A Standards-based Approach to Development of Clinical Registries - Initial Lessons Learnt from the Gestational Diabetes Registry

Conference Paper · November 2014

DOI: 10.13140/2.1.2861.7283

CITATIONS
7

READS
1,544

4 authors, including:

Koray Atalag
University of Auckland
41 PUBLICATIONS **336** CITATIONS

SEE PROFILE

Aleksandar Zivaljevic
Scion
7 PUBLICATIONS **22** CITATIONS

SEE PROFILE

Carl Leonard Eagleton
Counties Manukau District Health Board
13 PUBLICATIONS **398** CITATIONS

SEE PROFILE

A Standards-based Approach to Development of Clinical Registries - Initial Lessons Learnt from the Gestational Diabetes Registry

Koray Atalag
National Institute for Health Innovation, the University of Auckland
Private Bag 92019, Auckland 1142, New Zealand
k.atalag@auckland.ac.nz

Aleksandar Zivaljevic
The University of Auckland
Private Bag 92019, Auckland 1142, New Zealand
aziv835@aucklanduni.ac.nz

Carl Eagleton
Counties Manukau District Health Board
100 Hospital Road, Papatoetoe 202
Carl.Eagleton@middlemore.co.nz

Karen Pickering
Diabetes Projects Trust
PO Box 61144, Otara, 2159
manager@dpt.org.nz

Abstract

Gestational diabetes has implications for both mother and child with risk of complications during pregnancy, and type 2 diabetes later in life. This paper presents the initial lessons learned from the development of a clinical registry. The aims of the Registry are: 1) 100% successful diabetes screening within 3 months of delivery; 2) Annual type 2 diabetes screening; 3) Early warning in subsequent pregnancies.

We have employed the openEHR standard which underpins our national interoperability reference architecture to represent the dataset and also to build the web-based registry system. Use of this rigorous methodology to tackle health information is expected to ensure semantic consistency of Registry data and maximise interoperability with other Sector projects. The development work has been facilitated by the ability to transform the dataset automatically into software code – ensuring clinical requirements accurately translated into technical terms.

Dataset has been finalised, registry system has been developed and deployed for pilot implementation. Data entry is underway for participants after consenting.

This registry is expected to increase the screening of women leading to earlier detection of diabetes. It should provide a valuable picture of the condition and is intended for extension and wider roll-out after evaluation.

1. Introduction

Gestational Diabetes Mellitus (GDM) and Type 2 Diabetes (T2DM) are related major population health problems. They are driven ultimately by the complex interaction of genetic, behavioural and environmental factors and have an impact on large numbers of people. They tend to affect certain population groups more than others, leading to socioeconomic and ethnic inequalities in health outcomes. GDM is characterised by glucose intolerance with onset or first recognition during pregnancy and is identified through an oral glucose tolerance test (OGTT) during pregnancy (1). A repeat OGTT performed 6 weeks post-partum checks for resolution. If normal, an annual fasting glucose or glycosylated haemoglobin (HbA1c) screening test is recommended for T2DM, according to current New Zealand guidelines.

Only limited data are available on time trends in GDM in New Zealand, however, absolute numbers and prevalence rates appear to have been increasing in recent decades: the proportion of women with GDM diagnosed at National Women's Hospital increased from 1.4% of births in 1991 to >7% in 2010 (2). GDM has long-term, serious consequences. A study in NZ found 19% of 110 women with GDM developed T2DM after a mean follow up of only 2.4 years (3). Women with a history of GDM have an increased prevalence of cardiovascular disease (CVD) risk

factors such as hypertension, dyslipidaemia, and microalbuminuria, and of developing CVD (4). Furthermore these women are highly likely to develop GDM in a subsequent pregnancy. Finally, children born to women with GDM have been found to have increased rates of obesity and hypertension as adolescents (5).

