

All you need to know about new processes for data submission (almost)



Benjamin Ritchie, CORC Informatics Lead and Mark Hemsley, CORC Consultant unpack the details of data submissions for CORC members in the light of the new Mental Health Services Data Set. (November 2015)

The CORC team have received requests for user-friendly information on current major changes in national data collation.¹ In particular, members have asked us to explain how the launch of the Mental Health Services Data Set (MHSDS) relates to other sources of information and data submissions. This briefing aims to provide an accessible overview as well as an update on planned changes to the CORC+ route of data submission.

Please note this document is our best attempt to summarise a complicated and changing landscape and it represents our understanding of the situation as of **November 2015**.

Key messages

CORC data submission and reporting

- CORC will continue to collate data from members to provide bespoke outcomes-focussed reports and support research into children and young people's mental wellbeing
- From January 2016 **members following the CORC+ approach** (session by session outcomes data) should submit data directly to the CORC team in a pseudonymised form on an annual basis in their allocated submission month
- For **members following the CORC Snapshot approach** (outcomes data at fixed time points), there is no change to the data submission process
- CORC plan to develop infrastructure to accept data from members in the Mental Health Services Data Set (MHSDS) format, once a year of data in the MHSDS format have been collected

Health & Social Care Information Centre MHSDS submission

- From 1st January 2016 **services that are funded or part-funded by the NHS** should collect data compatible with the MHSDS format
- From 1st February 2016 **services that are funded or part-funded by the NHS** should begin submitting MHSDS data to the Health and Social Care Information Centre (HSCIC) on a monthly basis
- At a minimum, four tables of data should be submitted to the MHSDS:
 - MHSDS Header (MHS000) (containing a single row with data about the submission)
 - Master Patient Index (MHS001)
 - GP Practice Registration (MHS002)
 - Service or Team Referral (MHS101)
- Data in other tables should be submitted if collected by services; CORC encourage members to continue collecting outcome measures data, plus information on service user characteristics and interventions that assist in interpretation of outcomes

Further information and queries

- CORC data submission – Please contact us at CORC@annafreud.org
- Mental Health Services Data Set submission – Please see the full range of documentation at www.hscic.gov.uk/mhsds or contact enquiries@hscic.gov.uk

¹ The Child Outcomes Research Consortium (CORC) is a practice research network whose members come together to make best use of outcome data.

Contents

Key messages	1
CORC data submission and reporting.....	1
Health & Social Care Information Centre MHSDS submission	1
Further information and queries.....	1
Key terms explained.....	3
What are nationally collated data sets and what are their benefits?	4
Diagram 1: Examples of nationally collated data sets relevant to children and young people’s mental health.....	5
Which data sets should children and young people’s mental health services submit data to?	6
Diagram 2: Which nationally collated data sets should our service submit data to?	6
What is the process for CORC+ data submission from January 2016?	7
In what format should CORC+ data be submitted from January 2016?.....	7
What is the Mental Health Services Data Set (MHSDS)?	8
What does it mean for my service?	8
What have CORC been doing to support the MHSDS?.....	8
Where can the full range of information and updates on the MHSDS be found?	8
Frequently asked questions (FAQs).....	9
How do we decide which data to collect in our service?	9
What data are mandatory to submit to the MHSDS?	9
What other data can be submitted to the MHSDS?.....	9
Diagram 3: Which data should our service submit to the MHSDS?	10
How are data submitted to the MHSDS?	11
How can outcome measure/questionnaire data be submitted?	11
What is SNOMED CT?	11
How can the Current View tool be submitted?	11
Where can I obtain information on licences for outcome measures / assessment tools?	12
What analyses and reports will HSCIC produce from the MHSDS?.....	12
What analyses and reports will CORC produce for members?	12
Example ways of presenting the data analysis from CORC member reports	13

Key terms explained

As a starting point, we encourage readers to familiarise themselves with the key terms below.

- **Data** – Any recorded observations or measurements about the children and young people you see in your service.
- **Data set** – A collection of observations or measurements. May consist of more than one table of data, which can be linked together by a common field (e.g. PersonID). Note that the term ‘data set’ may also be used to refer to an *output specification* (e.g. on the HSCIC website).
- **Person-level data set** – Contains separate rows (or ‘records’) of data for each person. This is different to aggregate data, in which the data relating to individual people cannot be distinguished.
- **Clinical vocabulary** – A collection of definitions of clinical terms and (sometimes) their relationships to each other, which aims to represent clinical data consistently.
- **Data file** – A collection of observations or measurements that can be accessed using a computer and may comprise all, or part of, a *data set*.
- **Data set specification** – A collection of definitions of variables/fields and their attributes and relationships to one another. E.g.

