

***NHS WALES INFORMATICS
SERVICE***

**DATA QUALITY ASSURANCE
NATIONAL STATISTICS**

Version: 2.0
Date: 3rd November 2016

Document History

Document Location

The source of the document will be found on the Programme's folder under:

<\\Gig04srvipufs01\user data\ISD\Information Standards\IS3 Data Quality\IS3-5 Official Docs\DQ Assurance>

Revision History

Date of this revision: 03/11/2016

Version No.	Revision Date	Summary of Changes	Changes Marked
0.1	23/06/2014	First draft	n/a
0.2	23/06/2014	Addition of information from the Data Acquisition team	No
0.3	25/06/2014	Addition of procedures for aggregated flows	No
1.0	27/06/2014	Finalised document to share	No
1.1	03/10/2016	Updated to incorporate new additional data flows and standards	No
2.0	03/11/2016	Following review by ISD leads	No

Approvals

This document requires the following approvals:

Name	Date of Approval	Version
David Hawes (Information Standards Manager, NWIS)	24/06/2014	0.2

Distribution

This document has been distributed to:

Name	Date of Issue	Version
John Morris (Head of Health Statistics and Analysis Unit, WG)	27/06/2014	1.0

Contents

Document History	2
Document Location	2
Revision History	2
Approvals.....	2
Distribution.....	2
Purpose	4
Background.....	4
Data Quality Standards	5
Data Timeliness	5
Resubmitted Data.....	5
Monthly Resubmissions	5
Rolling Resubmissions	5
Annual Resubmissions.....	6
Data Processing	6
Data Quality Checks	6
Specific checks.....	7
Data Completeness	8
Regular Data Quality Monitoring	8
Reporting and Publishing Quality Assurance	9
Data Release.....	10
Further Information	10

Purpose

The Information Services Division (ISD) of the NHS Wales Informatics Service (NWIS) provides a central data processing, analysis and publishing service for NHS Wales. A key element of this process is to ensure that the data being processed is of suitable quality to maintain the integrity of the database which, in turn, enables the reporting of accurate and therefore meaningful health information.

One of the obligations of ISD is to provide information for the Welsh Government (WG) to publish National Statistics. This document describes the process for assuring the quality of this information.

Background

ISD processes data from a number of sources including patient-level and aggregated data sets including:

- Patient-level
 - **Admitted Patient Care data set (APC ds)** – hospital admissions
 - **Emergency Department Data Set (EDDS)** – A&E attendances
 - **EDDS Daily** – a daily submission of A&E attendance data
 - **Critical Care data set (CC ds)** – admissions to Intensive Care and High Dependency Units
 - **Outpatients data set (OP ds)** – outpatient appointments
 - **Outpatient Referrals data set (OPR ds)** – referrals to outpatients
 - **Postponed Admitted Procedures (PAP ds)** – cancelled or postponed admitted procedures
 - **National Community Child Health Database (NCCHD)** – birth and child health information
 - **Welsh National Database for Substance Misuse (WNDSM)** – substance misuse agency attendances and treatments
 - **Maternity Indicators data set (MI ds)** – initial assessment, delivery and birth data to support the national Maternity Outcome Measures and Performance Indicators
 - **Public Health Birth/Mortality File (PHBF/PHMF)** – registered births and deaths data provided by the Office of National Statistics (ONS)
- Aggregated
 - Referral to Treatment Times (RTT) – waiting times for completed patient treatment pathways and patients waiting to be treated
 - Diagnostic and Therapy Services Waiting Times (DATW) – waiting times for patients waiting for diagnostic or therapy services
 - QueSt 1(QS1) – hospital bed occupancy information and details of outpatient clinics held and cancelled

Elements of these data sets are used to compile National Statistics which include:

- Accident & Emergency Department and Minor Injury Unit attendances including time spent in department/unit (sourced from EDDS)
- Referrals to secondary care (OPR)
- Referral to treatment, and diagnostic and therapy waiting times (RTT/DATW)
- Bed occupancy, including average length of stay (QS1)
- Outpatient activity, including attendances and non-attendances for both new and follow-up appointments (QS1/OP)

ISD uses a number of procedures to assure the quality of the information required to compile this information.

