



TECHNICAL IMPLEMENTATION

Overview

INTRODUCTION

This overview provides information for implementing Standard Sets within the hospital. The mechanisms of data supply to the ICHOM Global Benchmarking Platform will be described in the appendix on 'Mechanisms of data supply'.

The ICHOM Global Benchmarking Platform helps healthcare providers worldwide to validate and compare outcome measures efficiently and at scale using quality assured ICHOM Standard Sets.

CONTENTS

Introduction	2
Contents	2
Implementation Model	3
How to upload data to the ICHOM Benchmarking Platform	4
Why batch delivery?	4
Batch process	4
Phase 1. Registration and account distribution	4
Phase 2. Set operational agreement; define upload scenario	5
Phase 3. Upload to connect-environment (connect phase)	6
Patient Identifier and pseudonym use	9

IMPLEMENTATION MODEL

The International Consortium for Health Outcomes Measurement (ICHOM) defines global Standard Sets of outcomes that matter most to patients to unlock the potential of value-based healthcare.

The technical structure around the definition of the Standard Set is articulated in the Implementation Model. Besides the variable definitions, the Implementation Data Dictionary consists of: relationships, validations, conditions, technical selection variables, keys and encryption. These features are relevant for secure and reliable data logistics.

The following documents are available to support the data upload:

- Implementation Data Dictionary: file with the structure and content of the registry in Excel to support batch-uploads via .CSV
- Example batch file: an example batch file with dummy data to illustrate how the batch-upload should be structured
- XML/ XSD (upon request)
 - XML: set expressed in machine readable format (XML)
 - XSD (XML schema document): schema definition of XML to arrange and validate batch-uploads

Attachments:

- Implementation Data Dictionary

Appendices:

- Mechanisms of data supply
- Patient identifier/pseudonym generation

HOW TO UPLOAD DATA TO THE ICHOM BENCHMARKING PLATFORM

Data can be uploaded to the platform via batch delivery. This appendix describes the batch data delivery process to provide tools for setting up and uploading a batch file successfully. This document is relevant for: healthcare professionals, data managers, quality control consultants and clinical informaticians.

Why batch delivery?

The goal of batch delivery is to extract and reuse as much data as possible from available information systems at the organization, like an electronic medical record. By reusing data, the administrative workload of healthcare professionals can be minimized. A batch is a collection of data derived from an information system put together in one file. The data dictionary defines for each Standard Set which data must be collected and uploaded.

BATCH PROCESS

The process for batch-deliveries consists of four phases:

- **Phase 1:** registration and account distribution
- **Phase 2:** set operational agreement; define upload scenario
- **Phase 3:** upload to connect-environment
- **Phase 4:** upload to production-environment

Phase 1. Registration and account distribution

During the registration phase, the organization requests access to our DataConnect environment. DataConnect is a secure web-based environment to upload data. Your main contact person can request additional accounts for the DataConnect environment via our service desk ichom@mrdrm.com with the following details for each license:

- Full name
- Job title
- Email address (within hospital domain)
- Mobile phone number (for multi factor authentication purposes)
- Standard Set

When the account has been created, a confirmation email will be received with instructions on how to log in and how the system works.

Phase 2. Set operational agreement; define upload scenario

To process the uploaded data, MRDM will configure your organization's settings, such as the upload scenario, the chosen file format and quality control. The data will be processed with an "update without NULL values scenario" (see "update scenario" below). This enables you to upload new dataset-records as well as existing dataset-records (changed or unchanged) without the risk of creating duplicates.

Update scenario

An '**update**' allows you to upload new and existing dataset-records (changed or unchanged). General principles of the '**update**' scenario:

- Existing dataset-records in the registry will be updated
- New records will be added to the database of the Standard Set
- The >dataset name<-id and id variables must be included in each upload as these id-variables contain the code that refers to your organization
- The latest delivered value (via batch or manual entry) on a variable is considered the truth and will be the value in the database
- Records that are already in the database cannot be overwritten with a NULL value (empty value). If you want to empty or delete a record you can delete this manually in the manual entry application.

Processing the batch data with an update means that it is also possible to simultaneously enter data manually. This can be useful if part of the data cannot be extracted from the EMR.