<i>Data item name</i>	<i>Data item description</i>
PersonID	A unique number that identifies a person within the data set
Name	The person’s first name
Height	The person’s height in millimetres

A data set specification defines what a data set should consist of, and aims to make it as complete, consistent and accurate as possible.

- **(Technical) Output specification** – A type of *data set specification* that defines how data need to be extracted (‘outputted’) from your service’s IT systems, often to allow it to be sent (‘submitted’) to another organisation E.g.

<i>Data item name</i>	<i>Data item description</i>
PersonID	A unique number that identifies a person within the data set
Height	The person’s height in centimetres (please round to the nearest centimetre)

- **Data collection** – The process of obtaining and recording data about the children you see in your service. Data are recorded on paper and in IT systems.
- **(National) Data collation** – The process of bringing together data collected by multiple services into one *data set*, often by requesting services submit data at regular intervals according to a standardised format (specified in an *output specification*).
- **Flowing data** – Another word for services sending or submitting data at regular intervals to another organisation.
- **Data validation** – The process of checking the extent to which data that have been submitted conform to the *output specification*.
- **Information** – Data that have been arranged, processed or analysed to provide a basis for people to create meaning from them and learn something new.

The definition of the key terms may change depending on the organisation and context, however these definitions should hold true in most situations.

What are nationally collated data sets and what are their benefits?

Organisations concerned with the wellbeing of children, young people and families collect and use data to help deliver their services, as well as meet regulatory and contractual requirements. For example, schools record the test results of their pupils to assist with tracking each pupil's progress. Data are collected and stored by services in a variety of different databases and IT systems.

There are other beneficial uses of the data service providers collect. For example, researchers may use anonymised data from a sample of schools to better understand the relationship between symptoms of mental health and educational attainment. These sorts of uses of data are separate from using data to help with the direct provision of services to children and young people, and thus are often referred to as 'secondary uses'. Further examples of secondary uses include:

At a local level:

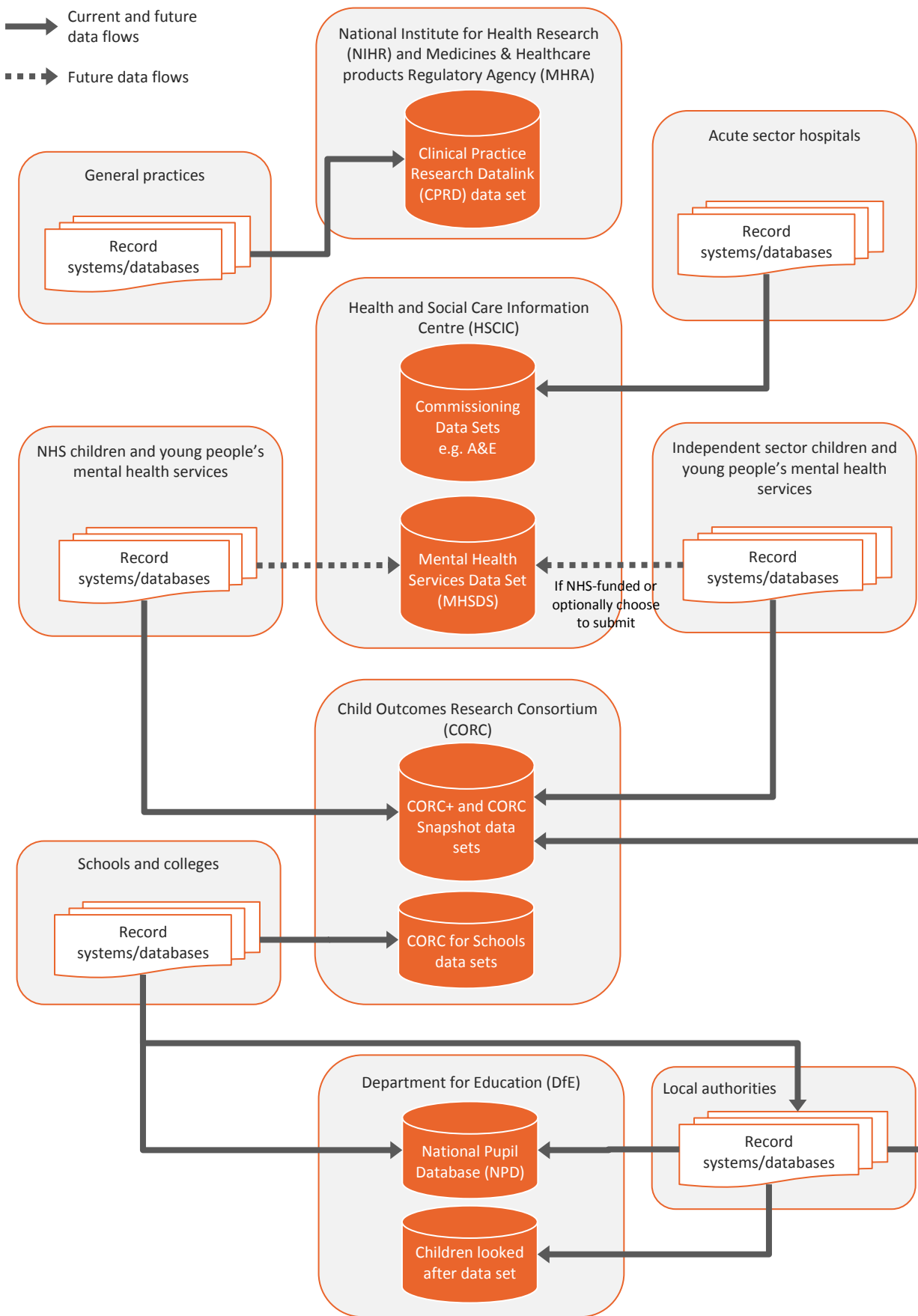
- Monitoring service provision
- Service benchmarking
- Evidencing good practice to stakeholders
- Informing service development
- Local research projects or audits
- Supporting payment systems

