pHealth 2020
B. Blobel et al. (Eds.)

doi:10.3233/SHTI200625

Semantic Cross-Mapping Execution of Data in the Perinatal Registry of the Netherlands

Linda VIS [b], Sanne KOOLE [b], Anneke GOOSSEN [a], Harrie HUISMAN [c], William GOOSSEN [a 1]
[a] *Results4Care BV, Amersfoort, the Netherlands*
[b] *Perined, Utrecht, the Netherlands*
[c] *Zorg & Welzijn Connect, Amsterdam, the Netherlands*

Abstract. Introduction: The electronically submitted data from midwives and hospitals to the Netherlands perinatal registry vary significantly in their data definitions, and electronic message versions. The purpose of this article is to describe the semantic cross-mapping tool and execution procedure to prepare the data for statistical analysis. Methods: requirements analysis, design, development and testing. Results: The tool for governance of versions of datasets, CIMs, data, and value sets is designed, developed, and tested. The test is based on the data-mart of version PRN 1.3 based data from 2019. Data are semantically cross mapped to current version perinatology data 2.2. Conclusion: The cross-mapping of PRN 1.3 data to perinatology 2.2 data are defined in the tool, testing revealed this mapping is successful.

Keywords. Perinatology, Clinical Data Warehouse, Semantics, Cross-mapping

Introduction

Since the establishment of the perinatal registration in the Netherlands in 1971, the processes for participation, content, format and data management have undergone significant changes [1,2]. Currently, midwives, general practitioners, obstetricians and pediatricians submit data about pregnancy, delivery, birth, maternity and neonatology to Perined, the national registry. All current data are based on the specifications of the national perinatal data set [2]. The intention is submitting data using the standard Health Level 7 version 3 Care Record (HL7v3CR) messaging, following the specifications of the national perinatal dataset [3,4]. Historically there were different specifications and submitting formats [2,4] and due to required revisions in the current national perinatal dataset there will be updated versions. Perined processes the different data versions manually into one semantic equivalent version and the diversity in datasets increased the time consummation and the complexity of the data processing. Automation and transparency with semantic cross-mapping are necessary for quick feedback to the health professionals [2,5].

[1] Corresponding Author. William Goossen, PhD, Results 4 Care BV, De Stinse 15, 3823 VM Amersfoort, The Netherlands; E-mail: wgoossen@results4care.nl.

The goal of this paper is to describe the semantic cross-mapping of data in one of the data-marts of Perined. This cross-mapping aims supporting Perined to handle the large variation of incoming data and therewith increase their data analytics capability.

1. Methodology

This project applied several methods to achieve cross-mapping of data in the perinatal registry of the Netherlands: requirements analysis for cross-mapping using Clinical Data Warehouse (CDWH), design, development and testing. The data send to Perined come from varied Electronic Health Record systems (EHRs) and are stored in an optimized format for complex queries, data reuse and analysis [6,7,]. The current Perined Data Warehouse (PWDH) is based on various CDWH principles [6,7,8,9,10,11,12]:

- attribute-value pairs and Detailed Clinical Models (DCM) for data storage, via linkage of clinical concepts to standard terminologies defining semantics and specifying data elements;
- a polyvalent database including structured data, various contexts, meta-information, standardized terminologies, and various queries;

The data submitted to Perined have different data sets and various submission formats [3,6,14]. Hence, the need for additional Extract Transfer and Load (ETL) procedures to semantically equivalent data.

During the finalization of PDWH several routes were explored. The cross-mapping, using semantification principles, that is to transform all varied source data into the actual national perinatal target data, proves the most adequate. The data are extracted from the raw data-marts, cross-mapped from source data to target data per data element and, if used, per coded value, and finally placed in a fully semantically equal datamart. For this a sequence diagram was developed, illustrating the various roles for researcher, database manager and applications (Figure 1). From that the various data management steps were developed in WEM [15], using DCM-services (ZIB-services).

In the WEM no-code platform a DCM-services application was constructed for perinatology. WEM requires only configuration of applications and reduces the need to program. WEM consists of a workflow engine, on top of which screens can be designed. For each workflow step and screen, data can be defined using the WEM internal data ontology or accessing the DCM-services. WEM has an open API, facilitates generic SOAP (Simple Object Access Protocol) and REST (REpresentational State Transfer), HL7 v3 messaging and HL7 FHIR data exchange.