Data Quality Standards

Data quality compliance is measured according to national standards. These have been developed by ISD in conjunction with stakeholders. These standards can be grouped into two distinct categories:

- **Data Validity** is a measure of whether the submitted data has been provided in the agreed format and, where applicable, whether it is populated with a nationally-agreed value. These national values are defined within the NHS Wales Data Dictionary¹ and the Welsh Reference Data Service². The content of these resources is approved and reviewed by the Welsh Information Standards Board (WISB) as part of the Information Standards Assurance process.
- **Data Consistency** considers the relationship between two or more data items and is a means of assessing whether related data items are consistent with one another (e.g. a record that indicates a male patient has given birth should be considered inconsistent and would require investigation).

Data Timeliness

There is a specified (usually monthly) submission deadline for each data set. These are set in accordance with reporting and publishing timescales. It is vital that files are received on time to align with the ISD national database refresh and to ensure that the data is made available for regular analysis, reporting and publishing, as well as for resource allocation within ISD.

A Submission Log is kept by the Data Acquisition team within ISD to record the date when each submission is received as well as the record count for each file. This enables the team to keep track of which files have been received as well as being able to compare the bottom line record count with the submission for the previous month. Any significant changes in these counts are queried directly with the submitting organisation immediately.

This information is then collated into a report which shows the number of submissions received late for each provider. Any missed deadlines are queried with the relevant organisation and regular transgression is escalated to WG policy leads and sponsors.

Resubmitted Data

Monthly Resubmissions

Where problems do arise with a particular data submission, permission can be sought to extend the deadline to 12pm the following day. This flexibility is designed to allow the data provider to resubmit files to improve data quality whilst ensuring that subsequent reporting deadlines are met. Where necessary, DATW data can be resubmitted up to one week after the submission deadline.

Rolling Resubmissions

Some data providers choose to resubmit data for previous months as part of their regular monthly submission, for example by submitting data for a rolling 6 month period each month. The purpose of this is to ensure that any data which was not available when the data submitted for the previous month was extracted is accounted for, and that any validation exercises undertaken after the previous submission date are reflected in national data. This helps to maintain a close match between the contents of local systems and data held centrally.

¹ <http://www.datadictionary.wales.nhs.uk/>

² <https://wrds.wales.nhs.uk/> (Accessible to NHS Wales users only)

Annual Resubmissions

In addition to the optional rolling resubmissions, there is also an opportunity for data providers to submit a rerun of their data for the entire financial year. This was first introduced for APC in order to capture clinically coded records which could take 3 months to be completed in accordance with national targets. Subsequently, organisations have also been able to submit full year resubmissions for Outpatients and EDDS in order to improve the quality of data held nationally following local validation. Again, this helps to maintain comparability between data held both locally and nationally.

Annual reporting of activity, such as for financial costing returns, is undertaken following receipt of these annual resubmissions in July each year.

Similarly, ONS resubmit birth and mortality information annually in August.

Data Processing

ISD manage a number of stages involved in processing the data from acquisition to publishing, including:

- Receipt of data submissions
- Loading the data
- Refreshing the national database
- Querying the database to extract the data required
- Presenting the information in reports

The aim is to correct invalid and inconsistent data at source. Submitting organisations are accountable for the quality of the data that they supply. To help them with this, these organisations need to be able to access any records they provide which are not compliant with national standards. Part of the data loading process is to facilitate this.

Data Quality Checks

Both patient-level and aggregated data extracts are submitted to the national database by Welsh healthcare provider organisations to NWIS using the NHS Wales Data Switching Service (NWDSS). In order to ensure that these data are of sufficient quality to meet reporting needs, it is necessary to employ data integrity checks to validate each data submission.

The Validation at Source Service (VASS) is a facility located on the NHS Wales intranet that enables data providers to identify errors in their data set extracts at the time of submission to NWDSS. It is the responsibility of these organisations to review these errors, correct them in their local systems and resubmit the data to NWIS.