At a national level:

- Production of national statistics
- Designing and monitoring service improvement projects
- Evaluating government policies
- Nationally commissioned research into people's mental wellbeing
- Supporting the development of payment systems

Nationally collated data sets facilitate secondary uses of data by ensuring data collected in the different databases among services are converted to the same standardised format and collated in one place, which makes data analysis and comparing data easier. They have been established by government departments, research institutes and networks of organisations such as CORC. Several nationally collated person-level data sets relevant to the mental health and wellbeing of children and young people are shown in Diagram 1, along with the paths data travel to reach them ('data flows').

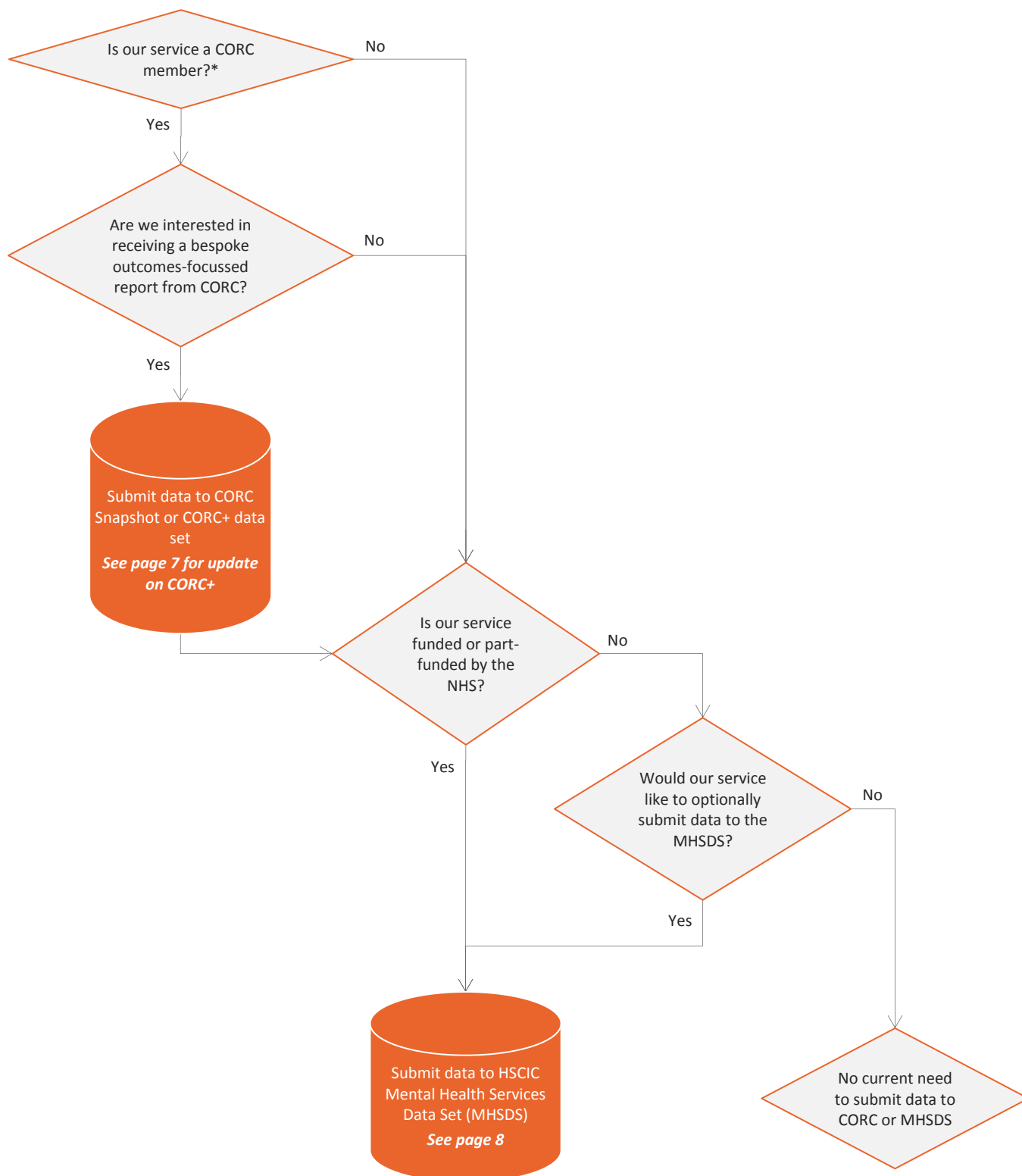
Diagram 1: Examples of nationally collated data sets relevant to children and young people’s mental health



Which data sets should children and young people’s mental health services submit data to?

This depends on several factors, including whether your service would like to receive a CORC report, and its sources of funding. Diagram 2 offers a simplified overview.

Diagram 2: Which nationally collated data sets should our service submit data to?



* A list of CORC members can be found at www.corc.uk.net/about/corc-collab-services. If you would like to understand how CORC membership can support you or those providing services in your area, please contact us at CORC@annafreud.org.

What is the process for CORC+ data submission from January 2016?

Many of you will be aware that the CORC team were commissioned by NHS England to centrally collate, analyse and report on the data from the Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) programme. This aspect of our role in this programme is reaching completion in December 2015. If you are members following the CORC+ approach, you will be aware that you have been able to utilise the same process of submitting data as CYP IAPT sites, which was via the secure data storage organisation, MegaNexus.

