset and data Item:

- ‘Dataset’
- ‘Data Item’
- ‘Plain English Description’
- ‘Definition of Validity’
- ‘Threshold for proportion of defaults’

See Core Field Descriptions in the ‘Core Field Descriptions’ tab of the DQMI publication accessible via <http://content.digital.nhs.uk/dq> for an explanation of each field.

## Suppression Rules

In order to address the data disclosure risk inherent in the presentation of small numbers, suppression has been applied to the data used in the DQMI. The first stage of suppression is that all individual data items with a total number of records less than 25 at the organisation level have been excluded from the data, except for when the corresponding coverage value is zero. Further suppression is then applied as detailed below;

### APC, AE and OP

1. Values between one and five replaced with an asterisk.
2. The corresponding percentages for the percentage of Valid and Complete records are suppressed when the suppression takes place in the ‘Number of Records’ or ‘Valid and Complete’ columns.



3. The corresponding percentages for the percentage of Valid and Complete records are recalculated to treat all 'Excess in Defaults' as valid (or zero) when the suppression takes place in 'Defaults' or 'Defaults in Excess' column.
4. When only the 'Defaults in Excess' column is suppressed then 'Defaults' are replaced with an asterisk.

## **DID**

1. Values between one and two replaced with an asterisk.
2. The corresponding percentages for the percentage of Valid and Complete records are suppressed when the suppression takes place in the 'Number of Records' or 'Valid and Complete' columns.
3. The corresponding percentages for the percentage of Valid and Complete records are recalculated to treat all 'Excess in Defaults' as valid (or zero) when the suppression takes place in 'Defaults' or 'Defaults in Excess' column.
4. 'Percentage of valid and complete' is presented to the nearest whole number.
5. When only the 'Defaults in Excess' column is suppressed then 'Defaults' are replaced with an asterisk.
6. All remaining integer values are rounded to the nearest five.

## **CYPHS, MHSDS, IAPT and MSDS**

1. Values between zero and four replaced with an asterisk.
2. The corresponding percentages for the percentage of Valid and Complete records are suppressed when the suppression takes place in the 'Number of Records' or 'Valid and Complete' columns.
3. The corresponding percentages for the percentage of Valid and Complete records are recalculated to treat all 'Excess in Defaults' as valid (or zero) when the suppression takes place in 'Defaults' or 'Defaults in Excess' column.
4. 'Percentage of valid and complete' is presented to the nearest whole number.
5. When only the 'Defaults in Excess' column is suppressed then 'Defaults' are replaced with an asterisk, with the exception for when 'Defaults in Excess' is equal to zero.
6. All remaining integer values are rounded to the nearest five.

Suppression rules differ to retain consistency with the standard Statistical reporting for these datasets, available in [Appendix 3: Section Links to Information and Reports for Individual Datasets](#).

## **Provider Codes**

For the purposes of the DQMI, all five character provider codes have been truncated to their first three characters and their data aggregated, with the exception of Hospices (five character codes commencing with an '8').

## **Changes since last publication**

## **Changes to the DQMI Calculation**

No changes have been made to the calculations in DQMI-5

## **Changes to the Data**

No changes to data have been made in DQMI-5

## **Changes to presentation**

No changes to presentation have been made in DQMI-5

## **Changes to the Field Validations**

No changes to the field validations made in DQMI-5

## **Changes to Power BI Interactive Report**

- New front page design
- Link to CSV added to all pages
- Organisation Type filter added
- Commissioning Region filter removed

## Proposed Changes for Future Publications

Proposed future changes are listed below. They are subject to confirmation, and other development requirements may take priority.

### Future Developments

#### Integrity

As the DQMI methodology matures, integrity measures will be added to the DQMI methodology. Integrity is the degree to which data satisfy the set of business rules that govern the relationships between fields, records and data assets. *For example: IF [Ward\_Type]='Maternity' AND [Gender]='Male' THEN this would be an error as a male patient would not be treated on a maternity ward*

There is no fixed timescale for when this will be implemented.