The VASS checks are primarily used for patient-level data and are made up of three main types:

- **Data Load** checks are used to protect the integrity of the database by identifying invalid values within a record. If a data load error is triggered, the whole record is rejected by the system, preventing it from being processed through to the national database. For example,
 - **Invalid Organisation Code (Code of Provider)** – The provider code cannot be blank or invalid when checked against the list of valid organisation codes.
 - **Duplicate Key Records** – Each dataset has a set of key data items, e.g. for APC it is the Organisation Code (Code of Provider), Spell Number and Episode Number. If two or more records are submitted in the same file with the same key data items they will all be rejected.

The only exception to this if is one record is a deletion and the other is a new or amendment record.

- **Data Validity** checks are based on the associated Data Validity Standards described above, for example,
 - **Invalid Date of Birth** – The date of birth must not be blank, in an invalid format or make the patient over 125 years old at the activity date.
 - **Invalid NHS Number** – The NHS number cannot be blank or less than 10 digits. It is also run against a check digit algorithm to confirm that it is an issued NHS number. Any records with an incorrect check digit will also be classed as invalid.
- **Data Consistency** checks are equally defined by their associated Data Consistency Standards. Examples of these checks are:
 - **Inconsistent Admission Date / Date of Birth** – The Admission Date must not be before the Date of Birth.
 - **Inconsistent Referrer Code / Referring Organisation Code** –The submitted Referrer Code must correspond with the Referring Organisation Code with which the referrer is associated.

Data providers can download the details of their errors from the NWDSS site. They can resubmit their file any number of times to check the quality of their data without formally submitting it to be processed through to the national database. Having reviewed and corrected any errors and resent their file, to formally submit the data, providers must sign off their file electronically via the NWDSS website³. This provides assurance to NWIS that the submitting organisation is satisfied with the quality of the data they have supplied.

If data quality issues are identified after the file has been processed through to the national database, data providers can resubmit incorrect/incomplete data in their next submission. They do this by submitting records with an ‘amendment’ or ‘delete’ flag to allow the system to recognise that these are changes to be made to existing data rather than new records to be added.

Multiple records can also be removed from the database using a ‘bulk delete header record’. This allows the submitting organisation to specify a date period where all records for their organisation will be deleted. This is generally used as the first record in a file containing the correct records to be added in replacement. For RTT and DATW, this happens automatically when the data provider states the date period applicable to the data to be loaded. As these particular data sets are used principally for high profile National Statistics where further quality assurance is required, the checks employed are primarily Data Load checks. As these data sets contain relatively few data items, it is easier for both NWIS and the submitting organisations to maintain the high level of data quality required.

Specific checks

The process for assuring some data sets include methods which are specific to that data flow, either due to the complexity of the data or where they are more appropriate for the method of transfer.

The RTT process has a facility called ‘Instant VASS’. This differs from the VASS system described above in that providers must add a comment with supporting information against each validation error. Without this information, the file will not be processed. A monthly report containing a list of these comments is

³ <http://nwdss.hsw.wales.nhs.uk/nwdssmerge/> (Accessible to NHS Wales users only)

sent to the Welsh Government's Health Statistics and Analysis Unit for review when all submissions have been received.

A separate web-based tool is used for QS1 submissions. Rather than submitting data in a file, users input data directly onto the system. When the data has been entered, the user must undertake a validation process. This produces a list of 'Validation Queries' which includes a comparison against the figures submitted for the previous month. A comment must be entered by the user against each of these to explain the reason for the difference. Only then can they sign off the figures to be processed through to the national database.

The recently-established flow of maternity data via the Maternity Indicators data set uses SQL views to transfer data directly from local data warehouses to the national database. This data is assured using a set of data integrity checks as part of the loading process.

Substance misuse data is transferred via XML into WNDSM. A combination of data integrity or "schema" checks and VASS checks are used to assure the quality of this data.

The only validation carried out on the births and mortality files received from ONS are to check for duplicate records. Where records are identified as duplicates, the most complete of these records is processed through to the national database.

Data Completeness

The Data Acquisition team within ISD are responsible for loading the submitted data. They are in an ideal position to observe any major issues with the files we receive and address them as early as possible. The team monitor data completeness issues using two types of reports:

- **Data Load Reports** highlight any submission with a significant number of records which will not be processed through to the national database. By identifying these at this stage, details can be fed back to the submitting organisation at the earliest opportunity to allow them to correct their data and resend it before it is loaded into the national database.
- **Data Completeness Reports** show the number of records submitted for each organisation by month. Any change in these numbers can easily be identified using this information.
- The team also perform a number of 'sense' checks on the submitted data which are compiled into **Data Comparison Reports**. They are focussed on trends in the number of submitted records and are intended to identify any sharp fluctuations, for example, if a data provider has deleted records in error.