From January 2016, you will be pleased to hear that the CORC team will be able to **accept your CORC+ data directly**.

Reporting will continue to be on an annual basis. CORC+ data submission will move from quarterly to **annually** (in line with the CORC Snapshot approach to data submission). Members are allocated a particular month of the year in which to liaise with the CORC team on data submission and development of their bespoke report.

Come 2017, if you are also submitting data to HSCIC you will have built up a year of data in the MHSDS format—we plan to develop infrastructure to receive those data from that point.

In what format should CORC+ data be submitted from January 2016?

For the purpose of providing data for the CORC team to develop an outcomes-focussed report, we request that the data are **pseudonymised** before submission to CORC. This is a type of anonymisation, which means that service users cannot be identified, although data relating to individual service users can be differentiated in the data set (i.e. it is an anonymised person-level data set).

An **updated CORC+ data set specification** will be made available to members from **mid-December 2015**. It will be very similar to the CYP IAPT data set specification that members following the CORC+ approach are familiar with, with a few changes to pseudonymise the data.

CORC are using their links with University College London (UCL) to develop the capacity within the UCL Data Safe Haven to safely store personally identifiable data for projects that require such data. If your service encounters challenges with pseudonymising your CORC+ data, please get in touch with the CORC team at CORC@annafreud.org to discuss the arrangements that would need to be made for submitting personally identifiable data (e.g. gaining informed consent from service users).

Please note there is no change to the data submission process for members following the CORC Snapshot approach.

What is the Mental Health Services Data Set (MHSDS)?