Better utilisation of health information is a necessity for delivering on the pressing requirements of modern clinical practice to provide the best available medical care for individuals, yet equitable and sustainable for the society over time. While the ultimate aim is to have the longitudinal and lifelong electronic health record (EHR) that is accessible whenever and wherever needed, because it doesn't exist, clinical registries are established to collect information about individuals in areas where improvement in practice is of high importance. These capture observations, diagnoses, procedures, clinical processes and most importantly outcomes. For example a clinical registry may include patients treated with a particular drug, device or surgical procedure (e.g. joint replacement), with a particular illness (diabetes), and utilising a specific healthcare resource (e.g. treated in ICU) (6).

The risk of developing T2DM can be substantially reduced by early identification in people at high risk (in particular those with impaired glucose tolerance) and then through targeted lifestyle and pharmacological interventions. These targeted interventions can also lead to better treatment (5,7). Sending out reminders has been shown to improve adherence and leads to better compliance with screening recommendations. The use of an electronic registry to automatically generate and send physician and patient reminders has been shown to be effective (8). In addition the Registry can be used to drive quality improvement and to enhance patient safety by identifying sub-optimum care and variations in processes and clinical outcomes. This can be reported at individual, provider organisation and regional/national levels (9).

Information is a precious asset in healthcare and deciding on what data to capture is not a trivial task. From complex clinical information systems with hundreds of data items to simple databases with only a few the same difficulties apply, albeit with varying levels, for the handling of data. The success of a registry depends on how well and to what extent collected data are able to answer the original questions. Therefore it is critical to have the right questions from the outset on which the definition of the final dataset will depend. The dataset definition process is usually the one most registry projects struggle and almost always one with compromise: the challenge is to find the right balance between ambition and what data are relevant and feasible to collect. Therefore it is arguably never perfect. It takes good clinical communication among other things like the effective use of collaborative methods and tools. It is not uncommon that some or all of dataset items may have already been defined previously by another project or by a national standard – so reinvention should be avoided.

A latent and often neglected issue in any health information system project is how to tackle the changeability of health information. For example a clinical system collecting imaging type of data may need to cater for new modalities at some stage, or a new clinical scoring system might come into practice. The ease with which these changes could be applied to a particular system and backward data compatibility should be considered firmly at the outset.

Currently some GDM related information is routinely collected by GPs, DHBs and community providers. However, not only are there different systems and multiple versions of the same systems being used at the moment but the content, completeness, data quality are known to be poor. Therefore it was imperative to establish an electronic registry that would be the single source of truth - adding significantly to our research capability by making much needed quality information available.

We are suggesting a standards-based approach in developing a clinical registry, starting from dataset preparation to the development of the software product including managing the evolution of both data and software over time. We have employed the openEHR formalism (10), an international and also national health informatics standard for representing health information and building EHR systems, from dataset development to software development and tackling changeability of clinical information (11,12). This will ensure the dataset is consistent and help align with other Sector projects in future. The Gestational Diabetes Registry is a joint effort among the National Institute for Health Innovation (NIHI) of the University of Auckland, Counties Manukau DHB and Diabetes Projects Trust (DPT). Funding was made available through CMDHB for a pilot study to undertake health informatics research and the technical development. DPT is supporting the clinical research and implementation of the Registry. Within the scope of the pilot study the Registry will send out six week OGTT, three monthly and annual HbA1c reminders to women and prompt their providers.

Overall aim of the study is to establish and evaluate a GDM registry and to investigate its ability to improve the preventive care women and their families have available to them. From a health informatics perspective, the development of a clinical registry traverses a few foundational subject areas in health informatics, including data/information representation, databases, health information systems, electronic health records, and healthcare terminology and interoperability standards. We are investigating the feasibility of establishing a GDM registry in general and in particular to what extent the dataset requirements of the Registry can be represented by existing standard clinical concept representations, or openEHR Archetypes, in Australian/New Zealand and international settings. In this paper we are reporting our preliminary results on dataset and software development.