These checks form the first step of the data quality assurance process.

Regular Data Quality Monitoring

When the data passes the initial load tests, it is passed through to the national database. The database is refreshed on a regular (usually monthly) basis – once for each data set.

With the data in this state, the Data Acquisition team produce monthly Data Validity and Data Consistency reports. These are made up of a number of indicators which have been developed as part of the Data Validity and Data Consistency Standards. The measures are derived from the VASS checks themselves but, rather than identifying invalid and inconsistent records, they show the percentage of applicable records which are valid and consistent according to the associated standards. There is a nationally agreed

performance target for each indicator and values that fall below these targets are highlighted in the reports. The reports show the calculated data quality performance value based on cumulative year-to-date data for each indicator by provider, as well as a combined value for Wales. The number of records is shown for each month and there is a row to show whether the data for the reporting month was submitted on time. The target for each indicator is shown and any values that fall below these thresholds are highlighted in red.

These reports are then passed to ISD Data Quality – part of the Information Standards team. The Data Quality team use these reports to monitor compliance with the national standards. If the value for any indicator falls significantly below the target (more than 4% below the relevant target), this initiates correspondence with the submitting organisation. Each quarter, the data provider is asked to explain why their data has failed to meet the required level of quality with details of the underlying cause and actions they propose to undertake to resolve the issue. Reasons can range from simple data entry errors to more technical data extraction or hospital system issues. This feedback is collated into the supporting information which accompanies the figures and the reports are published on the NWIS Information Standards intranet site⁴.

Data Validity and Data Consistency Standards have been developed for several data sets. Reasons why these standards have not been employed for some data sets are listed below:

- **EDDS Daily** – it is not practical to monitor data at this level on a daily basis. The quality of A&E data is monitored via the monthly submission.
- **OPR ds** – Data Validity Standards were implemented in April 2010. Data Consistency Standards will be developed in due course.
- **PAP ds** – all VASS checks applied to this data set are Data Load checks except the NHS Number Data Validity check. Invalid records are rejected during the load and are not processed through to the national database.
- **NCCHD** – this data is not processed through NWDSS.
- **RTT and DATW** – as with PAP, the checks for these data sets are predominantly Data Load checks which control the processing of invalid data.
- **QS1** – this data does not flow through VASS.
- **MI ds** – a new data set was introduced in April 2016. Some data integrity checks have already been implemented and Data Validity and Data Consistency Standards are planned for the next phase of the project.
- **PHBF/PHMF** – the Informatics Service has no responsibility for monitoring the quality of these files except to maintain the integrity of the database.

Reporting and Publishing Quality Assurance

Data completeness, validity and consistency are just some of the elements of the wider data quality picture. The procedures described above are designed to capture measurable changes in the submitted data but may not pick up some of the more obscure nuances. The ISD Business Intelligence team write and run scripts to extract the information required for ad hoc and regular information requests and for formal requirements such as National Statistics. As part of the quality assurance process for the reporting outputs, checks are made against previously run reports as well as other comparable published information. The scripts and resulting reports are also checked by a second analyst to double-check for any errors. Any queries that arise from these analyses are fed back to the ISD Data Quality team for

⁴ <http://www.nwisinformationstandards.wales.nhs.uk/home> (Accessible to NHS Wales users only)

further investigation. Reports which use data which is under investigation or contains any known data quality issues is caveated accordingly.

The methodology for analysing data for some of the more formal information requests are reviewed and accredited by WISB. These include RTT and time spent in A&E calculations. This provides assurance that the methods used are consistent and that they are (and continue to be) appropriate for the specific purposes for which the information is used.

Data Release

Only when the data has passed all of the relevant quality assurance checks described above is it deemed fit to be released.

Further Information

The Information Standards internet site⁵ contains more detailed information about national Data Quality Standards.

⁵ <http://www.nwisinformationstandards.wales.nhs.uk/data-quality>