#### Time series

As the DQMI methodology stabilises, a time series functionality will be added to the DQMI publication.

#### Additional datasets

Additional datasets will be added to the DQMI as it is developed over time.

## History of changes from Previous Publications

These are now to be found in previous versions of the methodology documents accessible via the NHS Digital website data quality page<sup>16</sup>

---

<sup>16</sup> <http://content.digital.nhs.uk/dq>

## SUS Data Quality Dashboards

It is acknowledged there are some differences between the DQMI and the SUS Data Quality Dashboards. The DQMI is a quarterly publication whilst the SUS DQ Dashboards are released on a monthly basis. Going forwards it is intended that the DQMI will be aligned to the SUS monthly post-reconciliation (freeze) date for the final month within each quarter, so the DQMI publication covering the period April-June will be extracted from SUS as at the June monthly post-reconciliation date. Both the DQMI and the SUS DQ Dashboards use the most recently available data at the time of production, so in that sense neither product is timelier than the other.

As much as possible we attempt to align the DQMI to the SUS DQ Dashboards. It is not always possible to do things in exactly the same way so there are some differences. One of the main differences is that the SUS DQ Dashboards report on different time periods to those used by the DQMI, with the dashboards being produced monthly using cumulative data for the financial year whilst the DQMI is a quarterly publication using only the months within the quarter being published as at the post-reconciliation (freeze) date for the last month within the quarter. Another subtle difference is in how the DQ measures are displayed, with the dashboards showing the number of invalid records whereas the DQMI displays the number of valid records. As much as possible we have tried to align the two and going forward we will continue to try and better align the two DQ products so that there are as few differences as possible between them.

Further information on the SUS Data Quality Dashboards is available here:

<http://content.digital.nhs.uk/article/1923/SUS-Data-Quality-Dashboards>

## IG Toolkit

It is also acknowledged there are differences between the DQMI and the [Information Governance Toolkit](#) Data Quality Requirements. The Data items to support IG Toolkit Requirement 507 are found within the [Secondary Uses Assurance Requirement - 507](#) PDF document. These differences will be addressed in the future to align these tools together. Key differences are:

- The IG Toolkit doesn't include the AE dataset
- There are OP and APC data items in the IG Toolkit that are not reported in the DQMI and vice versa (though there are also a number of data items that are present in both)
- The IG Toolkit gives weightings for the fields, whereas the DQMI gives equal weighting to every field
- The IG Toolkit reports on one activity month, whereas the DQMI is quarterly
- Where there are data items in common, the validity methodology is mostly the same. However, there are some small differences to address. E.g. IG Toolkit doesn't take sensitive records into account for the Date of Birth

## Featured Topics of Interest

Topics of interests can be included in DQMI publications. For example, a featured topic of interest (TOI) was included in the DQMI-3 publication of the DQMI to report the data quality of providers excluding a diagnosis of dementia within patient episodes.

DQMI-5 will not include a TOI.

The 3rd August publication of the Maternity Services Monthly Statistics Report by NHS Digital includes a special data quality feature. The feature investigates the coverage and reporting of birth episode statistics between the Maternity Services Data Set (MSDS) and Hospital Episode Statistics (HES) for 3 separate monthly reporting periods. The analysis considers the similarity and accuracy of the data being reported by the same organisations within each dataset and provides an interactive spreadsheet with which the differences between trusts can be compared.

