

Modelling and Mapping GEL Dataset

For North Thames Genomic Medicine Centre

Prepared by: Hildegard Franke, freshEHR Clinical Informatics Ltd.

Date: 21-Aug-2016

Table of Contents

1. Introduction 3

1.1. Partnership with Operon Ltd. 3

2. Modelling and mapping GEL data 4

2.1. What will freshEHR Clinical Informatics Ltd. deliver? 4

2.1.1. Archetypes and templates 4

2.1.2. Sandbox environment 4

2.1.3. Documentation 4

2.1.4. Support 4

2.2. When will we deliver? 5

2.2.1. Overall timescales 5

2.2.2. Exploratory phase 5

2.2.3. Phased delivery of remaining models 5

2.3. What do we require from North Thames Genomic Medicine Centre? 6

2.3.1. Access to domain experts / clinicians 6

2.3.2. Guidance on lab terminology and integration 6

2.3.3. Early guidance on Genetic investigations 6

2.3.4. Input on order of deliverables 6

3. Proposed next steps 7

3.1. Kick-off meeting 7

A. Appendix A - Tooling 8

# Introduction

The purpose of this short document is to describe the following:

* What freshEHR Clinical Informatics Ltd. propose to deliver to North Thames Genomic Medicine Centre with regard to mapping the GEL dataset to openEHR archetypes and templates.
* When and how freshEHR Clinical Informatics Ltd. propose to deliver the required artefacts and documentation.
* What freshEHR Clinical Informatics Ltd. require from North Thames Genomic Medicine Centre to enable the modelling and mapping work to proceed.

A previous document provided by freshEHR Clinical Informatics Ltd. gave an outline of the required modelling effort. The document is embedded below for reference purposes.



It is envisaged that the contents of this document will be discussed and agreed in a face-to-face meeting early in the project lifecycle.

## Partnership with Operon Ltd.

freshEHR Clinical Informatics Ltd. are partnering with Operon Ltd. (<http://operon.systems/>) in this project. Operon Ltd. will host the sandbox openEHR Clinical Data Repository environment and provide additional modelling resource.

We also intend to invoice through Operon Ltd. but this will have no impact on previous cost estimates.

# Modelling and mapping GEL data

## What will freshEHR Clinical Informatics Ltd. deliver?

### Archetypes and templates

freshEHR Clinical Informatics Ltd. undertake to provide the openEHR archetypes and templates required to model the GEL datasets for rare diseases and a small subset of cancer as described in the spreadsheet embedded below (for reference purposes).



This work will involve the use of both existing openEHR archetypes either unchanged or with small modifications (57% of total required archetypes) and the creation of new archetypes and some complex modifications (43% of the total required archetypes).

Archetypes will then be aggregated into templates, which form the deliverable for the development work outside the scope of freshEHR Clinical Informatics Ltd. involvement, other than in liaison with system developers who make use of the templates.

### Sandbox environment

freshEHR Clinical Informatics Ltd. will provide a sandbox openEHR Clinical Data Repository (CDR) environment, using the Marand Think!Ehr system, for testing purposes only. This will be hosted by Operon Ltd. This will be incrementally populated with the developed templates, and allow NHGMC and developers an early opportunity to work with the openEHR repository. A small set of sample instance data will be uploaded.

### Documentation

freshEHR Clinical Informatics Ltd. will provide a detailed dataset to archetype/template map as part of the deliverables.

### Support

freshEHR Clinical Informatics Ltd. will be available to provide support to developers during the implementation of the models.

We would strongly recommend one or two days technical training to cover specific technical aspects of openEHR such as AQL (Archetype Query Language), which will ensure that developers are familiar with these aspects.

Our cost proposal did not include any specific technical training, and we would expect this to be provided by the CDR supplier, e.g. Marand, but we are happy to provide/augment such provision if required at a cost to be agreed.

## When will we deliver?

### Overall timescales

It is envisaged that all required archetypes and templates will be delivered to North Thames Genomic Medicine Centre over a period of a maximum of 6 months. The content will be delivered incrementally with aspects ready to test and implement within a few weeks, and the overall timescale may be considerably less than 6 months, depending on the exact final scope. The precise start date is to be confirmed.

### Exploratory phase

We would like to propose an initial exploratory phase which will include a subset of archetypes and templates covering specific aspects of the dataset. We propose to cover the following:

* Clinical Genetics/ Family Pedigree
* Imaging (subset)
* Laboratory tests (subset)
* Performance measures
* Prostate cancer

This phase will provide the opportunity to test the approach and refine review and delivery processes, and it is envisaged that this phase will take one month from start date, although the Clinical Genetics/ Family Pedigree work may require some more discussion and refinement (see Section 2.3.3).