Note: do you manually change the data you have delivered in batch? This needs to be changed in the source system as well (e.g. EMR). This will prevent manual changes to be overwritten in a new batch upload with the (wrong/unchanged) value in the source system.

Phase 3. Upload to connect-environment in DataConnect (connect phase)

During the connect phase, data can be uploaded in a 'connect batch'. This batch-file is checked on compliance with the technical and (minimal) content requirements. This is also when organization specific settings are verified.

Connect phase steps

The connect phase consists of the following steps:

1. The organization delivers a batch-file to us via our DataConnect environment by choosing to upload a new data delivery with 'connect' selected as phase.
2. Based on this connect-batch the configurations for processing the batch are set. For example, within these settings the file format of the organization's preference is included (XML or .CSV).
3. The batch-file is processed using your organization's specific settings. During processing, the file is checked against required technical demands before it is validated. If the file does not comply to the technical demands, you will receive a report of the specific points that need to be corrected.
4. When the batch-file complies with the technical requirements, the (minimal) content-related requirements are validated. This step is also referred to as the batch validation.
5. After the content-related check (batch-validation), a validation report is created and shared with you. If needed, you can adjust delivered items that appear to be incorrect or incomplete according to the validation report.
6. If the batch file meets the minimal content-related requirements, you will receive a notification that the connect phase is finalized and production batches can be delivered. If the file does not meet the minimal content-related requirements, the batch needs to be adapted. Once you have made the required changes you can upload the updated file. This will be an iterative approach.

Note: during the connect phase no data will be imported to the database.

Technical and content requirements

When processing the connect-batch, the batch will be validated, based on technical and content requirements.

Technical requirements

The following technical requirements must be met for batches to be validated:

- The files always consist of the same file format as your first batch-delivery and complies to the requirements thereof (XML, .CSV)
- The extension of the file name corresponds with the file format. For example: a .CSV file with the name 'connect-delivery.csv' (not 'connect-delivery.xml')
- A consistent use of date formats when supplying .CSV files. Accepted date formats include:
 - yyyy-mm-dd
 - dd-mm-yyyy
 - When a date format is selected you must use this consistently with all batches.
- When delivering the file in .CSV-format, please use the same separator (;) and enclosure ("")

Content requirements

The content-related validation takes place after the batch has met the technical requirements. During the content-related validation we determine if the batch-file complies to the specifications of the Standard Set, as defined in the Data Dictionary (for .CSV), or as defined in the XSD file (the structure for XML-files). The requirements that must be met are:

1. The names of the variables must align exactly with the names of the variables as defined in the Data Dictionary and the XSD
2. A consistent use of data types and corresponding values. A few examples:
 - For an integer, use only digits
 - For a decimal, use digits and a dot (3.4)
3. For variables where an option needs to be selected, these must match the numbered option sets as defined in the Data Dictionary. E.g. available options in the Data Dictionary are 1, 2, 3 and 4. Option 5 cannot be delivered. It's not possible to choose multiple options for one variable.

Note: what is described above are import blocking errors which will prevent the entire batch from being processed. There may be other errors that potentially influence the completeness and quality of data. These are mentioned in the validation reports.

Phase 4. Upload to production-environment in DataConnect (production phase)

Once the connect-delivery complies to the minimal technical and substantive requirements, you will receive a message that you can upload the data onto the production-environment. This will start the production-trajectory. The production batch file needs to meet the same technical requirements as the previously delivered batch file in the connect phase. If this is not the case, the batch will be declined for technical reasons. To change your organization's technical specifications, you must submit a request via the service desk. Changing the specifications during the process means that the organization specific settings need to be adapted, which will require the batch to be run via the connect-procedure again.

Note: Standard Sets are updated regularly (typically once per year). If the Standard Set that you participate in has been updated, the batch process will be restarted. This means we will re-iterate the connect phase so we can update the settings based on the specifications of the new Standard Set. The batch-delivery needs to comply to the specifications of the new release.