The MHSDS is a national data set that will be collated and stored by the Health & Social Care Information Centre (HSCIC) from early 2016. CORC have been working closely with colleagues at HSCIC to develop and refine this important new data set and ensure it includes outcome data in ways that support best practice. The MHSDS will contain data relating to people of all ages in contact with mental health and learning disabilities services in England.

The MHSDS supersedes the following data sets:

- Mental Health and Learning Disabilities Data Set (MHLDDS)
- Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) Data Set
- Child and Adolescent Mental Health Services (CAMHS) Data Set (N.B. this existed as an output specification only, as data submissions were not implemented)

Services which have submitted data to the MHLDDS or CYP IAPT Data Sets should submit data to the MHSDS from early 2016 (see section below for the key dates). The output specification for the MHSDS is called the 'Technical Output Specification'.² This defines the data to be extracted ('outputted') from services, in order to be sent ('submitted') to HSCIC each month. Please note that the term 'Mental Health Services Data Set' may be used interchangeably to refer to both the output specification, and the data set itself. A service's IT systems may store a data item in a different format to the corresponding data item in the MHSDS output specification. In these cases the MHSDS data item can be derived by 're-formatting' or 'mapping' the data (e.g. re-formatting a date from 15/10/15 to 2015-10-15).

What does it mean for my service?

If you are funded or part-funded³ by the NHS: **From 1st January 2016** your service should collect data to submit to the MHSDS. **From 1st February 2016**, your service should begin submitting these data to HSCIC on a monthly basis. Please see the Frequently Asked Questions section below for suggestions on how to decide which data to submit. We recommend CORC members talk with their IT staff and system suppliers about the technicalities of flowing data.

If you are not funded by the NHS: Organisations that provide mental health services, but are not funded by the NHS, may optionally submit data to the MHSDS.

Services may review various organisational assessment and planning tools to help implement processes of data collection, extraction, submission and use relating to the MHSDS.⁴

What have CORC been doing to support the MHSDS?

CORC have been, and continue to, liaise with NHS England and HSCIC to ensure important elements of the CYP IAPT data specification are appropriately carried forward into the MHSDS output specification, and that the perspectives of the children and young people's mental health community are taken into account in the data set design.

In recent months, CORC looked into how the CYP IAPT questionnaires were being incorporated. Several potential issues were discussed with HSCIC and NHS England, and action is underway to resolve them.

Where can the full range of information and updates on the MHSDS be found?

Please refer to www.hscic.gov.uk/mhsds and www.hscic.gov.uk/mental-health/whats-new
For any enquiries email HSCIC at enquiries@hscic.gov.uk

² The current version of the Technical Output Specification and User Guidance may be downloaded from www.hscic.gov.uk/mhsds

³ The [Information Standards Notice](#) states that "Data submission is mandatory for all instances of care that is [sic] wholly or partially funded by the NHS"

⁴ An Implementation Planning Template, Readiness Assessment Tool and System Conformance Checklist may be downloaded from www.hscic.gov.uk/mhsds

Frequently asked questions (FAQs)

How do we decide which data to collect in our service?

A number of factors influence decisions on the data your organisation collects. These include the capacity to collect data (in terms of IT infrastructure and human resources), and responses to various local and national needs for data (including commissioner or other funder requirements). The latter include needs for data to (i) directly support the provision of advice, help and treatment for service users (e.g. their contact details, clinical notes), (ii) meet regulatory and contractual obligations (e.g. as agreed with commissioners), (iii) support current initiatives and projects (e.g. quality improvement, research), and (iv) contribute to the refinement of national standards.⁵

It is important to note that with the exception of the 'mandatory tables', the MHSDS output specification is not intended to compel services to collect particular data in its own right. This has been a source of confusion for some services in the past.

What data are mandatory to submit to the MHSDS?

This has caused much confusion on the ground. **There are only four 'mandatory tables' that must be submitted.** If they are not, the submission will fail data validation and will be rejected.

The four 'mandatory tables' are:

- MHSDS Header (MHS000) (Note that this contains a single row with data about the submission)
- Master Patient Index (MHS001)
- GP Practice Registration (MHS002)
- Service or Team Referral (MHS101)

The MHSDS Technical Output Specification and accompanying User Guidance contain further details.⁶

What other data can be submitted to the MHSDS?

There are 44 'non-mandatory tables'. Together with the four 'mandatory tables' these make up the whole data set specification. 'Non-mandatory tables' should be submitted if the data items within them are collected in your service. The primary diagnosis table (MHS604) is one example. If your service collects primary diagnosis data for service users, then this table should be submitted. If your service does not collect primary diagnosis data, then this table does not have to be submitted.