- <http://content.digital.nhs.uk/catalogue/PUB21280>

## Appendix 1: Example of DQMI Calculation

Dataset	Data Item	Number of records	Valid and complete	Defaults	Defaults in Excess (calculated)	Threshold for proportion of defaults (%)	Data Item Score (%) (calculated)	Coverage of Dataset (calculated)
1	1	2,500	2,400	100	50	2.0	<b>94.0</b>	<b>1.0</b>
	2	2,500	2,500	100	0	100.0	<b>100.0</b>	<b>1.0</b>
	3	2,500	2,000	100	100	0.0	<b>76.0</b>	<b>1.0</b>
2	1	1,000	900	900	900	0.0	<b>0.0</b>	<b>0.3</b>
	2	1,000	800	800	0	100.0	<b>80.0</b>	<b>0.3</b>
	3	1,000	900	450	450	0.0	<b>45.0</b>	<b>0.3</b>
	4	1,000	1,000	100	65	3.5	<b>93.5</b>	<b>0.3</b>
3	1	10,000	10,000	-	-	-	<b>100.0</b>	<b>0.3</b>
	2	10,000	10,000	-	-	-	<b>100.0</b>	<b>0.3</b>
4	1	-	-	-	-	-	-	-
5	1	-	-	-	-	-	-	-
7	1	50,000	50,000	100	0	2.0	<b>100.0</b>	<b>0.7</b>
	2	50,000	49,000	5,000	0	100.0	<b>98.0</b>	<b>0.7</b>
	3	50,000	45,000	5,000	3,250	3.5	<b>83.5</b>	<b>0.7</b>

Note: '-' denotes that no data was expected

### Formulae

- Defaults in excess =  $Total\ Defaults - (Threshold * Number\ of\ Records)$
- Data Item Score =  $\left(\frac{Number\ of\ valid\ and\ complete\ records - defaults\ in\ excess}{Number\ of\ records}\right) * 100$
- $DQMI = \frac{1}{n} \sum_{i=1}^n \left(\frac{Number\ of\ valid\ and\ complete\ records - defaults\ in\ excess}{Number\ of\ Records}\right)_i * \frac{1}{n} \sum_{i=1}^n (C)_i * 100$

Where  $n$  is the number of data items for which data was submitted,  $i$  is the index number of each of those data items and  $C$  is coverage.

### DQMI Calculation

$$DQMI = \left[ \left( \frac{1}{12} \right) * (0.94 + 1.0 + 0.76 + 0 + 0.8 + 0.45 + 0.935 + 1.0 + 1.0 + 1.0 + 0.98 + 0.835) \right] * \left[ \left( \frac{1}{12} \right) * \left( 1 + 1 + 1 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{2}{3} + \frac{2}{3} + \frac{2}{3} \right) \right] * 100$$

$$DQMI = \left[ \left( \frac{1}{12} \right) * (9.7) \right] * \left[ \left( \frac{1}{12} \right) * (7) \right] * 100 = 47.2$$

## Appendix 2: Caveats

Caveats
<b>General</b>
All percentages are displayed to whole numbers
Percentages are calculated from the original unrounded number and presented to whole numbers
Providers with a completeness denominator <25 are not reported in the DQMI because small volumes of data have high levels of volatility.
Provider names labelled “**INVALID - SEE METHODOLOGY DOCUMENT FOR CAVEAT” in the DQMI mean that the organisation is not registered or recognised by the Organisation Data Service (ODS).
Provider names labelled with (C) at the end have closed before the time period the data it is reporting on.
Where an organisation has changed its organisation code and appears on our data using the old and new codes simultaneously, the organisation with a coverage of zero will show a tilde on their DQMI score.
Provider codes have been limited to three characters. Typically a three character code will identify an organisation whilst the two trailing characters of a five character code will identify a specific site.
For the purposes of the DQMI, our intention is to present data aggregated at the organisation level. We, therefore, choose to make use of three character codes where possible.
The rule applied is that five character codes commencing with an ‘8’ (Hospices) are retained. However, all other five character codes have been truncated to their first three characters and their data aggregated. Valid provider codes and their names have been obtained from the Org Daily corporate, a NHS Digital standard organisation reference data table by selecting organisations that were open during the reporting period.
Fields have only been checked for validity, they have not been checked for accuracy in that the individual has actually been issued with the correct value.
POSTCODE OF USUAL ADDRESS: Inconsistent default definition between the datasets. There is no clear definition for ‘Not Known’ in the NHS Data Dictionary. We are working with Data Standards to clarify this, at which point we will be able to align the default validation rule.
<b>Admitted Patient Care (APC)</b>
For Commissioning (CDS) Datasets APC, OP and AE, the data reported has been sourced from Secondary Uses Service (SUS).
Exclusions: Only finished general episodes are counted (CDS type 130) because this keeps the results in line with the existing SUS Data Quality Dashboards (which reports each maternity dataset separately).
PRIMARY DIAGNOSIS (ICD): The denominator only includes episodes within finished spells as coding often occurs only after discharge. This is the same methodology used for the SUS DQ Dashboards.
POSTCODE OF USUAL ADDRESS: Fields submitted without a space between the outward and inward components are counted as invalid in the DQMI, but valid in the SUS Data Quality Dashboards. Going forward the SUS DQ Dashboards will be amended to be consistent with this.