2. Methods

This is an end-to-end implementation of openEHR specifications - from modelling of the Registry dataset to the development of the front-end web-based application and the back-end clinical data repository. This international standard underpins our national Interoperability Reference Architecture (13) and is the depicted standard for establishing a reference library of common clinical concept definitions (14). At the heart of openEHR lies the Archetype paradigm which is a way of constructing information models by means of putting together and configuring building blocks defined in a common reference model (RM). Thus the methodology is commonly known as Two-Level or Multi-Level Modelling (MLM) which was adopted by ISO and CEN (12).

RM consists of a small set of object oriented classes which depict generic characteristics of health records and necessary information items to meet ethical, medico-legal and provenance requirements. These carefully engineered classes can faithfully capture results of all medical entries; in other words it is guaranteed to store any type of current or future healthcare data in a computable way. RM also includes ability for versioning of medical transactions and managing change-sets to handle input errors, simultaneous multi-user read/write access to the record, and provides necessary information 'hooks' for ensuring privacy, confidentiality and security.

Archetypes bring together and constrain RM classes and form computer processable models. It is possible to define labels, data structures, data types, prescribed value ranges and value sets for data items. In addition Archetypes can be linked to external terminology/ontology systems and bibliographic databases. This ensures complete and non-ambiguous expression of domain knowledge by expert clinicians which is ready for consumption in the technical development environment. openEHR further provides the ability to tailor existing models to meet specific needs without breaking original semantics by a method called Archetype Specialization. Here, extensions and modifications that would result in more strict constraints are allowed that would continue to make data backward compatible.

A third component of the methodology is the openEHR Template, which assembles archetypes into larger structures like a screen form, document, report or message, and further constrain them for local use. They may add further local constraints on archetypes, including removing or mandating optional sections, and may define default values.

While RM is of concern to IT professionals, Archetypes and Templates are targeted to clinicians. By using this approach software development and evolution can proceed independently from volatile clinical information requirements which can bring significant cost and time savings (15).

2.1. Dataset development

Initial dataset requirements were captured by an Excel spreadsheet by the clinical team but we transformed it into a mindmap for ease of understanding and communication (Figure 1).

A web-based collaborative clinical modelling and publishing tool, the Clinical Knowledge Manager (CKM), was used to facilitate dataset development that natively supports openEHR (16). This allowed for rapidly refining the dataset requirements by iterative review mechanism – similar to peer-reviewed scientific journal editorial process (Figure 2).

Next stage was to create a flat CLUSTER type Archetype faithfully representing the mindmap including appropriate data types, enumerated values and cardinality constraints, default values and basic validation rules. The purpose was to feed this Archetype into the CKM for review. We should emphasise that this is a 'hack' into mainstream openEHR modelling where the resulting model is expressed using purpose-specific CLINICAL ENTRY type archetypes which come with a lot of contextual aspects. By introducing the flat model we were able to get started with the review quickly and hence deferred dealing with complexity of the formalism to a later stage when a reasonable level of maturity of the dataset content is reached.

We conducted two CKM reviews to finalise the dataset. First the clinical team members were asked to register to the CKM and then they were invited to the review using the editorial functions of CKM. They were given a timeframe of two weeks to complete the review. A total of four clinicians have provided feedback to individual data elements and also responded on specific questions raised. All in all the dataset development took a total of 8 weeks.

