**Open Standards for Clinical Data at North Thames Genomic Medicine Centre —**   
**Strategic Analysis**

**June 2016 Jonathan Edwards**  
 **The IT Health Partnership**

Table of Contents

1. Executive Summary 3

2. Background: What are we seeking to achieve? 3

3. Methodology and sources 4

4. Analysis 4

4a) Open standards for clinical data — Frequently asked questions 4

Which initiatives provide open standards for representing clinical data? 4

What is the goal of openEHR and ISO 13606? 4

How are openEHR and ISO 13606 structured? 5

What are the main differences between openEHR and ISO 13606? 5

How far has openEHR developed? How is it governed? 6

Does openEHR mean open-source? 6

If openEHR is such a good idea, why hasn’t it taken off faster? 7

What are the potential gaps in openEHR? 7

What is the difference between openEHR and other standards in use in healthcare? 7

4b) What will be the benefits of using openEHR for NTGMC? 8

Data sharing 8

Data control 9

openEHR and the EPR: Enterprise-wide and “best-of-breed” approaches 9

Conclusion: Benefits of openEHR for NTGMC 10

4c) What is needed to implement openEHR for a repository, and who can provide this? 10

openEHR execution engine / platform 10

Marand 11

Ocean Informatics 11

EPR vendors working with openEHR 11

4d) Who has deployed openEHR for clinical data repositories? 12

4e) What is the best approach to implementing openEHR? 13

4f) What benefits would the GMC’s use of openEHR have for the 100,000 Genome project as a whole? 14

4g) Options appraisal 14

1. Who provides the repository? 14

2. What sort of repository is provided? 15

5. What needs to happen next? Action plan for Phase 2 15

# 1. Executive Summary

* This document assesses the rationale for using open standards for clinical data to build the data repository at North Thames Genomic Medicine Centre (NTGMC), and the optimal approach to doing this.
* The benefits of using open standards for the repository at NTGMC will be
  + Facilitating data sharing within and between organisations
  + Allowing greater control over data, by separating data from technology and applications
  + Supporting those organisations pursuing an enterprise-wide EPR
  + Supporting those organisations with a “best-of-breed” approach to EPR.
* Alternative approaches — building a data repository using a proprietary data model, or procuring a commercial data repository using a proprietary model — will not provide those benefits.
* The leading open standard for clinical data is openEHR. openEHR has matured in recent years, has been proven in several large-scale deployments and has been shown to work well with other standards. In fact, openEHR is already being used at Moorfields Eye Hospital as part of the OpenEyes system.
* Using openEHR for the central repository at NTGMC and the related repositories at the NHS Trusts that are part of NTGMC would be a low-risk initiative.
* A repository based on openEHR would generate benefits both for NTGMC and for the rest of the 100,000 Genome project, while creating a legacy that would extend beyond the end of the project.
* This would position NTGMC as an innovator and would associate the biomedical research centres of GOSH, Moorfields and UCLH with an innovative initiative.
* To proceed down this path, NTGMC will need to create a detailed implementation plan; learn from the experiences of organisations that have deployed openEHR; select a supplier of an openEHR execution engine; assemble technical and clinical resources, especially clinical modellers; and enlist support and resources from stakeholders such as the Farr Institute, Genomics England, NHS Digital and NHS England.

# 2. Background: What are we seeking to achieve?

Genomics England, a wholly owned company of the English Department of Health, is delivering the 100,000 Genomes project, one of the largest national sequencing projects in the world. The 100,000 Genome project will sequence 100,000 genomes from two categories of NHS patients: patients with cancer, and patients with rare diseases as well as their families. The project, which was procured by NHS England, is intended to enable breakthroughs in medical research, diagnosis and treatment. England has been divided into 13 regions, each of which has a Genomics Medicine Centre that is responsible for collecting gene samples and accompanying clinical (phenotypic) data, and sending this data to Genomics England. North Thames Genomic Medicine Centre (NTGMC), which covers north London, is responsible for collecting around 15,000 genomes. NTGMC needs to build a repository for the clinical data that will be collected from participating hospitals. The Genomics England data set comprises 278 data points for cancer and approximately 1,000 data points for rare diseases. NTGMC is interested in using open standards to store the clinical data. Therefore NTGMC commissioned the current report to assess the rationale, value and feasibility of using open standards, and to provide initial guidance on how to proceed down this path.