Validation steps

Batch validation of a production batch follows these steps:

1. Upload a batch on DataConnect and choose the 'production' phase
2. If the technical specifications are met, your batch file will be validated on its content
3. After the content-related validation, you will receive a message with the validation reports attached. The message will state whether the uploaded batch file has a positive or negative assessment:
 - Negative assessment: we will communicate what needs to be adjusted and you are required to upload a new batch with the necessary corrections.
 - Positive assessment: the data will be imported in the database. Data-upload is successful.

Note: The minimal content-related requirements are batch import blocking errors. There may be additional errors in the delivery, that could potentially influence the quality and completeness of the delivered data. These are not blocking errors, but they are mentioned in the validation reports for the organization's consideration.

Import

When the uploaded batch file meets the content-related requirements, the data will be imported into the database and an import report will be sent to the health care provider, stating the number of records per dataset we've processed and imported.

Note: Although we import the uploaded data, there may still be some substantive errors in the uploaded data. These errors can be found in the quality report file that's part of the import report. We do urge you to look through the import report when the data has been imported to avoid any surprises later on in the process (e.g. the numbers in the reporting environment are not as expected).

PATIENT IDENTIFIER AND PSEUDONYM USE

This appendix provides information on the use of patient identifiers within the platform. In addition, technical instructions are provided on the various accepted methods and their technical implications. This information is intended for technical staff or privacy officers. This includes data engineers, IT management or any other IT and/or data related positions.

Patient identifiers within the ICHOM Global Benchmarking Platform

A patient identifier (also known as a Medical Record Number (MRN)) is a vital part of the organization's administrative systems. It consists of a unique patient-specific numeric or alphanumeric scheme, identification information and encryption mechanisms. Within the ICHOM Global Benchmarking Platform, the patient identifier is used for several functionalities. It is expected that the patient identifiers associated with the delivered data are unique patient-specific (alpha-)numeric values. The patient identifier is used for several features of the platform:

1. Prevent data duplication
2. Enable incremental updates to existing data
3. To provide a cross reference of the included patients on the platform

All patient data is encrypted according to privacy standards (HIPAA, ISO) during data handling by the data processor and anonymized during analytics through aggregation.

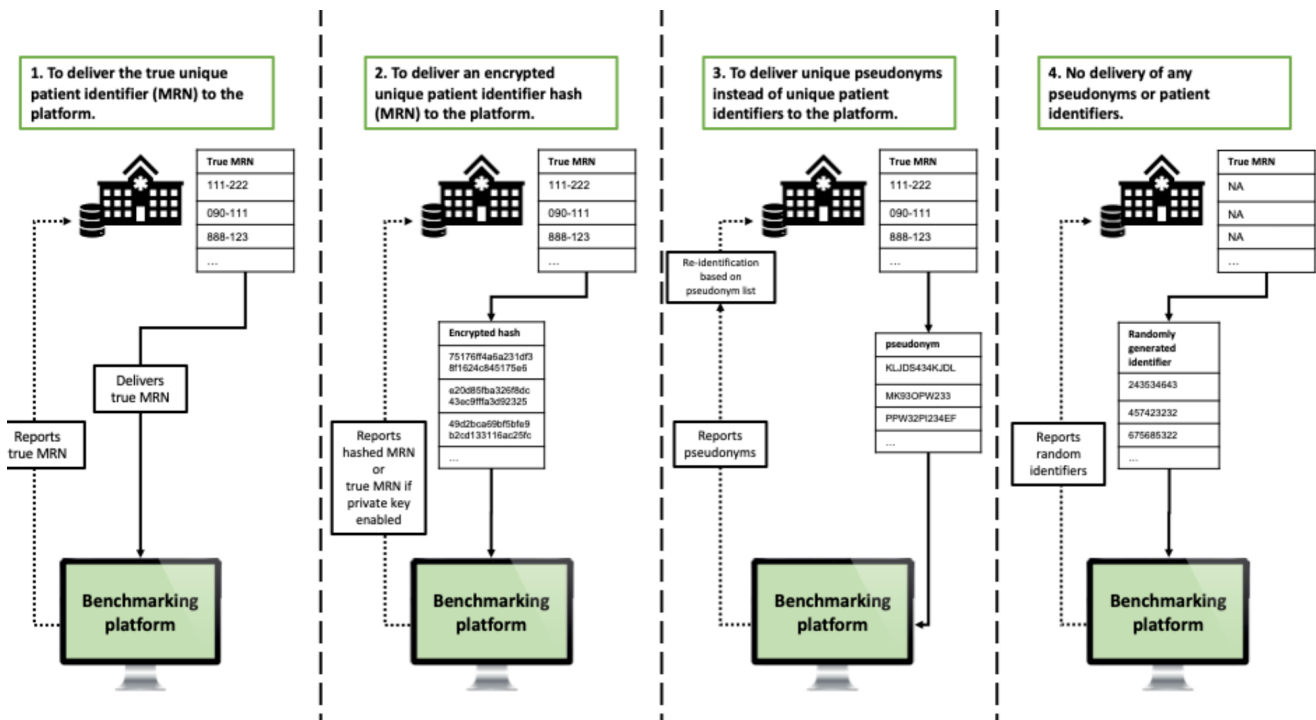
Note: the patient identifier is only shared with the data processor, not with any other parties.