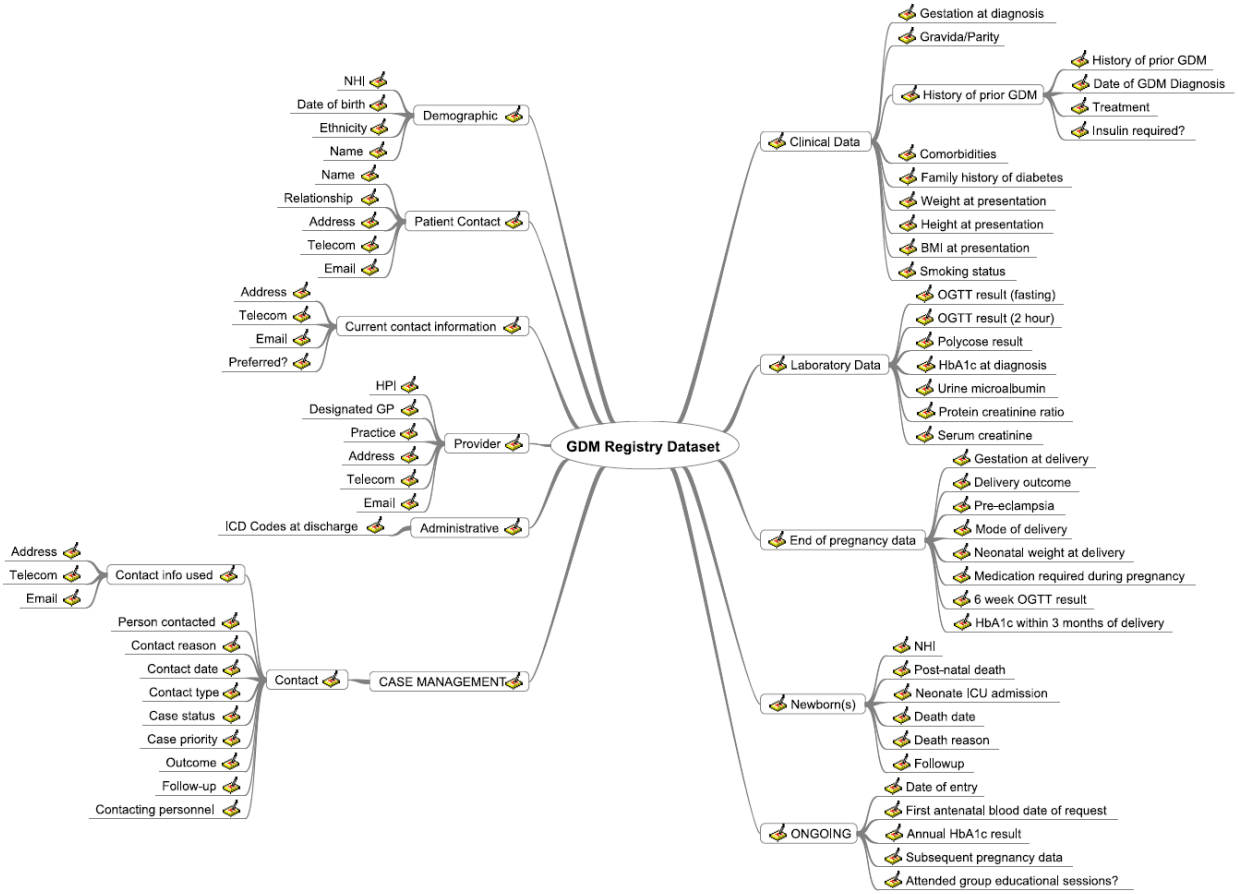

Figure 1. Mindmap representation of the GDM Registry dataset

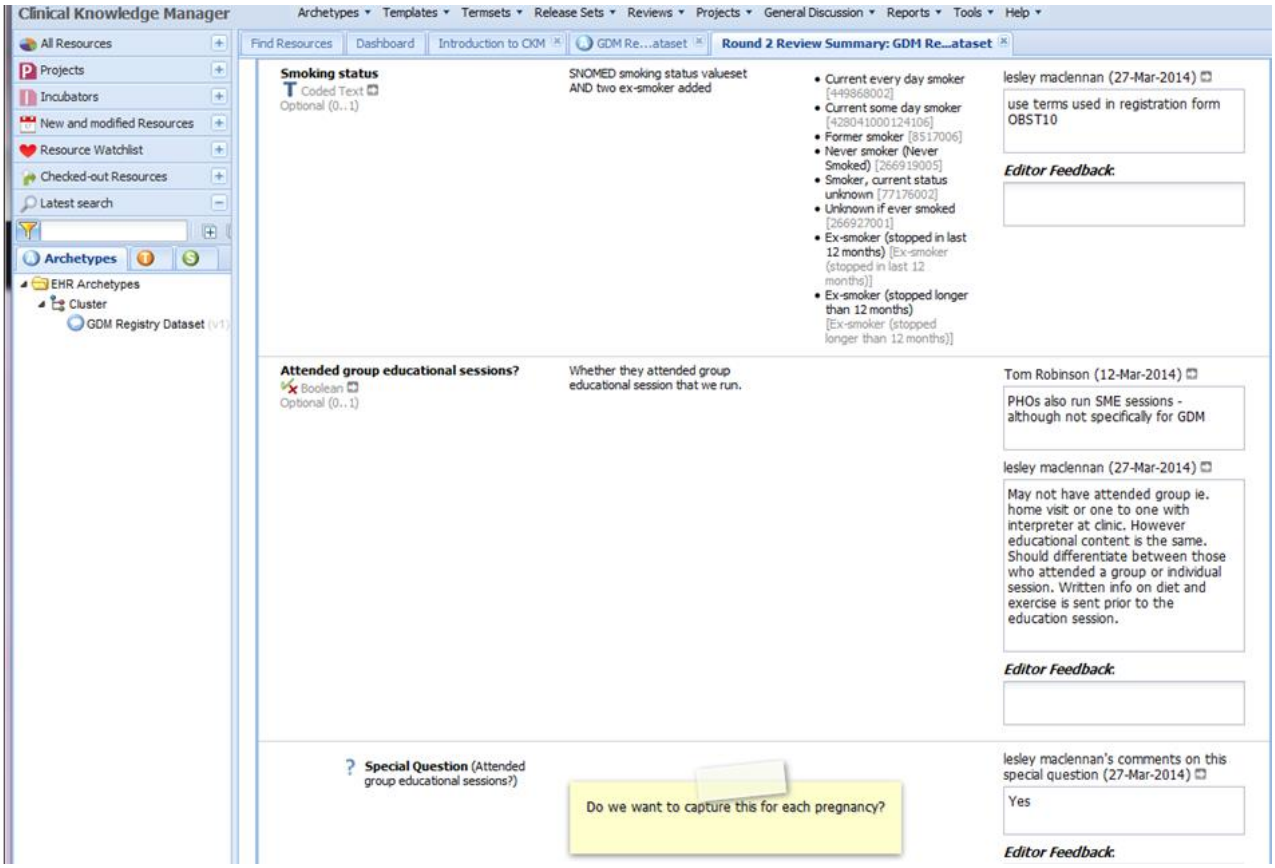

Figure 2. CKM editorial process during dataset development

Next step was proper modelling of the dataset by reusing existing archetypes and other clinical resources as much as possible. Information model reuse is considered time and cost saving operation and is not specific to the openEHR only. Literature offers examples of reuse of ontological models (17), HL7 models (18) and many others (19).

Information models surveyed for reuse were Archetypes created by other authors as part of their software development initiatives or publicly available generic models, including the models from the openEHR foundation (20) and NEHTA DCM (21) repository. One new Archetype was created to model information not covered by existing archetypes using the Ocean Archetype Editor (22). Chosen Archetypes were then assembled under a parent COMPOSITION archetype and further constrained as an openEHR Template using Ocean Informatics' Ocean Template Designer (TD) which was the last step in modelling.

2.2. Development of the Registry System

openEHR Templates drive software development by providing computable programming artefacts into software development environment. TD interface offers drag-and-drop modelling functionality and Visual Studio type interface for manipulating models' properties. The models can be exported to several different forms, including XML Schema, HTML, CSV and software code in the form of a C# class called Template Data Object (TDO). We have used the TDO (a.k.a. domain model) and included in our software project. Export process allows for easy namespace setting and the resulting artefact is a .cs file containing the class which can directly be included in Microsoft Visual Studio 2010 development environment. The Registry System is based on the MVC 2.0 design pattern, running on the Dot.Net framework version 3.5 with SQL Server 2008 R2 database. We utilised the OceanEHR framework (through an academic license) which provides application programming interfaces (API) to complex openEHR data operations and also ensures that data conforms to openEHR specifications. We have also used and extended a portal front-end template, from Ocean Informatics, to build the GDM Registry secure web application which was made available under Apache 2 open source license.