The DCM services handles five levels of data management. The most granular level is the management of many value sets, allowing meta information (value set name, unique object identifier (OID), version and responsible party to be identified, as well as each term (value) and semantic code, increasingly coming from SnomedCT. Then, the management of single data can be specified with similar features as value sets. On level 3, full DCM's, e.g. Apgar Score, diagnoses, pregnancy can be applied. On level four, compositions can be created, for example, the subsets with data and DCM of preceding pregnancies, or birth data, or delivery reports. Finally, all perinatal data together form the top level, that is a complete perinatal dataset of one version [2,6,12,14]. As an add on to this application, a mapping definition component was created. This facilitates cross-mapping on the levels of values, data, DCM, and compositions from source dataset to target dataset (Figure 1).

It is possible to create multiple types of mappings, some of which are simple and others more complex. One of the simpler types of mapping is the date of birth of the child (date and time to date-time-stamp mapping). In this case there are two source elements (i.e. date of birth and time of birth) that need to be merged to one target date-timestamp of the date of birth of the child. In some cases, the professional has registered a value that is a dummy, as the true value is unknown. For the source element date of birth these dummy values are 1-1-1900 and 31-12-1900, and for the source element time of birth this is 99:99 o'clock. These values are entered as dummy values and are automatically mapped to a missing value in the target dataset. Another simple type of mapping is the position of the child (value set to value set mapping). With this type of mapping, actual lines must be drawn between the values of the source and the target elements (Figure 2). There are also some more complex types of mappings where a long list of source values of different source element needs to be mapped to one target element (due to the long list not shown). Following this pilot, another 9 source data-marts will be mapped to perinatology 2.2 data and then be an important database for the data analytics by Perined. And these data marts will be the input for further data processing by Perined.

2. Results

2.1. The Cross-Mapping Execution

The architecture that can handle such diversity of national perinatal data management is described in Goossen [6]. The process described in Figure 1 illustrates how the source data-marts are prepared via cross-mapping to create one large semantically equivalent Perined database. This is a variation on the original architecture. This step was included as an additional approach and the resulting Perined database is in fact the final version of the PDWH.

Perined established a set of requirements for the cross-mapping. These include the components that should be mappable (values in value sets and data elements), the meta information for each component, and cardinalities. Cardinalities include mapping 1 to 1, 1 to many, many to 1, and many to many. And in the mapping both data elements and value set definitions, and the actual instance data need to be included in order to carry out the transform. First is that the data expert at Perined defines per data-element and value in value sets how it shall map from source to target data. This step is depicted in Figure 2. The cross-mapping process is depicted in Figure 1.

2.2. Information Security

Perined takes adequate data protection measures into account. These include adherence to the General Data Protection Regulation and include measures around transparency, informed consent obtained by the professionals, purpose-specification for data uses, legitimate basis in line with the purpose of improving perinatal care on national level and follows (inter)national legislation. In addition, retention measures are taken to store the Perinatal data for trends analysis. Finally, Perined holds ISO 27001 and NEN 7510 certifications based on effective measures for data security.