# 3. Methodology and sources

The report used a combination of desk research and interviews.

Interviews were conducted with individuals from NHS England, NHS Digital, Farr Institute, Genomics England, UCLH, North Thames GMC, Helicon Health, EuroRec, CGI, Ripple Health, openEHR Foundation, Marand, Ocean Informatics, DIPS and Cambio.

The author is Jonathan Edwards, Partner at the IT Health Partnership. The report was completed in June 2016.

# 4. Analysis

# 4a) Open standards for clinical data — Frequently asked questions

## Which initiatives provide open standards for representing clinical data?

The two global initiatives that provide open standards for representing clinical data are Open Electronic Health Record (openEHR) and the closely related ISO 13606 standard.

## What is the goal of openEHR and ISO 13606?

The goal is semantic interoperability, meaning that clinical data from one system can be transferred to another system, and the receiving system can understand it fully and accurately. The challenge with most clinical systems on the market today is that clinical data is coded in the proprietary formats of commercial vendors. This means that moving from one system to another involves a highly costly process of data migration, with the risk of data corruption and data loss. Existing standards for health information exchange such as HL7 are used primarily to send messages between systems; openEHR and ISO 13606 go a step further by attempting to standardise clinical content.

## How are openEHR and ISO 13606 structured?

The structure of openEHR is best understood using the analogy of Lego. There are four levels to understand:

1. openEHR Reference Model. This is a set of rules that define how openEHR data points are constructed. It is analogous to the rules that Lego engineers follow when constructing Lego blocks. The reference model is fixed and cannot be modified.
2. openEHR Archetypes. The archetypes are all the data points that are used in electronic health records, for example blood pressure readings, diagnoses, symptoms and laboratory values. They are analogous to Lego blocks, with their different sizes, shapes and colours. The archetypes are intended to be fixed and can be reused with minor modifications.
3. openEHR Templates. Templates are documents that are composed of various data points (archetypes). Documents can be simple, with just a few archetypes, or complex. Examples of documents include discharge summaries and referral letters. The approach is to take the maximal data set of the archetypes and constrain it to show only the data needed for a particular clinical purpose. Continuing with the Lego analogy, templates are analogous to things that are built out of Lego blocks, such as Lego houses, cars, ships etc. Templates are not fixed; any number of templates can be created as needed using various combinations of archetypes.
4. openEHR Repository. The repository is an assembly of templates. It can also include data sets, queries, terminology mappings, decision support rules etc. A repository is analogous to a city built of Lego, including houses, roads, cars etc.

## What are the main differences between openEHR and ISO 13606?

The structure of ISO 13606 is a simplified version of openEHR. ISO 13606 includes Level 1 (Reference Model) and Level 2 (Archetypes); it does not include Templates and it typically does not include a repository.

In addition to structural differences, the two standards are fundamentally different in purpose. Whereas openEHR is intended for data storage, ISO 13606 is primarily intended for data transfer (interoperability). There are very few examples of repositories based on ISO 13606.

The two standards are close cousins: many of the same individuals were involved in the creation of both standards, and archetypes are interchangeable. For example, it is possible to build a repository using primarily openEHR archetypes while including some ISO 13606 archetypes.

openEHR has several features that make it more suitable than ISO 13606 for large-scale implementations. openEHR has a sophisticated modelling tool, it has commercial execution engines (discussed in Section 3 below) that have been demonstrated to work at large scale, and it has an Archetype Query Language that enables querying of openEHR archetypes.

For these reasons, the rest of this report will concentrate on openEHR in preference to ISO 13606.

## How far has openEHR developed? How is it governed?

openEHR is coordinated by the openEHR Foundation, which was founded in 2000 and is based at University College London.

openEHR’s Archetype Definition Language and its associated Archetype Object Model are ISO standards. Other core openEHR standards include the openEHR Information Models, the Archetype Query Language, and a set of APIs including REST services generated from archetypes.

Therefore, openEHR provides three fundamental things: a standard way to represent clinical data (the Reference Model); models of clinical content and terminology (the Archetypes); and methods of access to data and models (Query Language and APIs).