Within the **non-mandatory tables** there are some mandatory **data items** that must be included, **should you decide to submit that table.**

For example, if you are submitting the primary diagnosis table (MHS604), then you must submit the following mandatory data items within it:

- Service Request Identifier
- Diagnosis Scheme in Use
- Primary Diagnosis (Coded Clinical Entry)

⁵ E.g. the Access and Waiting Time Standard for Children and Young People with an Eating Disorder in 2016-17

⁶ The current version of the Technical Output Specification and User Guidance may be downloaded from www.hscic.gov.uk/mhsds

Other data items in the Technical Output Specification are designated as ‘required’ data items. These should be submitted if they are collected in your service. Continuing the primary diagnosis table (MHS604) example, if you collect primary diagnosis date, which is a required data item, then it should be submitted. However, if you collect primary diagnosis but not primary diagnosis date, then the primary diagnosis date does not have to be submitted.

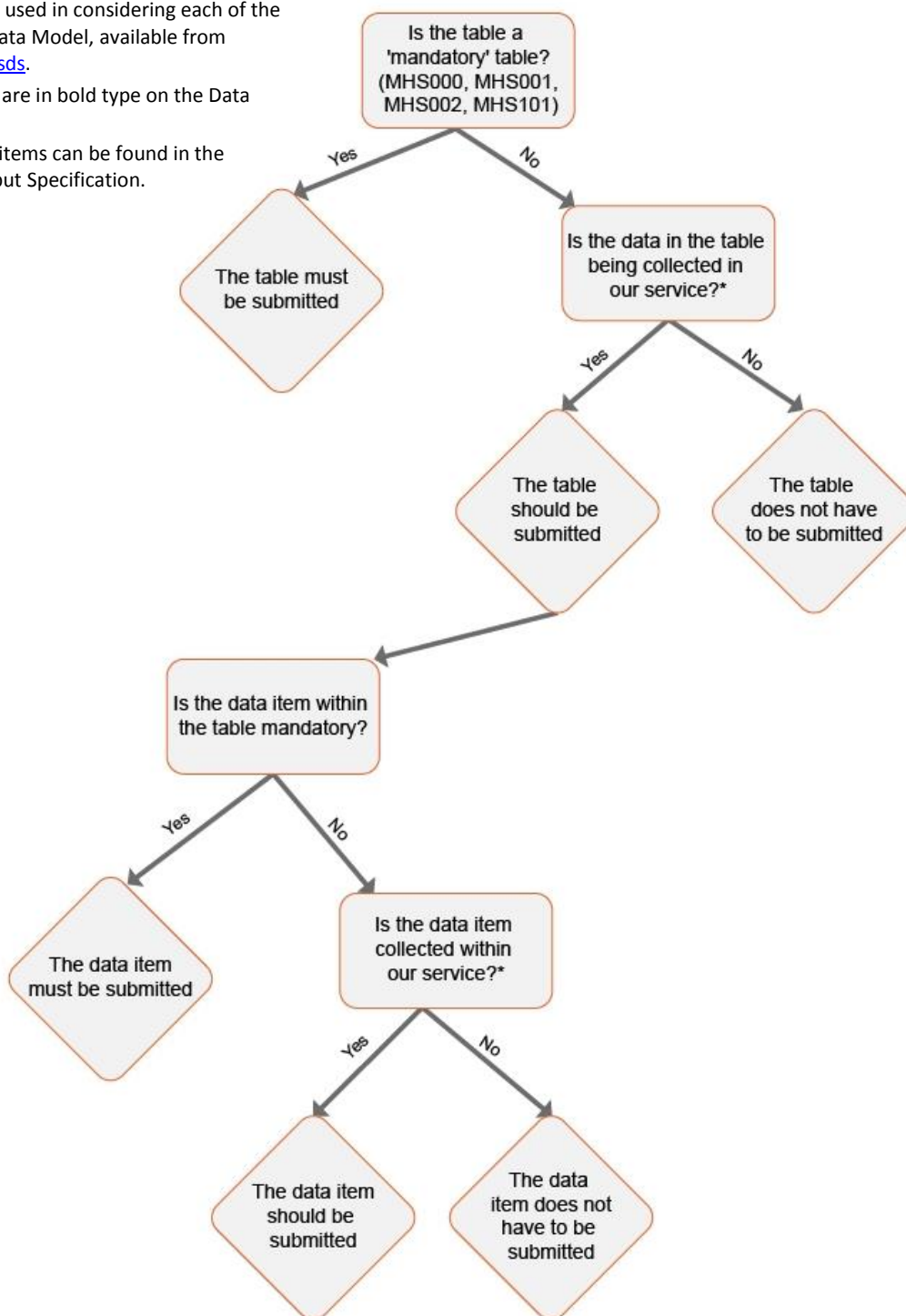
Diagram 3 aims to summarise the frequently asked question of which data should be submitted to the MHSDS.