<b>Caveats</b>
<p>ORGANISATION CODE OF COMMISSIONER: Codes 'YDD82' and XMD00' are listed as a default on NHS Data Dictionary, however not included in the DQMI Default validation as they are being counted as invalid.</p> <p>The <a href="#">NHS England 16/17 Commissioner Assignment Method Guidance Document</a> (page 14) states YDD82 should no longer be used as an organisation code for CDS data. NHSE states that XMD00 is invalid, and only should be used during times of war conflict. This is the same methodology used for the SUS DQ Dashboards.</p>
<b>Accident and Emergency (AE)</b>
<p>Exclusions: No exclusions are applied to AE counts.</p>
<p>POSTCODE OF USUAL ADDRESS: Fields submitted without a space between the outward and inward components are counted as invalid in the DQMI, but valid in the SUS Data Quality Dashboards. Going forward the SUS DQ Dashboards will be amended to be consistent with this.</p>
<p>ORGANISATION CODE OF COMMISSIONER: Codes 'YDD82' and XMD00' are listed as a default on NHS Data Dictionary, however, are not included in the DQMI Default validation as they are being counted as invalid.</p> <p>The <a href="#">NHS England 16/17 Commissioner Assignment Method Guidance Document</a> (page 14) states YDD82 should no longer be used as an organisation code for CDS data. NHSE states that XMD00 is invalid, and only should be used during times of war conflict. This is the same methodology used for the SUS DQ Dashboards.</p>
<b>Outpatients (OP)</b>
<p>Exclusions: Administrative events (FIRST_ATTENDANCE = '5') or indication of future appointment (ATTENDED_OR_DID_NOT_ATTEND = '0') are excluded from the DMQI.</p>
<p>POSTCODE OF USUAL ADDRESS: Fields submitted without a space between the outward and inward components are counted as invalid in the DQMI, but valid in the SUS Data Quality Dashboards. Going forward the SUS DQ Dashboards will be amended to be consistent with this.</p>
<p>ORGANISATION CODE OF COMMISSIONER: Codes 'YDD82' and XMD00' are listed as a default on NHS Data Dictionary, however not included in the DQMI Default validation as they are being counted as invalid.</p> <p>The <a href="#">NHS England 16/17 Commissioner Assignment Method Guidance Document</a> (page 14) states YDD82 should no longer be used as an organisation code for CDS data. NHSE states that XMD00 is invalid, and only should be used during times of war conflict. This is the same methodology used for the SUS DQ Dashboards.</p>
<b>Children and Young People's Health Services (CYPHS)</b>
<p>The valid numbers quoted in the DQMI compare with the Valid numbers in the standard CYPHS DQ reporting, of which Other and Default are sub-categories.</p>
<p>Coverage is evaluated based on previous submissions to the dataset as there is not a master list of CYPHS providers</p>
<b>Diagnostic Imaging Data Set (DID)</b>
<p>DID Data is extracted from Radiology Information Systems (RIS). Not all RIS have access to full patient information, so not all trusts are able to submit fields such as NHS Number Status Indicator within their DID extract. However, data providers are requested to include these fields where they are available. This may account for the poor coverage of that particular field.</p>