Much of the work of creating archetypes and templates is devolved to national initiatives. The Clinical Knowledge Manager (CKM) is a collaborative venture intended to maintain stable versions of archetypes and to share them within and between countries. Different countries have their own CKM groups, and it is up to each group to synchronise its efforts with the international CKM. openEHR allows for specialisation of archetypes, meaning that a national group can select an international version and customise it for the needs of its country, while maintaining compatibility with the international version. The national groups are expected to share their customised archetypes with the international CKM, hence facilitating the refinement of the international versions of the archetypes.

openEHR now includes more than 500 archetypes, corresponding to approximately 7,000 clinical data points.

Inevitable challenges with collaborative initiatives such as openEHR include duplication, lack of control and variable quality. As openEHR has matured, it seems that these challenges have become less significant than in its earlier days.

## Does openEHR mean open-source?

No; although openEHR is an open standard, to implement it you need an execution engine, and most of the execution engines are closed-source (see below). There are initiatives in the UK and elsewhere to create open-source execution engines for openEHR; none of these initiatives has yet resulted in a large-scale deployment.

## If openEHR is such a good idea, why hasn’t it taken off faster?

There are a few principal reasons.

1. Many years of initial work were needed to create the reference model and the archetypes; this has now been completed.
2. In the meantime, vendors offering proprietary data models have improved their product offerings and have continued to gain market share. This includes EPR vendors such as Cerner and Epic, as well as vendors such as InterSystems, NantHealth and Orion that offer a shared health record.
3. The vendors of execution engines for openEHR are relatively small and have lacked the resources for rapid growth. However, this is now changing.

## What are the potential gaps in openEHR?

The Archetype Query Language can only query data created in openEHR format, not any other data. This is not necessarily a problem, since one could create enriched query sets that extend AQL to query data created in other formats as well as openEHR.

Although genotypic data is not included in openEHR at present, this is not a problem for the work of Genomics England, as the genotypic data is stored elsewhere.

Social care data is also not included in openEHR at present.

## What is the difference between openEHR and other standards in use in healthcare?

openEHR is the only open standard for storage of fine-grained clinical data.

openEHR can work well alongside other leading healthcare standards. These include

—HL7 messages, which are used for data exchange;

—HL7 Clinical Document Architecture (CDA), which is used for storing clinical documents but not for storing detailed clinical data;

—HL7 version 3 Reference Information Model, which is a model for storing detailed clinical data, but has not been widely adopted due to perceived difficulties with programming in it;

—Fast Healthcare Interoperability Resources (FHIR), which is an API developed by HL7 for exchanging clinical data between different systems;

—IHE standards such as XDS, which are are widely used for exchanging clinical documents;

—Systematized Nomenclature of Medicine — Clinical Terms (SNOMED CT), which is a standard vocabulary of terms relating to the care of an individual.

# 4b) What will be the benefits of using openEHR for NTGMC?

First, it is important to understand the two main benefits of using open standards such as openEHR for data storage: data sharing and data control.

## Data sharing

openEHR is the only data model that allows healthcare providers to store rich, fine-grained clinical data in an open format. Therefore, openEHR helps address several challenges that have historically made it difficult to share healthcare data between computer systems.

*Complexity.* Healthcare data is rich and complex. One example is the timing of medication doses. A clinician could specify dosage frequencies in at least the following ways

—every time period, for example “every 4 hours”

—n times per time period, for example “three times per day”

—n per time period, for example “two per day” or “six per week”

—every time period range, for example “every 4-6 hours” or “2-3 times per day”

—maximum interval, for example “not less than every 8 hours”

—maximum per time period, for example “to a maximum of 4 times per day”.

—time-specific, for example “take after breakfast and lunch”

—at specific times of day, for example “take at 6am, noon and 8pm”

—after/before events, for example “take before lying down”

—n time period before/after event, for example “take 3 days before travel”

—n time period range before/after event, for example “take on days 2-3 after symptoms begin”.

*Speed of expansion of medical knowledge.* The volume of medical knowledge is growing exponentially in two directions: increasing breadth (new information) and increasing depth (more details). Our understanding of diseases and treatments is also changing rapidly.

*Variety of ways to represent healthcare data,* including free text, structured data, images, graphs, questionnaires, and multimedia such as ECG and videos. Moreover, clinicians record data differently depending on their training, preferences and roles.

