

CDM Inspection Report

Template

EHDEN Taskforce

15-11-2020

V1.1

# Executive Summary

This document describes the minimum content of the inspection report that must be created for a Data Partner by a certified Small to Medium-sized Enterprise (SME) to fulfil the inspection task.

The goal of the inspection report is to provide insight into the completeness, transparency and quality of the performed Extraction Transform, and Load (ETL) process and the readiness of the data partner to be onboarded in the EHDEN and OHDSI data networks and participate in research studies.

The template should be shared with the Data Partner at the start of the mapping process, to make sure they are able to fulfil the requirements at the end.

If the SME that is performing the inspection was not involved in the ETL implementation we advise to use a two-stage inspection process. A first inspection report can be made to provide recommendations to the Data Partner on how to improve the ETL and processes, if necessary. Ideally, this includes a site visit of the SME after providing instructions on the content of the report. The Data Partner can share this draft report with EHDEN to obtain additional input. Once the improvements have been made the final report can be created by the SME and sent to EHDEN for approval.

After this step the percentage of the grant related to the inspection milestone is released to the Data Partner.

# General information

*This section should provide general information about the data partner and the SME that is performing the inspection task.*

1. Name of the Data Partner (Legal Entity)
2. Database Name and Acronym
3. Details Contact Person Data Partner
4. Short description of the database and scope of mapping exercise
5. CDM Version
6. Last ETL execution date
7. Name of the SME performing the inspection task (Legal Entity)
8. Details Contact Person SME
9. Describe the level of involvement of the SME in the development and implementation of the ETL and the installation of the analytical infrastructure.
10. Date of the inspection report
11. Signature of the Data Partner and SME

# ETL development

*This section describes the ETL process and provides insight into the Quality Control. EHDEN will provide a CDM Inspection R package that contains quality control and performance checks on the CDM and will provide standardised output to be included in the inspection report.*

1. ETL documentation.
   * Needs to be shared as an Appendix to this report.
   * Approve the quality of the ETL documentation with respect to its completeness and level of detail per data domain. Ideally it is based on the Rabbit-in-a-Hat mapping definition document. If a staging table approach is used, its creation needs to be described in detail.
   * Does it contain enough detail on the applied business rules and are they following the THEMIS rules?
   * Compare the ETL documentation with the shared ETL code to make sure it is a correct representation of the implementation. Ideally, end-to-end tests using the Rabbit-in-a-hat testFramework.R is implemented and results are shared.
   * If allowed share the White Rabbit scan report
2. Vocabulary Mapping Process
   * Needs to be described in the ETL documentation, i.e. for which domains custom vocabulary mappings had to be performed, how this was done (e.g. were translations made, was Usagi used etc.) and quality controlled.
   * Insight needs to be provided on the coverage (% of codes and % of data) as generated by the Data Quality Dashboard.
   * The CdmInspection R package will contain queries to extract the top 50 most frequent codes for the custom mapped domains to be checked by the SME (note this is less important for vocabularies that are already part of the standardised vocabularies).
   * In addition, all the custom mappings need to be shared with the report as Excel file, to allow for random checks by EHDEN’s vocabulary team. Ideally these lists are sorted by source code frequency.
   * The CdmInspection R package will extract the vocabulary version and included list of vocabularies; this output needs to be added in the report.
3. Describe the status of the ETL code:
   * Described the technology used for implementing the ETL (SQL,R, Python etc).
   * Is the ETL code executable fully automatically or are there manual steps? If there are manual steps these need to be explained in documentation.
   * Provide feedback on the level of commenting and code structure. The minimum level of commenting contains an explanation of the sql query, R function, etc. See also [this](https://ohdsi.github.io/Hades/codeStyle.html) guidance provided by OHDSI. Code structure refers to a logical structure of the SQL/R files. We recommend that the files are name as their target table and contain all code related to that domain, e.g. insert\_person.sql, insert\_condition\_occurence.sql.
   * How is the ETL code tested? Discuss the quality controls steps or ideally share the code that executes this. Have all checks been passed? Is there a comparison available of the person count on the source and CDM and are the differences explained?
   * Is there a version control mechanism in place?
   * Share the ETL code for review by the EHDEN team (as zip file or as a link to a GitHub page).
4. Quality Control
   * Show that the Data Quality Dashboard results are 100% and check if the thresholds have been changed by doing a diff with the default.
   * Discuss with the Data Partner why the thresholds have been changed and add an overview to the report.
   * Share the DQD result zip file with the EHDEN team.
5. Maintenance
   * Describe briefly the process the Data Partner implemented to keep the data in the OMOP CDM up-to-date when new source data will become available, if the local coding systems are updated, or if new versions of the CDM will be released. Describe how versions of the CDM will be maintained over time.

# Technical Infrastructure

*This section describes the necessary input to test that the technical infrastructure is setup correctly and ready for network research*

1. Execute the performance checks in the CdmInspection R package. This will run a number of quick and longer running queries to check the performance of the system.
2. Is the cdm\_source table populated with the necessary meta data? (part of the CdmInspection R Package, and DQD)
3. Report on the technical infrastructure, e.g. database server specifications (CPUs, memory, disks), backup facilities, specifications webserver hosting ATLAS, testing environment if available etc.
4. Run script to check the installed R packages including HADES methods libraries. (part of the CdmInspection R Package)
5. Check that the following tools are available and functional: ATLAS, ACHILLES report executed (part of the CdmInspection R package), (ARACHNE is optional for now). Functionality needs to be tested by the execution of a study package (part of the CdmInspection R package), design of cohort in Atlas, generation of cohort counts in ATLAS, execution of a simple cohort characterisation in ATLAS. Instructions will be provided by EHDEN.
6. Describe the maintenance process put in place by the data partner for the tool updates.
7. Is the data source added in the EHDEN Database Catalogue and has the Achilles Export been uploaded for the visualizations? Also describe if a process has been agreed for updating this information regularly.

# Scientific preparedness

This section contains several items related to the interaction with the EHDEN/OHDSI community and training after the mapping process.

1. Describe how the Data Partner will train and educate the different users of the system in their organizaton and what the current status is of the expertise in the team.
2. Are they able to execute the ongoing OHDSI/EHDEN network studies, e.g. are there governance issues, lack of resources, etc.
3. Are there already plans to initiate research studies?
4. Are there plans to participate in OHDSI Working Groups?

# Questions

If there are questions about the creation of the inspection report do not hesitate to contact the EHDEN team using the EHDEN forum (<https://forum.ehden.eu/c/sme-forum/>)