The Composition representing the whole dataset is expressed within the TDO. As the object instantiated from the class is too complex to be used for data entry and information presentation purposes directly, a multitude of smaller objects were created to satisfy user interface requirements (hence the smaller objects we call *view* objects further as per the MVC paradigm). Mapping functions are used for populating view objects with data retrieved in the composition. Accordingly, before the data is persisted, the Composition is populated with values contained in view objects.

As much as the model informed software development, the process of software development also had a significant impact on the final model. For example we started out with two templates: one to capture persistent information and the other to capture episodic information from pregnancies. Later it became apparent that a single template was more appropriate from the development point of view and that capturing persistent information for each pregnancy was actually crucial to be able to show the progression over time. For example smoking status, family history and gravida/parity type information can potentially change over time independent from pregnancy history and we initially intended to capture them separately. However since the Registry will not be used as a point-of-care system where critical clinical decisions are to be made we preferred to persist these as part of the last recorded pregnancy.

Another point that became apparent during software development was the practicality of using a single generic archetype for representing all different laboratory tests as opposed to using dedicated laboratory test archetypes for each result type. For this reason we have used the NEHTA Pathology Test Result archetype to cater for oral glucose tolerance test (OGTT), polycose, HbA1c, urine microalbumin, protein creatinine ratio, serum creatinine and pregnancy test results. The constraints applied programmatically at the runtime for laboratory test results include change of name and units properties as well as constraining temporal events to match time series patterns such as baseline or two hours measurements for OGTT.

The Registry will receive updates (on a monthly basis) about demographic information (e.g. address change) or any new laboratory results. Intervention actions will be triggered by incoming data and TXT reminders will be sent via a commercial gateway system. We have used the Archetype Query Language (AQL) to interrogate collected data.

3. Results

We have finalised the Registry dataset and deployed the Registry. After conducting a privacy impact assessment and the Institutional Review Board approval eligible women are currently being consented and entered into the Registry. At this stage regional ethics approval is not needed as the Registry is perceived as part of care delivery within the DHB.

The Registry System consists of four main parts: 1) Demographics; 2) Clinical view and data entry (Figure 3); 3) Interventions; and 4) Administration. Access is role based (clinical and/or admin) and records can be marked as: a) Temporary: record still being populated (import/data entry), not included in actions; b) Active: records visible to all and complete; and c) Inactive: only admin can see, suspended/opt-out.

