

Modelling EB Datasets

For DEBRA Ireland

Prepared by: Hildegard Franke, freshEHR Clinical Informatics Ltd.

Date: 25-Aug-2016

Table of Contents

1. Introduction 3

1.1. Explanatory notes 3

2. Analysis 4

2.1. Archetypes 4

2.2. Templates 5

3. Scope and estimated effort 5

3.1. Minimum scope 6

3.2. Full scope 6

# Introduction

This document provides details of an analysis carried out by freshEHR Clinical Informatics Ltd. on a number of datasets used by the Epidermolysis Bullosa community with a view to creating a set of shareable clinical models (archetypes and templates) based on the openEHR methodology. The document describes two different use cases (and associated work estimates):

* A reporting dataset which aligns the data content of the contributing registries and is sufficient to capture the minimum content agreed at the Dublin meeting.
* A maximal and more complex dataset intended to cover all data points in the contributing registries and support a full registry data entry application such as openApp.

The analysis is based on the following datasets (as provided by DEBRA Ireland):

* BADGEM (British Association of Dermatologists, Dermatology and Genetic Medicine)
* EB Registry Freiburg
* French EB Registry
* EB House Austria
* French Minimum Dataset
* Italian Core Dataset

The documents provided by DEBRA are embedded below for reference purposes.

## Explanatory notes

A couple of points require some additional clarification:

1. The ‘EB Registry datasets’ document contained an additional dataset (Birmingham EB database). Unfortunately, the provided screenshot only depicted a small aspect of this dataset purely related to Genetics. We therefore made the decision to omit the Birmingham EB database from our analysis.
2. Only two of the datasets provided more detail than just the names of the data points, and for the other datasets assumptions have been made what these data points actually mean. This will probably require further discussion and analysis once the modelling work starts.
3. The French Minimum Dataset (which together with the Italian Core Dataset provided extra detail) appears to be the most comprehensive dataset. Most of the content of the other datasets appears to be included in the French set, although in several cases the value sets and/or data types are not aligned.
4. We had already done a very detailed analysis of the EPIRARE Dataset as part of the PARENT project, and it is envisaged that including this dataset in the overall modelling work will not require any additional effort.

# Analysis

The details of the analysis of the different datasets are summarised in a mindmap which can be accessed at this link <http://www.xmind.net/m/7Ezm> .

We have grouped the 79 individual data points into 11 sections:

Patient Demographics

HCP/Location

Diagnosis

Pedigree

Genetics/Labs

Consent

Death Details

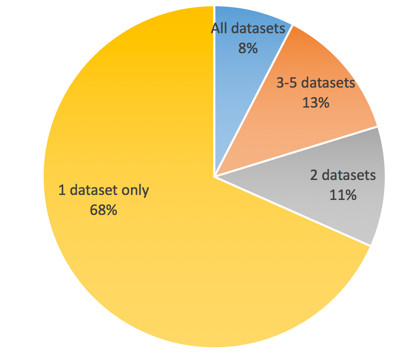
Encounter Details

Medication

Other patient data

Study/enrolment

This grouping is somewhat arbitrary, and there may be different views on whether individual data points are listed in the right section, but this is largely for analysis and illustration purposes and has no impact on the overall modelling effort required.

Each of the 79 data point has been given a marker to indicate which of the six datasets it belongs to, and the summary is as follows:

Used in all 6 datasets = 6 (8%)

Used in 3 to 5 datasets = 10 (13%)

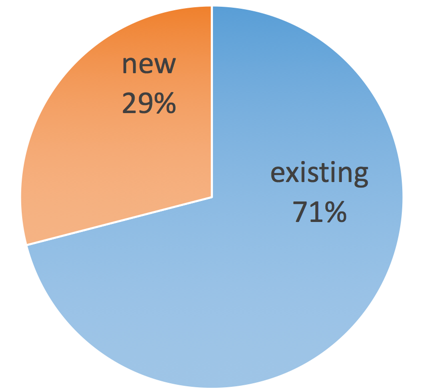
Used in 2 datasets = 9 (11%)

Used in one dataset only = 54 (68%)

## Archetypes

openEHR archetypes are shareable models of distinct clinical concepts such as diagnosis, laboratory tests, height, weight etc. Administrative concepts such as name, address, telecoms, healthcare professional details and organisation are also modelled as archetypes. Archetypes attempt to capture maximal datasets to allow for sharing across multiple clinical sectors.

Our analysis shows the following:

1. It is likely that **31 archetypes** will be required to model the entirety of the six datasets included.
2. A total of **22 archetypes** already exist as a result of the ongoing international, European and UK modelling efforts. This amounts to 71% of the total archetypes required.
3. The remaining **9 archetypes** (29%) will need to be created from scratch. When this work is undertaken, it is possible that this number may change slightly, either by combining or further splitting newly created archetypes. While the final number may be slightly different, it is not envisaged that the estimated effort (see Section 3 for detail) will change.

It should be noted that some of the new archetypes tackle some complex issues, especially around Pedigree and Genetics. It is likely that these will require significant clinical input and negotiation, which is reflected in the time estimate for these archetypes. We are aware of other clinical groups with similar interest, and there may be an opportunity for collaboration.

We estimate that a total of 6-8 days would be needed to create these archetypes, depending on scope (see Section 3 for further details).

## Templates

Templates are aggregations of archetypes to reflect datasets for specific clinical purposes such as a reporting dataset or data entry form.

The total number of required templates will depend on the scope, and will vary from a single large template (for simple reporting purposes) to 7 or 8 smaller, but more complex templates (as would be required to support a full registry application, such as openApp).

Complex or large templates usually take between one and two days to create, so the estimated effort would be between 1 and 10 days, depending on scope.

# Scope and estimated effort

We have provided an analysis of two use cases:

* The absolute minimum to provide a usable model for reporting purposes.
* The effort required to model all the datasets in their entirety and capable of supporting registry app development.

## Minimum scope

We would propose that the following is required as a minimum for basic reporting purposes:

|  |  |  |
| --- | --- | --- |
| **Topic** | **New archetypes required** | **Estimated modelling effort** |
| Single Template | N/A | 1 day |
| Basic demographics | No | included in template effort |
| HCP/Location | No | included in template effort |
| Diagnosis | Yes | 1 day |
| Pedigree | Yes, but minimal [[1]](#footnote-1) | 0.5 days |
| Genetics | Yes, but minimal 1 | 0.5 days |
| Consent | Yes | 1 day |
|  | **Total modelling effort** | **4 days** |

In addition to the modelling effort we would add a further 1 day for discussion and review.

## Full scope

The full scope would provide models to cover the content of all six datasets that were analysed for this document, and would be appropriate for registry app development, such as openApp. Further discussion with the app developer would be required to optimise the modelling.

|  |  |  |
| --- | --- | --- |
| **Topic** | **New archetypes required** | **Estimated modelling effort** |
| Multiple Templates | N/A | 10 days |
| Full demographics | Yes | 0.5 days |
| HCP/Location | No | included in template effort |
| Diagnosis | Yes | 1 day |
| Pedigree | Yes | 2 days |
| Genetics | Yes | 2 days |
| Consent | Yes | 1 day |
| Death | No | included in template effort |
| Encounter | No | included in template effort |
| Medication | Yes | 0.5 days |
| Study/enrolment | Yes | 1 day |
|  | **Total modelling effort** | **18 days** |

In addition to the modelling effort we would add a further 3 days for discussion and review.

1. As discussed in Section 2.1, modelling Pedigree and Genetics fully requires significant clinical input and negotiation. It is, however, possible to create a much simpler version for minimum scope purposes. [↑](#footnote-ref-1)