Diagram 3: Which data should our service submit to the MHSDS?

This flow chart may be used in considering each of the tables in the MHSDS Data Model, available from www.hscic.gov.uk/mhsds.

Mandatory data items are in bold type on the Data Model.

Further detail on data items can be found in the MHSDS Technical Output Specification.



*Please include consideration of data collection in your service from which the MHSDS data item(s) could be derived through re-formatting, or mapping the data values to the MHSDS data values.

How are data submitted to the MHSDS?

Please refer your IT staff to HSCIC's Technical Guidance, which details the infrastructure and information governance requirements.⁷

In brief, data must be submitted through a website known as the Bureau Service Portal, and the steps to gaining access to it are set out in section 3 of the Technical Guidance,⁷ including the requirement to have an N3 internet connection.⁸ An organisation with an N3 connection may submit data on behalf of another provider organisation. In this scenario of data submission, the provider organisation and the submitting organisation are identified through the MHSDS Header (MHS000) table.

In order to upload data to the Bureau Service Portal services must place the data inside a Microsoft Access database known as the MHSDS Intermediate Database, which can be obtained from HSCIC by emailing enquiries@hscic.gov.uk. For a particular submission, services must place all of the data into one MHSDS Intermediate Database file before uploading it. Section 4.2 of the Technical Guidance contains further details.⁷

How can outcome measure/questionnaire data be submitted?

Via two different tables:

- MHS606 – coded assessment (referral) – For measures that are completed outside of a 'care activity'/session. E.g. If service users complete an Experience of Service Questionnaire (ESQ) after their last session.
- MHS607 - coded assessment (care contact) – For measures given during a 'care activity'/session. E.g. if an Outcome Rating Scale (ORS) questionnaire is used in a session with a care co-ordinator.

There is also a table called 'anonymous self-assessment' (MHS608). This may be employed in future for the submission of data from questionnaires completed anonymously.

Outcomes measure data must be submitted using the associated SNOMED CT codes (see next FAQ).

What is SNOMED CT?

SNOMED is a clinical vocabulary that is being supported/used by NHS England and various countries around the world. The idea is to provide a commonly used language that will traverse across disciplines, database providers and even countries.

SNOMED codes, known as 'concept IDs', should be used to submit outcome measures data to the MHSDS. This is achieved through a one-to-one mapping between total, subscale or item scores from questionnaires and concept IDs. For example, the SNOMED concept ID that identifies a Children's Global Assessment Scale score is '860591000000104'. A table of the scores, SNOMED concept IDs, and permitted values is contained in the MHSDS Technical Output Specification.⁹ We would recommend CORC members talk with their IT staff and technical suppliers to ensure their outcome measures data can be mapped to the SNOMED concept IDs and submitted in this format.

If you would like to learn more about SNOMED CT, HSCIC provide a number of resources.¹⁰

How can the Current View tool be submitted?

Current View data can be submitted via the coded assessment (referral) and coded assessment (care contact) tables (see FAQ above on "How can outcome measure/questionnaire data be submitted?"). Each item within the Current

⁷ The current version of the Technical Guidance may be downloaded from www.hscic.gov.uk/mhsds

⁸ <http://n3.nhs.uk>

⁹ The current version of the Technical Output Specification may be downloaded from www.hscic.gov.uk/mhsds

¹⁰ <http://systems.hscic.gov.uk/data/uktc/training>

View tool has a corresponding SNOMED CT code, which can be found in the MHSDS Technical Output Specification.⁹ SNOMED CT codes that map to the Current View complexity factors are not on there at the moment. It is expected that these will be included in a future update.

Where can I obtain information on licences for outcome measures / assessment tools?

