FDM Review process

Version: 1.2

Last modified : 2023/04/20

Author: John Birkinshaw

## Purpose of document

Describes the process by which a FDM goes from being completed to being finished (ready for use).

Applies to: All SME’s and Analysts responsible for building and checking a FDM.

Roles: Subject Matter Expert (SME) , Analyst (Reviewer)

### Why do we need to review the FDM ?

The Production of a FDM can range from very simple (for single simple data tables) to extremely complex due to the nature of the tables being built into that particular FDM. The review is essential to ensure that all the stages have been completed.

## Introduction

The SME will build the FDM as per the [specifications](https://github.com/ConnectedBradford/FDM_Principles) and using their own knowledge of the subject area. Where they are not familiar with the “subject” the SME will need to use their own initiative and familiarity with the techniques developed to build the FDM. In this case the FDM may be functionally ready but some of the ancilliary data (such as the data dictionary) may not be as complete)

The following is both a checklist for the FDM that will need completing and a “how to” guide on how to go about doing each part.

|  |  |  |
| --- | --- | --- |
| FDM Name |  | |
| FDM Build date |  | |
| SME |  | |
| Analyst/Reviewer |  | |
| **Item** | **Description/process** | **Y/N** |
| FDM on GCP | Open bigquery and logon with you yhcr account |  |
|  | Is there a FDM folder on GCP ? |  |
|  | person table ? (M) |  |
|  | observation\_period table ?(M) |  |
|  | care\_site (M) |  |
|  | visit table ? (O) |  |
|  | data table(s) – prefixed with tbl\_ ? |  |
|  | data table(s) - lowercased fieldnames? |  |
|  | data table(s) - tablename\_start\_date ? |  |
|  | data table(s) - tablename\_end\_date ? |  |
|  | No tmp tables exist ? |  |
| Unit Testing | Go to Juypter Notebook and Open FDM\_*AA*Test\_v1.2.ipynb  Available on Github <https://github.com/ConnectedBradford/FDM_Specifications/>  The code contains 2 sections :  Section 1 is the configuration. Here you need to amend the targetdb to the name of the FDM being tested. Then run the section  Section 2 – runs the tests. For each test it places an entry in a temporary table: tmp\_unit\_tests\_results. This is simply the test number and the pass/fail status. Any Fails must be investigated and referred back to the SME developer |  |
|  | Unit Test scripts run and completed successfully ? |  |
| Dashboards | Is the Dashboard for the FDM available ? |  |
|  | Dashboard url: |  |
| Dashboard note | Once each FDM is completed the SME can use Sam’s script to build the dashboard views. The dashboard is a sanity check. It takes just 15 mins to build and using it the developer can quickly see if some of the data doesn’t make sense. It shows a very basic overview of the demographics of the FDM.  Note: We are looking at alternative methods of building these, for now we are using Looker (Google Data Studio) – Copying an existing report and “pointing” it at this FDM’s data. | |
| Githib folder available (In Connected Yorkshire repository) | Is there a GitHub folder for this FDM ?  There should be a folder on [Github](https://github.com/orgs/ConnectedBradford/repositories) for the code and documentation relating to this FDM |  |
| Docs on Github | Does the document folder contain the Data Dictionary ? This document describes the data , in so far as we can describe it. In many cases we simply don’t have the details of what some fields are, what the data is. Where possible we can refer the data dictionary to published sources. For example Secondary Care (SUS) data is similar to published SUS data so we can infer from that.  The data dictionary will also explain what we don’t know – and any inferences we have made. For example a column may have been dropped (that exists) in the source data if its values are all the same value , and we don’t know what this value is. |  |
| Code on Github | Does the code folder contain the actual code used by the developer to build the FDM ? |  |
| Code reviewed \*  *For Bert team - Need to think about what’s realistic here* | Have you read the code through ? |  |
|  | Could you run this code and rebuild the FDM if required ? |  |
|  | If the source data were updated with new (Delta) would the FDM code allow for this ? |  |
|  | Note : As most data will be refreshed periodically it is essential that ALL processes followed by the SME to build the FDM are repeatable and contained within the code. If there are assumptions made by the developer are these in the comments within the code (or the data dictionary) | |
| Source Data reviewed | Have you compared the FDM with the source Data ? |  |
|  | Does the FDM contain the data tables that are in the Source Data ?  This will give you a feel for the scope and scale of the FDM as well as a better understanding of what the SME has tried to do with it. This is NOT a detailed review. |  |
| Completed | If all the above is completed satisfactorily the FDM can be signed off as ready for use | |
|  | FDM checks completed : |  |