<b>Caveats</b>
Trusts have up to 6 months to submit data relating to a particular month. For instance, anything relating to January can be submitted up to the end of July. Therefore, it is possible that the DQMI figures pertaining to January 2016 through to March 2016 - for which the deadlines for final amendment are 31 <sup>st</sup> July 2016 and 30 <sup>th</sup> September 2016 respectively - could change.
<b>Improving Access to Psychological Therapies (IAPT)</b>
The valid numbers quoted in the DQMI compare with the sum of the Valid, Other and Default numbers in the standard IAPT DQ reporting.
GENERAL MEDICAL PRACTICE CODE (PATIENT REGISTRATION): The denominator differs from other IAPT measures as all Scottish GP codes (beginning with S) are excluded from the IAPT data quality assessment.
PERSON BIRTH DATE: This field is rejected on submission if invalid, so is always 100% valid, which is why it is not included in DQMI.
POSTCODE OF USUAL ADDRESS: Note where POSTCODE_DISTRICT = 'ZZ99' is included as a Default in the DQMI; however it is not listed as a default on NHS Data Dictionary.
COMMISSIONER CODE: Note no default codes are currently defined. This will potentially change in the future.
<b>Mental Health Learning Disability Data Set (MHSDS)</b>
GENERAL MEDICAL PRACTICE CODE (PATIENT REGISTRATION): The denominator differs from other MHSDS measures as all Scottish GP codes (beginning with S) are excluded from the MHSDS data quality assessment.
ORGANISATION CODE (CODE OF COMMISSIONER): The denominator can differ from other demographic data items as it is taken from a different table in the dataset. The demographic information is submitted as part of the Master Patient Index table and is stored separately from the Commissioner History data which includes a record for each commissioner assignment.
A rolling six-month window is being used to calculate "expected to submit" coverage with the first ever MHSDS submission being in Jan-16 and MHLDDS finishing in Dec-15
PERSON STATED GENDER CODE: Note this is the only dataset to report from this Gender field, all other datasets report from PERSON GENDER CODE CURRENT.
<b>Maternity (MSDS)</b>
The valid numbers quoted in the DQMI compare with the sum of the Valid, Other and Default numbers in the standard MSDS DQ reporting
POSTCODE OF USUAL ADDRESS: Note where VALID POSTCODE FLAG = 'Y' and POSTCODE DISTRICT (MOTHER) starts 'ZZ99' is included as a Default in the DQMI, however, it is not listed as a default on NHS Data Dictionary.

## Appendix 3: Supplementary Information

### Links to Information and Reports for Individual Datasets

Information about the data quality assurance and reporting processes for individual datasets included within the DQMI is available on the NHS Digital website:

#### Children and Young People's Health Services

- Landing page: <http://content.digital.nhs.uk/maternityandchildren>;
- Commissioning Datasets (Admitted Patient Care, Outpatients, and Accident & Emergency) – Landing page: <http://content.digital.nhs.uk/sus>

#### Diagnostic Imaging Dataset –

- Landing page: <http://content.digital.nhs.uk/did>

#### Improving Access to Psychological Therapies (IAPT) –

- Landing page: <http://content.digital.nhs.uk/iaptmonthly>
- Please see the document at the following link for more information: [http://content.digital.nhs.uk/media/19211/IAPT-DQ-Measure-Rules/xls/IAPT\\_DQ\\_Measure\\_Rules.xlsx](http://content.digital.nhs.uk/media/19211/IAPT-DQ-Measure-Rules/xls/IAPT_DQ_Measure_Rules.xlsx)
- An additional document which may be useful is: <http://content.digital.nhs.uk/media/20844/IAPT-month-metadata/xls/IAPT-month-metadata.xlsx> using the “Related DQ measures” tab.