*Hierarchical nature of healthcare data.* Clinicians need multiple levels of detail depending on their roles.

*Need to be able to reuse knowledge created by clinicians elsewhere,* since medical science is by its nature a collaborative global field.

Because each of the the clinical information system vendors on the market today uses a proprietary data model, it is difficult to exchange anything more than summary data between systems. By using openEHR, healthcare providers can easily exchange rich detailed clinical data.

## Data control

In addition, openEHR addresses the problem of “lock-in”. Most commercial applications use proprietary data models, meaning that the healthcare provider’s data is locked into a format that is determined and controlled by a commercial vendor. Making changes to the data model, or writing new interfaces to extract the data, requires paying the commercial vendor. Switching to a different vendor requires spending large amounts of money on data migration and data conversion, with the inevitable difficulties of translating between two proprietary data models.

By separating data from technology and applications, openEHR allows healthcare providers to retain control of their data and to exchange that data with other healthcare providers using a globally recognised format.

## openEHR and the EPR: Enterprise-wide and “best-of-breed” approaches

Next, it is important to consider the value of openEHR alongside clinical applications such as the EPR.

openEHR is a standard for data storage. It does not replace the need for transactional applications such as the EPR. The EPR is essential to performing critical functions such as e-prescribing, order communications, electronic documentation, clinical decision support and workflow. Many NHS trusts have embarked on ambitious initiatives to implement EPRs in order to improve clinical quality, safety and efficiency.

The two main options for providing an EPR are

1. Purchase an enterprise-wide electronic patient record (EPR) application; leading vendors of such applications include Cerner, Epic, Allscripts and Meditech.
2. Maintain a multiplicity of “best-of-breed” applications and use interface engines to connect them, with a clinical data repository at the centre.

openEHR is useful for both options.

* With option 1, no EPR vendor will be able to provide all the functionality needed by a healthcare provider. The organisation will have to maintain multiple specialist applications, and this will most likely require a clinical data repository. By using an open standard format for the repository, it will be easier to aggregate data in a way that is clinically useful and to exchange data within the organisation and between organisations.
* With option 2, the healthcare provider will have to retain hundreds of applications. This makes it even more important to have a repository in which the data can be stored using an open standard format.

## Conclusion: Benefits of openEHR for NTGMC

In summary, the benefits of using openEHR for the clinical data repository at NTGMC will be

* Data sharing within and between organisations
* Data control by separating data from technology and applications
* Support for an enterprise-wide EPR
* Support for a “best-of-breed” EPR

# 4c) What is needed to implement openEHR for a repository, and who can provide this?

## openEHR execution engine / platform

To build a clinical data repository using openEHR, the principal requirement is for an openEHR “execution engine”, also known as a “platform”. This performs the following functions.

- Stores data in openEHR format in a database.

- Translates data from source systems into openEHR format. This includes interfaces for inputting and extracting data, and support for global standards including HL7 and FHIR.

- Provides an application development environment that enables applications, registries etc to be built on the database using openEHR data.

- Provides a query interface that supports sophisticated querying. openEHR contains an Archetype Query Language (AQL) that extends the capability of SQL and allows searches using hierarchical relationships. AQL was developed by Ocean Informatics and is in the process of being integrated into the openEHR specifications. Using AQL, any data item created using the openEHR reference model can be searched. Therefore, complex searches using multiple parameters can be conducted.

Although it is possible to build an openEHR execution engine, this would require a substantial investment of time, money and resources. It would make more sense to acquire one.

Two vendors, Marand and Ocean Informatics, have an openEHR execution engine. In addition there are two EPR vendors, DIPS and Cambio, that have rewritten part or all of their applications onto openEHR.

## Marand

Marand is based in Ljubljana, Slovenia. The company has annual revenue of €18m and was founded in 1991. Marand has 120 staff, including 80 software developers.

Marand’s openEHR execution engine is called Think!EHR.

In addition to Think!EHR, Marand has the following:

- A clinical information system (Think!Clinical) and a medication management system (Think!Meds); Think!Clinical and Think!Meds both use AQL to retrieve the necessary clinical data from Think!EHR.

- ThinkEHR Explorer, a suite of tools for application development and for defining care coordination protocols.