Although the focus of this work is to be Rare Diseases, we felt that inclusion of one area of the Cancer dataset would allow us to explore the overlap of the two datasets and to take advantage of parallel modelling work at UCL/UCLH around prostate cancer, as a means of exploring the challenges of aligning point-of-care data to reporting data requirements.

### Phased delivery of remaining models

Once the exploratory phase has been concluded, we propose monthly ‘drops’ to deliver the remaining models. The exact size of each ‘drop’ will vary depending on complexity of the models included in that drop, and the broad content of each drop will be confirmed during an initial kick-off meeting.

## What do we require from North Thames Genomic Medicine Centre?

### Access to domain experts / clinicians

Throughout the project lifetime, we will require access to domain experts and clinicians to assist with clarifications and reviewing the archetypes and templates. We do not propose to go through a formal review process with all the artefacts, but input will be required at various times.

The following format might be a good starting point, with the option to review during the exploratory phase:

* Email access at all times for quick clarifications.
* Weekly remote meetings for more detailed review and clarifications (Skype / GoToMeeting).
* Monthly face-to-face meetings around ‘drops’ of deliverables.

### Guidance on lab terminology and integration

There are a number of questions and issues around laboratory results, and these will need to be considered and clarified at an early stage:

* Which terminology is being used?
* Is there more than one terminology in different labs?
* Is there a view to start using SNOMED CT?
* How will integration with the laboratory systems be achieved?

### Early guidance on Genetic investigations

The GEL dataset contains sections on Genetic investigations and Family Pedigree, but there was some discussion around the completeness of those sections from a point-of-care perspective. Early clarification will be helpful. This is one area where very specific advice and input from clinicians will be required and may be more demanding in resolving differing clinical perspectives to achieve the optimum result. We can also draw on international initiatives / expertise in this area.

### Input on order of deliverables

Some of the decisions around which sections of the dataset are delivered before others will depend on the complexity of the models, and we will make some recommendations. However, we also require some guidance on prioritising some aspects over others.

# Proposed next steps

## Kick-off meeting

An early kick-off meeting would be helpful to confirm the following:

* Agree scope, deliverables and timescales as laid out in this document.
* Clarify issues around content of dataset and integration.

###### Appendix A - Tooling

A number of tools will be used during the modelling process. Please note that this section is largely for information, it is not envisaged that the authoring tools in particular will be used by North Thames Genomic Medicine Centre.

This link provides detailed information for any interested stakeholder on how to install and configure the tools:  
  
<https://github.com/freshehr/openehr_tool_docs>

XMind

This is a free-to-use mind-mapping tool which can be downloaded from this link [www.xmind.net](http://www.xmind.net).

We use XMind to map out details of the datasets and corresponding archetypes, and there may be times when these maps will be used in face-to-face and virtual meetings to agree scope and content.

openEHR Archetype Editor

The Archetype Editor facilitates authoring of openEHR clinical and administrative archetypes in ADL 1.4 and XML format. It is open-source, free to use and can be downloaded from this link <http://www.openehr.org/downloads/archetypeeditor/home>.

Please note that this tool can only be used with Windows, and it will be used by freshEHR Clinical Informatics to author/modify the required archetypes. It is not envisaged that any of the North Thames Genomic Medicine Centre domain experts / clinicians involved in the project will have to use this tool.

openEHR Template Designer

The Template Designer is required to create the openEHR templates. It is a closed source tool designed by Ocean Informatics, but it is free to use and can be downloaded from this link <http://www.openehr.org/downloads/modellingtools>.

Please note that this tool can only be used with Windows, and it will be used by freshEHR Clinical Informatics to author/modify the required templates. It is not envisaged that any of the North Thames Genomic Medicine Centre domain experts / clinicians involved in the project will have to use this tool.

GitHub

A GitHub repository will be set up to store and provide access to all the artefacts and documents used in the project.

Clinical Knowledge Manager

Clinical Knowledge Manager is an online collaborative tool to facilitate the review of archetypes and templates and act as a repository in which international and national archetypes can be viewed and downloaded for use. It is a commercial tool developed and licensed by Ocean Informatics (<https://oceaninformatics.com/>) .

We would not expect North Thames Genomic Medicine Centre to own an instance of CKM, but it is envisaged that one the UK instances (<http://www.clinicalmodels.org.uk/ckm>) will be available at times to facilitate review and discussion. This will require further discussion with the Apperta Foundation (<http://www.apperta.org/>) who are funding the UK CKM.

CKM is a very useful tool where we need to get good guidance from a broader range of clinical experts. It will not need to be heavily used for this project but might be helpful for the genomics and family pedigree work. This will have no cost implications for the project.