Please see the website <http://systems.hscic.gov.uk/clinrecords/nccr> or email HSCIC directly at enquiries@hscic.gov.uk for information on licences to allow providers and IT suppliers to incorporate outcome measures / assessment tools into electronic systems.

What analyses and reports will HSCIC produce from the MHSDS?

At the present time, the level of information available on outputs of the MHSDS is limited and they are at different stages of development. The lowest level of granularity anticipated from the proposed reports is the level of the overarching organisation.

Data submission and quality notices will likely be provided to data submitters within a short time frame after submission, and will include counts of data items submitted and other information to help judge data quality. Processed data extracts may be referred to as 'reports' but a more accurate description would probably be 'processed data files'. It would be up to local services to decide on how these may be used, which will depend on the capacity and capabilities available locally for analysis and interpretation. An online tool provided by HSCIC (iView) allows users to generate tables, charts, dashboards and comparisons. HSCIC plan to make MHSDS aggregate data available in iView through a staged approach, based on feedback from stakeholders. Thus the particular information available and granularity will become clearer over time. Details of the content and format of reports on HSCIC's analysis of MHSDS data are not yet available. Until the quantity of data in the MHSDS builds up, it is likely that these will focus on data quality and some exploratory analyses.

What analyses and reports will CORC produce for members?

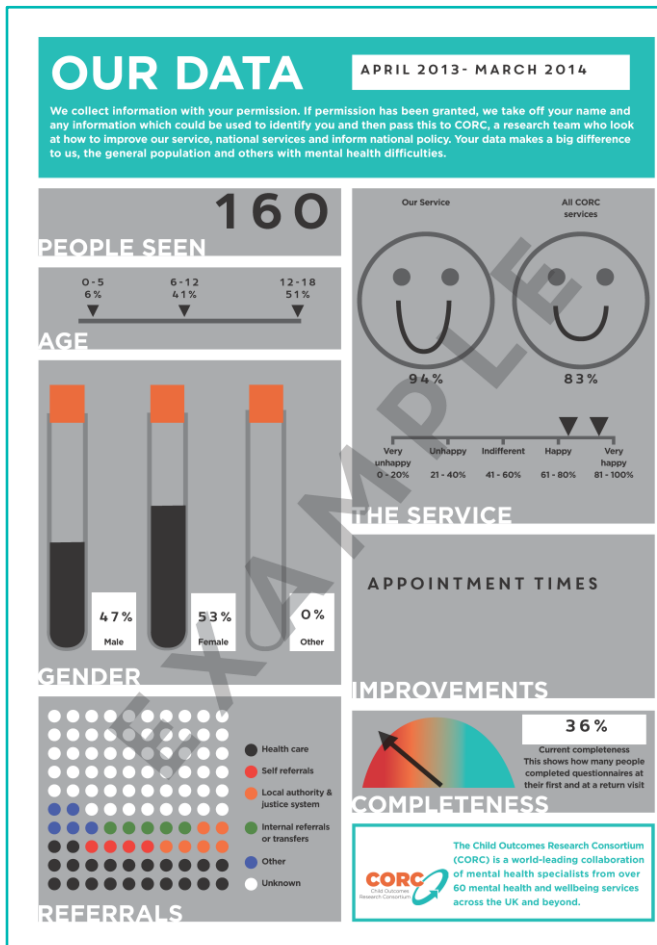
CORC provides a range of reports and analysis. The team work with members, young people, statisticians and designers to make reports accessible and useful for informing quality improvement cycles, discussions between funders and providers, and other initiatives.

Annual **Service-level** and **Team-level Reports** aim to analyse, interpret and present data in such a way as to facilitate discussions on how services may be improved. The CORC Snapshot and CORC+ data sets support analysis that allows members to see how their service users are doing compared with other member services. The reports include details of teams within services and can potentially consider differences between them subject to need. Each report is bespoke and agreed with CORC members to fit with their circumstances, whether they are an NHS, voluntary sector or other provider. Examples of ways of presenting the information provided in a report are shown on the following page.

In 2016 the CORC team will be piloting an **Area Report for Commissioners**, which looks across providers' data on an area level. This will draw on CORC's expertise and experience in supporting CYP IAPT collaboratives with dashboard reports. CORC is also developing **specific school reports** to assess mental health across schools and the impact of mental health provision in schools.

For further information on how these services can help your organisation please contact your local CORC Regional Support Officer or CORC@annafreud.org.

Example ways of presenting the data analysis from CORC member reports



The graph below displays the parent reported ESQ for Erinsborough.