Methods for providing the patient identifiers

There are four distinct methods for providing the patient identifier. It is up to the organization's preference and local legal regulations which method to use. However, not all methods support all functionalities of the platform as explained in more detail per method. The methods are described in order of support from all to less features.

The four methods are:

1. To deliver the true unique patient identifier (MRN) to the platform
2. To deliver an encrypted unique patient identifier hash (MRN) to the platform
3. To deliver unique pseudonyms instead of unique patient identifiers to the platform
4. No delivery of any pseudonyms or patient identifiers



1. True unique patient identifier (MRN)

This method provides the most straightforward way of delivering the patient identifiers to the ICHOM platform. It consists of uploading the unique patient MRN to the platform as they are present in the electronic medical record (EMR). The platform processes the delivered data based on this patient identifier and reports the results using the same patient identifiers as were delivered by the organization.

This method enables all the platform’s features to be available to the organization.

2. Encrypted unique patient identifier hash (MRN)

This method consists of delivering an encrypted patient identifier hash to the platform based on the MRN. Encryption should be performed with the SHA-2 family of asymmetric encryption functions and not with any symmetric encryption functions (like the Rijndael function) to ensure data confidentiality. The number of bits used should be at least 256 bits, utilizing the SHA-256, SHA-384 or the SHA-512 algorithms. The security of the hash function can be increased even more by using a keyed version of the hash (HMAC).

Using one of the mentioned encryption algorithms, the organization encrypts its patient identifiers into hashes before delivery to the ICHOM platform. The platform will process the encrypted patient hashes and reports the results using the same encrypted patient hashes as were delivered by the organization. The organization can interpret the results by decrypting the used encrypted patient hashes with their private key.

This method enables all the platform's features to be available to the organization.

3. Unique pseudonyms as a patient identifier

This method consists of delivering a pseudonym of the patient identifier instead of the true patient identifier. Pseudonymization involves the process of disidentifying the patient identifiers into patient pseudonyms by using a pseudonymization method preferred by the provider. An important characteristic of the generated patient pseudonyms is that they can be re-identified. To facilitate this re-identification of the patient pseudonyms, the organization should maintain a local list of pseudonym codes that map to the true patient identifiers.

Using pseudonymization, the organization pseudonymizes its patient identifiers into pseudonyms before sharing them with the ICHOM platform. The platform will process the patient pseudonyms and reports the results using the same patient pseudonyms as were delivered by the organization. The organization can interpret the results by re-identifying the used patient pseudonyms with their locally maintained pseudonym list.

This method enables all the platform's features to be available to the organization.

4. No patient identifiers

This method is less preferred than any of the above methods, since the absence of patient identifiers excludes many of the features of the platform. This method is in case that the organization does not desire to deliver patient identifiers. The platform will generate random identifiers, which consist of randomized numeric values, and use them throughout the platform. The platform will process the random identifiers and reports the results using these random identifiers to the organization.

With every new upload, the organization needs to deliver the new data and all previously shared data to prevent duplication of patients within the platform. Each data upload will overwrite the previous upload.

The organization cannot interpret the results by re-identifying or decryption, and thus incremental updates, patient specific reports, data deduplication and many other features will not be supported.