#### Mental Health Services Dataset (MHSDS) –

- Landing page: <http://content.digital.nhs.uk/mhsds>
- (Please note: information relating to the MHLDDS can be found via the following link - <http://content.digital.nhs.uk/mhldsmonthly>)

#### Maternity Dataset (MSDS) –

- Landing page: <http://content.digital.nhs.uk/maternityandchildren>
- Data quality information is available at - <http://content.digital.nhs.uk/catalogue/PUB21280/msms-mar16-exp-meta.xlsx>

### Scope of commissioning dataset submissions (CDS) and AE) and Mental Health CDS Activity

Although Mental Health activity is out of scope for (Commissioning dataset) CDS, all out-patient, admitted patient care or AE activity is required to be submitted to SUS, which includes if this activity is provided by a Mental Health organisation.

More details below:

<http://content.digital.nhs.uk/hes>

HES website (which is a direct data feed from SUS)

*“It is a records-based system that covers all NHS trusts in England, including acute hospitals, primary care trusts and mental health trusts.”*

[http://content.digital.nhs.uk/media/22271/SUS-Essentials-v83/pdf/SUS\\_Essentials\\_v8.3.pdf](http://content.digital.nhs.uk/media/22271/SUS-Essentials-v83/pdf/SUS_Essentials_v8.3.pdf)

Page 8 – SUS Essentials

*“The Department of Health mandates the collection and submission of all NHS-commissioned activity (Acute and Mental Health), including services provided for the NHS by the Independent Sector. This represents a wide range of organisations of varying sizes.”*

## SUS XML Validation

SUS can only accept data submitted as **XML** (Extensible Mark-up Language) which is a text based language for encoding structured information. It allows consistent error checking based on **NHS Data Dictionary** definitions which are expressed in an XML ‘schema’.

Data senders requiring the use of an XML translation service must select one supplier from the list of accredited suppliers before they can submit data to SUS. The terms of this contract are negotiated between the sender organisation and the XML supplier.

When submitted, interchanges are validated against the XML schema. If an interchange passes validation it is transmitted to SUS where additional validations, referred to as ‘SUS business rules’, are performed to ensure that the data can be processed.

Definitions of the Commissioning Data Sets (CDS) types and validation rules can be found on the NHS Data Dictionary and SCCI websites as follows:

The **NHS Data Dictionary** describes the structure and content of each CDS type. It includes codes that denote whether data is mandatory and the level of XML validation (such as whether format or values are validated) and the SUS business rules that are applied to each data item.

XML schemas are also available on the **NHS Data Dictionary** website:

Changes to the definitions are documented via Information Standard Notifications (ISNs), formerly known as Data Set Change Notices (DSCNs), which can be found on the SCCI website:

<http://content.digital.nhs.uk/isce/publication>

## SUS Business Rules

Where an interchange fails validation of the business rules, a SUS Interchange Failure Notification is sent to the data submitter and gives the details of the submission and the reason for failure.

The full list of SUS Business rules (rules S1 to S14) are held at the NHS Data Dictionary at **BUSINESS RULES: Secondary Uses Service Business Rules**.

## Scope of Community and Mental Health Submissions (CYPHS, IAPT, MSDS and MHMDS)

The ISNs for Community and Mental Health Datasets (CYPHS, IAPT, MSDS and MHSDS) specify the submission format for each dataset and describe the alignment of data items with the NHS Data Dictionary. Each dataset’s Technical Output Specification is also published on

the NHS Digital site (see [Links to Information and Reports for Individual Datasets](#)). This contains additional details of validations, applied at the Open Exeter submission portal, which provide instantaneous DQ information to providers of data.

Providers can use this feedback to make repeated improved submissions up to the close of the submission window. Providers are also given additional feedback in the form of a DQ notice generated by the Community and Mental Health Team. For MSDS this allows providers to identify consistency and DQ issues across periods. For IAPT and MHSDS, the providers also get this report for Primary submissions, so they have an additional opportunity to correct any DQ issues in their Refresh submissions. The data in the DQMI are based on the latest data received for each period.