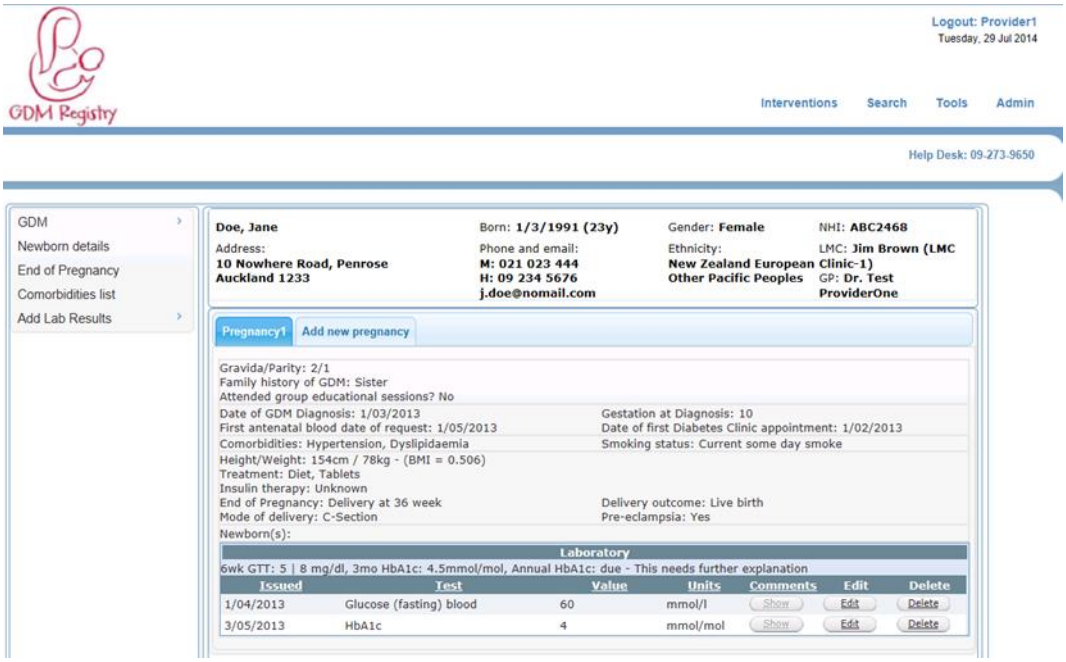

Figure 3. Clinical view form of the GDM Registry System

4. Discussion

The rapid development and deployment of the Registry lends itself to strong clinical championship and effective governance as well as the technical benefits of using openEHR. Dataset was defined and finalised in a matter of weeks by the help of CKM reviews and the ability to change software by means of modelling due to dataset alterations proved to be advantageous. We believe using a fully-fledged EHR methodology to underpin the Registry will bring about many benefits, including better expressivity in data representation, meta-data support, interoperability, semantic querying (easy and very fast), standards compliance and leveraging existing tooling and international content - hence future-proofing registry (23).

An often overlooked critical factor is to ensure that the registry users are willing to enter data – structured and in good quality. Increasingly health information is collected by different systems, more and more as part of the care delivery process, and healthcare professionals and patients don't want to enter the same data more than once. Therefore it is imperative to look at integrating systems with such systems before attempting to deploy a solution. Current limitation of the Registry is that we have not undertaken any integration work during the pilot phase. We will look at implementing API based interfaces with external systems such as the NHI, TestSafe and Maternity Clinical Information System.

NIHI's existing technical infrastructure and data management capabilities will ensure Registry data will be handled securely ensuring that patient privacy is protected. Indeed, as an independent not-for-profit organisation, NIHI is increasingly recognised as an appropriate institution to host and provide stewardship to clinical datasets, such as the Growing Up in New Zealand, SPARX and the New Zealand Cardiac Registry (ANZACS-QI). Because the openEHR approach is clinical domain agnostic, we can leverage this platform for rapidly and cost-effectively building other registries.

5. Conclusion

This registry is expected to increase the screening of women leading to earlier detection of diabetes. It should provide a valuable picture of GDM in CMDHB and is intended for extension and wider roll-out after evaluation. The standards based approach to dataset and software development will ensure consistency and comparability as well as enable data integration without any loss or distortion of semantics.

6. Acknowledgments

We would like to acknowledge the members of the Gestational Diabetes Registry Development Group, including Tom Robinson, Chris Bullen, John Baker, Brandon Orr Walker, Shekhar Sehgal, Timothy Kenealy, Leslie MacLennan, Lesley McCowan, Heidi Baxter and Caran Barratt-Boyes, and our Maori and Pacific Health colleagues. The study was funded by Counties Manukau DHB and Diabetes Projects Trust. We are thankful to Ocean Informatics Pty Ltd. for providing the OceanEHR framework and template, and for training and technical support.

7. References

1. Metzger BE, Gabbe SG, Persson B, Lowe LP, Dyer AR, Oats JJN, et al. International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy Response to Weinert. Diabetes Care. 2010 Jan 7;33(7):e98–e98.

2. Auckland District Health Board. National Women's annual clinical report. Auckland; 2010.

3. McGrath NM, Evans C, Holgate A. Post-partum follow-up of women with gestational diabetes mellitus from Northland, New Zealand. Diabet Med J Br Diabet Assoc. 2007 Feb;24(2):218–9.