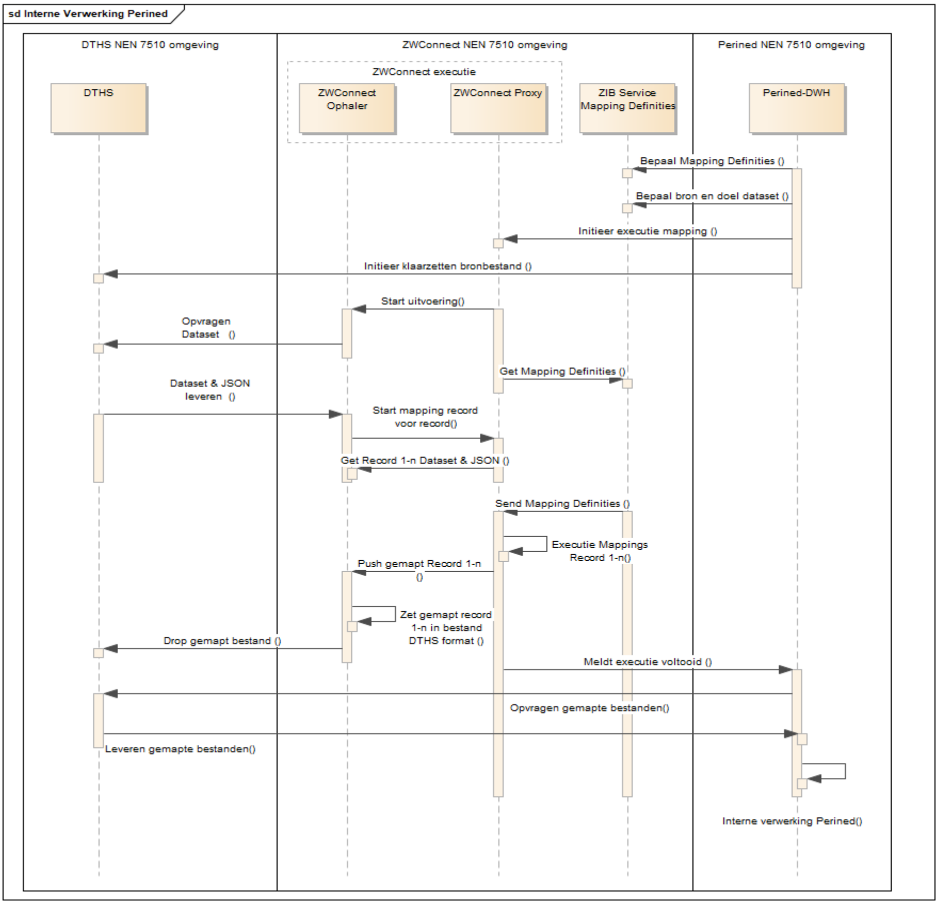

Figure 1. The process sequence from source, via cross-mapping execution, to target.

2.3. Testing

Specific components have been tested and approved, e.g. the data transfer from Perined to WEM and the extraction of a single record. Next the WEM mapping execution could pick up the source data. The mapping definition can be accessed through the internal WEM API leading to a proper transform. The transformed data can be inserted in the target database with mapped and semantically adequate perinatal data.

Acceptance test consisted of a set of 2,000 real world records, each consisting of about 200 distinct data elements, roughly half with value sets. The records were extracted from the data-mart with PRN 1.3 based source data. The data were extracted on a record by record base, entered into the mapping execution application and based on the mapping definitions the data were transformed from the PRN 1.3 format to the target perinatology 2.2 format. The thus transformed data were injected into the target database for the PDWH. Table 1 describes the issues encountered on this level.

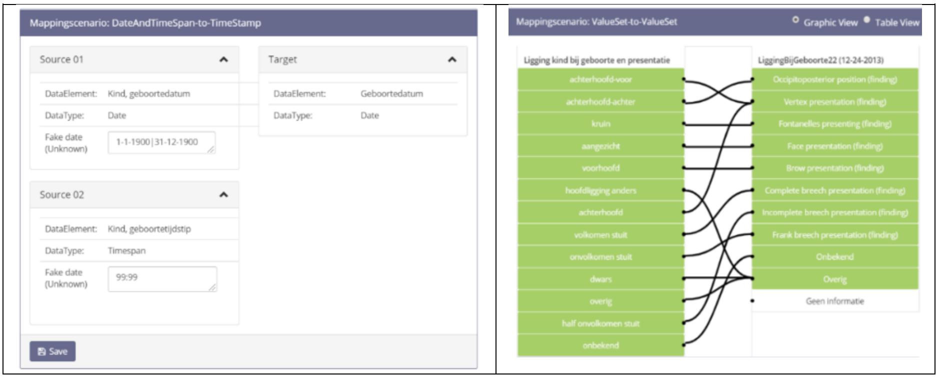

Figure 2. An example of the cross-mapping definition from source to target.

Table 1. Test results from the acceptance test of 2,000 records and 200 data elements

Issue #	Issue	Solution
1	Missing data definitions for data.	Add to either or both source and target data.
2	Time stamp format in string.	Improve datatype to TS
3	Many data in string format.	Correct datatypes
4	Booleans not correct	Change from string to True/False
5	Date and time combined	Separate out date and time in two variables.
6	Data with child birthdate all empty	Data definition checked and time stamp corrected
7	Individual records source and target not linkable	Make the unique record number visible and accessible/ usable in mapping definition step.
8	A Terme date missing value not correctly mapped to 'missing' in target	Checked the data definition, data type and corrected the mapping
9	Coded values 3130, 3140, 3150 missing in transform	Reengineered the mapping definition options from many to many for this data element and codes. Checked for whole set.
10	Composition labels included, but not properly mapped	In target dataset these are obsolete, removed from source data and mapping definition.
11	The data elements g_ddvliesscheur_p22_map, g_ddpers_p22_map & g_ddstartww_vlgzv_p22_map have many outlier values 03-01-1901.	Check on source data, change data type to TS, reengineered the mapping to appropriate target data element.
12	Some data are not defined in the data definition PRN 1.3	Explanation given about the history of versions of PRN 1.4 datasets and changes not properly documented. Update definition.
13	Data elements on smoking are filled with data but not mapped. Definitions could not be changed	Reengineered the definitions for the mapping.
14	Child mortality has an extreme high rate.	Corrected the data type format for the target database.
15	We need besides the record ID also an file identifier	Included at both definition and execution level and for source and target.