- Think!EHR business process modelling layer, for definition and deployment of clinical pathways and order sets.

According to Marand, Think!EHR has been tested and certified on 20 million patients, 1 billion documents, and 25 terabytes of data. The average query response time was less than 1 second. Marand claims a maximum of 39,942 AQL transactions per second, and 99,999% uptime in its Slovenian deployments.

Marand supports the main commercial databases including Oracle, IBM DB2 and Microsoft SQL Server.

Marand is often implemented alongside IHE XDS components for document sharing.

See section 4 for descriptions of Marand’s deployments.

## Ocean Informatics

Ocean Informatics is an Australian company that provides software and services to support openEHR. It has deployed openEHR for shared health records, infection control, primary-care, and integration between primary and acute-care.

See below for descriptions of Ocean Informatics’s deployments.

## EPR vendors working with openEHR

DIPS is a Norwegian EPR vendor that adopted the Marand execution engine and has been rewriting its EPR application into a new EPR called Arena that is entirely based on openEHR. DIPS is used by three of the four health regions of Norway.

Cambio is a Swedish EPR vendor that has rewritten part of its application (clinical decision support) using openEHR. In fact, Cambio has incorporated its new tools for clinical decision support into a new domain of openEHR called Guideline Definition Language (GDL). Cambio is used by more than half of the 20 county councils of Sweden.

# 4d) Who has deployed openEHR for clinical data repositories?

The largest deployments of openEHR have been in the city of Moscow and in Slovenia.

Moscow EHR — “Integrated Medical Information System”

This deployment covers 12 million citizens with 161 million patient visits per year. It has 130,000 users including 45,000 physicians. It has been live since 2012. The Marand execution engine was combined with a user interface developed by a Russian firm called Infinnity Solutions and with consulting support from Ocean Informatics.

Slovenia — national EHR

This deployment covers more than 2 million citizens, 13 hospitals, 441 GPs and 159 specialist clinics. The repository includes 1.5 million documents in the repository. The deployment was completed within 6 months and has been live since 2013.

The Marand Think!EHR is used for exchange of clinical documents (discharge letters, outpatient encounter notes, laboratory results etc); vaccination registry; patient summary; and e-prescribing.

Other notable deployments of openEHR include the following.

Australia — Queensland infection control

The state of Queensland is using Ocean Informatics’s MultiPrac product running in more than 140 hospitals, under contract to the Queensland Health Centre for Healthcare Related Infection Surveillance and Prevention (QH CHRISP). MultiPrac is used for surveillance of hospital-acquired infections.

Australia — Northern Territory Department of Health EHR

Since 2012, Northern Territory has had a shared EHR based on openEHR, serving more than 50,000 remote indigenous residents of the Northern Territory and neighbouring jurisdictions of Australia.

Brazil

In 2012, the SPAsaude health plan deployed an ambulatory EHR system for chronic care management.

Netherlands

The Dutch company Code24 has a few implementations using openEHR for medication management, electronic referrals etc.

Portugal — IdealMed

IdealMed is a private hospital in Coimbra that deployed a repository and a clinical trials management system using openEHR in 2011.

UK — Leeds Care Record

The Leeds Care Record is an open-source shared health record using openEHR and other standards that allows hospital clinicians to see GP records and GPs to see records from other GP practices. Initial deployment occurred in 2013.

UK — NHS Digital

NHS Digital (formerly the NHS Health and Social Care Information Centre) has an initiative to create messages based on openEHR archetypes for transfers of care.

UK — Moorfields Eye Hospital

Moorfields Eye Hospital is using openEHR for its OpenEyes electronic patient record system, which is based on an open-source implementation of openEHR.

# 4e) What is the best approach to implementing openEHR?

Although it is beyond the scope of this report to outline a detailed implementation plan, some important priorities have emerged from the research conducted thus far.

A critical point that differentiates a successful implementation of openEHR is clinical modelling and mapping. The vendor of the execution engine must deploy competent clinical modellers to work with clinicians at the customer site and help them understand how they will use the data. While it is essential to have close involvement of clinicians in building and adapting archetypes and templates, the clinicians need to be guided by technical experts with a strong understanding of computer programming so as to ensure that whatever the clinicians wish to develop is technically feasible.