4. Shah BR, Retnakaran R, Booth GL. Increased Risk of Cardiovascular Disease in Young Women Following Gestational Diabetes Mellitus. Diabetes Care. 2008 Aug;31(8):1668–9.

5. Graziano Di Cianni AG. Gestational diabetes mellitus: an opportunity to prevent type 2 diabetes and cardiovascular disease in young women. Women39s Health Lond Engl. 2010;6(1):97–105.

6. Gliklich RE, Dreyer NA, Leavy MB, editors. Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd ed. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 [cited 2014 Aug 12]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK208616/

7. Baker MK, Simpson K, Lloyd B, Bauman AE, Singh MAF. Behavioral strategies in diabetes prevention programs: a systematic review of randomized controlled trials. Diabetes Res Clin Pract. 2011 Jan;91(1):1–12.

8. Brown CH, Berndt D, Brinales JM, Zong X, Bhagwat D. Evaluating the evidence of effectiveness for preventive interventions: using a registry system to influence policy through science. Addict Behav. 2000 Dec;25(6):955–64.

9. Molina-Ortiz EI, Vega AC, Calman NS. Patient Registries in Primary Care: Essential Element for Quality Improvement. Mt Sinai J Med J Transl Pers Med. 2012;79(4):475–80.

10. Beale T. Archetypes: Constraint-based domain models for future-proof information systems. Eleventh OOPSLA Workshop on Behavioral Semantics: Serving the Customer. Seattle, Washington, USA: Northeastern University; 2002. p. 16–32.

11. Kalra D, Beale T, Heard S. The openEHR Foundation. Stud Health Technol Inform. 2005;115:153–73.

12. ISO Technical Committee 215: Health informatics, Working Group 1: Health Records and Modelling Coordination. ISO 13606 - Health informatics-Electronic health record communication. ISO; Report No.: ISO 13606.

13. Hay D, Maitre AL, Kenworthy A, Atalag K. New Zealand Interoperability Reference Architecture [Internet]. Health Information Standards Organisation (HISO); 2012. Report No.: 10040. Available from: http://www.ithealthboard.health.nz/sites/all/files/Interoperability%20Reference%20Architecture%20v%201.0.pdf

14. Health Information Exchange Architecture Building Blocks [Internet]. Health Information Standards Organisation (HISO); 2012 [cited 2012 Jul 30]. Available from: http://www.ithealthboard.health.nz/content/health-information-exchange-architecture-building-blocks

15. Atalag K, Yang HY, Tempero E, Warren JR. Evaluation of software maintainability with openEHR–a comparison of architectures. Int J Med Inf [Internet]. [cited 2014 Aug 13];0(0). Available from: http://www.ijmijournal.com/article/S1386505614001415/abstract

16. openEHR Clinical Knowledge Manager [Internet]. [cited 2012 Jul 30]. Available from: http://www.openehr.org/knowledge/

17. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing? Int J Hum-Comput Stud. 1995 Nov;43(5–6):907–28.

18. Fieschi M, Coiera E, Li Y-CJ. Medinfo. IOS Press; 2004. 810 p.

19. Goossen W, Goossen-Baremans A, van der Zel M. Detailed Clinical Models: A Review. Healthc Inform Res. 2010;16(4):201.

20. Kalra D, Beale T, Heard S. The openEHR foundation. Stud Health Technol Inform. 2005;115:153–73.

21. NEHTA Clinical Knowledge Manager [Internet]. [cited 2012 Jul 30]. Available from: http://dcm.nehta.org.au/ckm/

22. Ocean Informatics. Ocean Archetype Editor. 2006; Available from: http://www.oceaninformatics.biz/archetype_editor/ArchetypeEditor_download.html

23. Atalag K. Using a single content model for eHealth interoperability and secondary use. Stud Health Technol Inform. 2013;193:282–96.