3. Conclusion

The design of the Perinatal Data Warehouse facilitates handling the variability of data submitted nationally to Perined. The Perined Information Architecture works well for the data submission, ETL procedure and the storage. However, the semantification of the raw data-marts is an important final step before the analytics can take place [6, 13]. The cross-mapping of PRN 1.3 data to perinatology 2.2 data could be defined in the DCM services tool. The secured mapping execution procedure could handle 2.000 real world cases. Acceptance testing revealed that the mapping, after corrections in the process, the definitions, and the execution, successfully placed the transformed data in the final PDWH database. Several manual steps are no longer required. The next step will be to do this with 180.000 production records, define the automated process and to improve the performance. Additionally, achieving this for PRN 1.3 to peri 2.2 does not guarantee similar mapping results for additional data-marts.

To handle the complexity of the different data definitions and datasets, Perined embarked on semantic cross-mapping. Based on this first test with 2.000 records of 1 dataset, we hope to have created a valid and useable tool for mass automated semantic cross-mapping.

References

[1] Van Hemel, O. J. S., Schutte, F., van Bemmel, J. H., & Chang, A. M. M. Feed-back in an Obstetric Data Base. In Anderson, J., (Ed.) Medical Informatics Europe, 1978, 78,1: 783-796.

[2] Goossen WT., Arns-Schiere AM. Information Architecture for Perinatal Registration in the Netherlands. J Obstet Gynecol Neonatal Nurs. 2017 Mar - Apr;46,2:310-321.

[3] Goossen, W. T., Jonker, M. J., Heitmann, K. U., Jongeneel-de Haas, I. C., de Jong, T., van der Slikke, J. W., & Kabbes, B. L. Electronic patient records: domain message information model perinatology. International Journal Medical Informatics, 2003; 70,2-3: 265-276.

[4] Advanced requirement tooling for data elements, codes, OIDs and rules. Advanced requirement tooling for data elements, codes, OIDs and rules (ART-DECOR.), 2016. From: https://art-decor.org/mediawiki/index.php?title=About

[5] Perined & Mediquest, Perinatal Quality Indicators and Benchmark,2020. From: https://home.mediquest.nl/mijnperined-gaat-geboortezorg-inzicht-geven-in-kwaliteits-en-benchmarkinformatie-van-een-overview-tot-heel-gedetailleerd/

[6] Goossen W. Architecture for variable data entry into a national registry. In: Blobel B, Goossen W., (Eds). Proceedings pHealth 2017 Conference, 2017, Amsterdam, IOS Press, pp 81-90.

[7] Johnson SB. Generic data modeling for clinical repositories. J Am Med Inform Assoc. 1996 Sep-Oct;3,5:328-39.

[8] Nadkarni PM, and Marenco L. Easing the transition between attribute-value databases and conventional databases for scientific data. Proc AMIA Symp. 2001:483-7.

[9] Goossen, W. Detailed clinical models: representing knowledge, data and semantics in healthcare information technology. Healthcare Informatics Research, 2014; 20,3: 163-172.

[10] ISO/TS 29585 Health informatics — Deployment of a clinical data warehouse. Geneva, ISO.

[11] Boterenbrood, F., Krediet, I., & Goossen, W. Building a high-quality medical data architecture for multiple uses in an integrated health care environment. Journal of Hospital Administration, 2014; 3,5. Retrieved from http://www.sciedu.ca/journal/index.php/jha/article/view/3904.

[12] Perined. Inventory of Data Files. [In Dutch] Utrecht, Stichting Perined, 2014.

[13] Goossen-Baremans A, Collins S, Park HA. Semantification in Connected Health. Stud Health Technol Inform. 2017;232:133-151.

[14] National Information and Communication Technology Institute for Health (Nictiz). Perinatology dataset. 2016. From: http://decor.nictiz.nl/perinatologie/peri20-html-20150713T115647/dataset.html.

[15] WEM. No-code software. Retrieved from: https://wem.io/ May 6, 2020.