In addition to the openEHR repository, it will be essential to have a meta-data registry that maps all the systems that are contributing data to the repository. This is especially important with large academic medical centres that may have hundreds of specialised systems.

It will be important to determine how to extract data from the source systems into the openEHR repository. This will require firstly, exposing APIs that the source systems will use to transfer the data, either in bulk on a regular basis or for each record separately. The next step will be to create an openEHR template which will define the schema for the imported records, including identifying the required archetypes. Finally, it will be necessary to define a mapping from source to target, to identify how the legacy data fields are translated to the fields in the template.

# 4f) What benefits would the GMC’s use of openEHR have for the 100,000 Genome project as a whole?

The previous sections discuss several benefits of using openEHR for clinical data storage, including semantic interoperability, minimising vendor lock-in, and helping organisations cope with the rapidly expanding volume and complexity of medical knowledge.

Another benefit will be to create a valuable legacy extending beyond the end of the 100,000 Genome project. The hospitals that participate in NTGMC, and other hospitals that choose to use instances of the openEHR repository, would gain immediate benefit by having a clinical data repository that they could use for data storage, with their data held in an open format, at minimal cost.

In addition, a successful implementation of openEHR at NTGMC could be replicated by the 12 other GMCs of England and by the new GMCs of Wales and Scotland.

The GENIE sample tracking system developed by University Hospital Birmingham has not yet built a clinical data repository for the phenotypic data that GENIE is collecting. GENIE could use openEHR for its repository. By doing so, GENIE would gain a repository that could also be used for other purposes.

# 4g) Options appraisal

NTGMC needs to provide a clinical data repository to each participating hospital to store the clinical phenotypic data to accompany the genome samples. Up to this point in the report, we have been assuming that NTGMC would provide the data repository itself. However there are actually several options to consider, using two variables: who provides the repository, and what sort of repository is provided.

## 1. Who provides the repository?

1a) NTGMC provides a repository, with multiple instances each hosted at the participating NHS trusts. This seems like the most promising option at present.

1b) Each NHS trust provides its own repository. This does not seem a sensible option.

1c) University Hospital Birmingham (UHB) provides a repository for NTGMC as well as for the other GMCs that use the GENIE system. This may be an option, depending on whether UHB is interested in following this path.

## 2. What sort of repository is provided?

2a) Build an openEHR repository using an openEHR execution engine such as Marand or Ocean Informatics. This is the recommended option.

2b) Build a repository using a proprietary data model. This is not recommended, for reasons discussed above.

2c) Procure a commercial repository that has a proprietary data model. This is not recommended, for reasons discussed above.

Therefore options 1a) plus 2a), or 1c) plus 2a), appear to be the best combination.

|  | **Build or procure an openEHR repository using an execution engine** | **Build a repository using a proprietary data model** | **Procure a commercial repository with a proprietary data model** |
| --- | --- | --- | --- |
| **NTGMC provides repository with one instance per trust** | Yes | No | No |
| **Each trust provides its own repository** | No | No | No |
| **UHB provides a repository for NTGMC trusts and for other GMCs that use GENIE** | Yes | No | No |

# 5. What needs to happen next? Action plan for Phase 2

Phase 1 of this initiative involved the research and writing of this strategic analysis.

Phase 2 will involve the following activities.

* Refine the strategic analysis with input from key stakeholders including NTGMC, NHS England, NHS Digital, GENIE, GOSH, Genomics England, Farr Institute, UCLH, and the vendor community.
* Create a detailed implementation plan for openEHR at NTGMC. We need to understand the level of effort required to build the repository and map the data sets into openEHR, the likely time and cost involved, and the overall feasibility of doing this.
* Interview leaders at end-user organisations that have implemented openEHR, including Moscow, Slovenia and Northern Territory, and learn from their experiences.
* Conduct a detailed evaluation of vendors of openEHR execution engine, and select one.
* Assemble technical and clinical consulting resources needed, especially clinical modellers.
* Obtain input from Farr Institute and GeCIPs to shape the narrative around openEHR, to promote NTGMC’s approach to other influencers, and to assemble resources — for example, for the work necessary to map the Genomics England data set to openEHR.
* Obtain input from the standards team at NHS England.
* Obtain input from Genomics England, and assemble resources for mapping from the GEL dataset to openEHR.