R. Röhrig et al. (Eds.)

doi:10.3233/SHTI190808

Towards a Federation of Metadata Repositories: Addressing Technical Interoperability

Hannes ULRICH[a,1,*], Jori KERN[b,*], Ann-Kristin Kock-Schoppenhauer[a], Martin LABLANS[b,†] and Josef INGENERF[a,c,†]

[a] *IT Center for Clinical Research, University of Lübeck, Germany*
[b] *Federated Information Systems, German Cancer Research Center, Heidelberg, Germany*
[c] *Institute of Medical Informatics, University of Lübeck, Germany*

Abstract. The utilisation of metadata repositories increasingly promotes secondary use of routinely collected data. However, this has not yet solved the problem of data exchange across organisational boundaries. The local description of a metadata set must also be exchangeable for flawless data exchange. In previous work, a metadata exchange language QL^{4}MDR was developed. This work aimed to examine the applicability of this exchange language. For this purpose, existing MDR implementations were identified and systematically inspected and roughly divided into two categories to distinguish between data integration and query integration. It has been shown that all the implementations can be adapted to QL^{4}MDR. The integration of metadata is an important first step; it enables the exchange of information, which is so urgently needed for the further processing of instance data, from the metadata mappings to the transformation rules.

Keywords. Metadata, Federation, GraphQL, QL^{4}MDR

1. Introduction

The secondary use of electronic health records can improve processes and daily routines of healthcare actors due to the analysis of common events and the possibility of cross-sectoral data exchange [1]. To prepare clinical data for meta-studies or further process analysis, the information from various heterogeneous EHR systems must be integrated. Comprehensible and semantic data integration is necessary, but also an error-prone and time-consuming challenge [2]. Metadata can be used for validation and transformation of instance data: harmonised metadata at schema level generate the transformation of the corresponding instance data, using matchings and mappings between different metadata sets, as conceptually shown in Figure 1. Metadata and MDRs are established tools in the field of medical informatics for years, and metadata representation is often standardised by the international ISO 11179 [3,4].

The standard defines a metamodel and basic attributes for describing metadata but does not provide implementations. As a result, several MDR systems have been

[1]Hannes Ulrich, IT Center for Clinical Research, Lübeck;
Telephone: +49 (0) 451 - 3101 5607;
Fax: +49 (0) 451 - 3101 5604; E-mail: hannes.ulrich@itcr.uni-luebeck.de
*,†These authors contributed equally to this work.

developed, each with an independent interface. Interconnecting these various systems would facilitate joint use of definitions such as mappings and transformations and, ultimately, cross-sectoral data exchange. However, because of technical and syntactical heterogeneity, such a federation does not exist. In a previous study, we designed a graph-based query language, QL⁴MDR [5] enabling federation of various MDR implementations and proved the feasibility by building a prototype into Samply.MDR [6]. The aim of this study is to investigate if and how QL⁴MDR can be integrated into other existing MDR implementations, too.

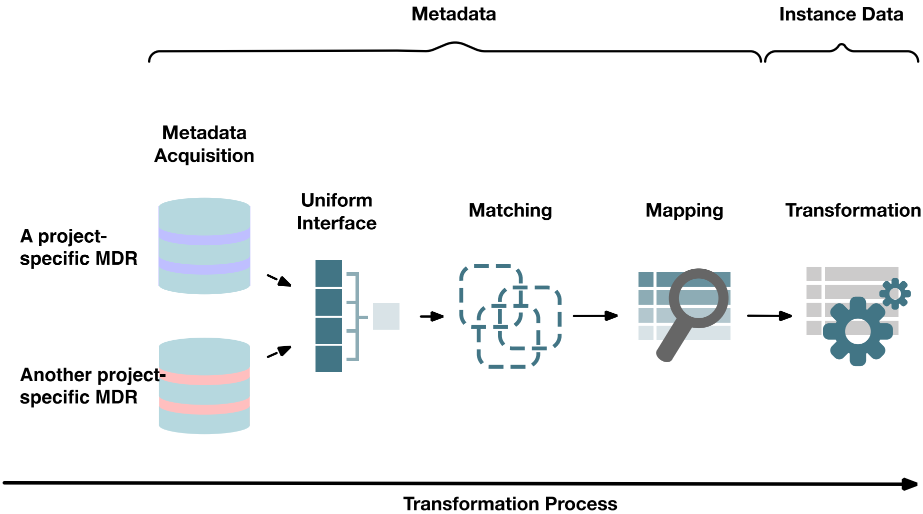

Figure 1: Using metadata to support the integration of healthcare instance data. The process consists of four stages: the metadata acquisition stage with a uniform interface enables the reuse of information which is stored in project-specific MDRs. The matching stage aligns the metadata and identifies potential correspondences. The mapping stage creates transformation rules, which are used in the transformation stage. The first three stages only process metadata, whereas the last transformation stage includes healthcare instance data [5].

2. Methods

After investigating existing MDRs, the technical and administrative components of the MDR implementations included in the review were analysed according to the requirements of QL⁴MDR.

2.1. QL⁴MDR

The developed QL⁴MDR interface [5] is a GraphQL-based API which is based on the ISO/IEC 11179-3 standard to enable metadata queries in a unified way. The underlying schema consists of objects, fields, queries, and mutation types. The definition of entry points corresponding to the key elements of the ISO 11179-3 within the schema allows defining a path through the graph, inquiries, and mutations. QL⁴MDR promotes the accessibility of metadata (one of the FAIR principles [7]) by providing a uniform query interface in a modern web-based interface.

2.2. Review of MDR implementations

In this work, we consider MDR implementations which we found by a systematic online review and met our criteria: healthcare-related, freely accessible, and actively maintained. Afterwards, a technical review was done by a systematic technical review of available information from websites or manuals and by two experts who were involved in the development of the QL^{4}MDR. Four types of distinctions were identified as influencing the way the MDR is being enabled to federate its content: source code access, data source access, ISO 11179 conformity, an existing query interface, comparable communication patterns. In this study, we included six different MDR implementations as seen in Table 1: the Common Data Element Browser of the National Institute of Health [8], the METeOR of the Australian Institute of Health and Welfare and its successor the Aristotle Metadata Registry [9,10], the MDM Portal [11] and ART-DECOR [12]. For the purpose of completeness, we have also included Samply.MDR [6] in the consideration.

Table 1 This table show the results of the systematic review of included metadata repositories. The greyed systems were excluded due to the missing unmet criteria.

	Healthcare-related	Freely accessible	Actively Maintained	Open Source	Data Source access	ISO 11179	Query interface
Common Data Element Browser	Yes	Yes	Yes	No	Yes	Yes	Yes
METeOR	Yes	Yes	Yes	No	Yes	Yes	Yes
Aristotle	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MDM Portal	Yes	Yes	Yes	No	Yes	No	Yes
ART-DECOR	Yes	Yes	Yes	Yes	Yes	No	Yes
Samply.MDR	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Kairos MDR	Yes	**No**	Yes	No	Yes	No	Yes
Clinical Knowledge Manager	Yes	Yes	Yes	No	Yes	No	**No**
Semantic MDR	Yes	Yes	**No**	Yes	Yes	Yes	Yes

3. Results

The analysis showed that there are two primary options to integrate QL^{4}MDR into MDRs, either data integration or query integration. Whereas data integration at least partly requires access to source code or data sources, query integration takes place at a higher level. Due to neither data nor source code access, the query integration needs to be established outside of the MDR and therefore, a requestable interface must be provided.

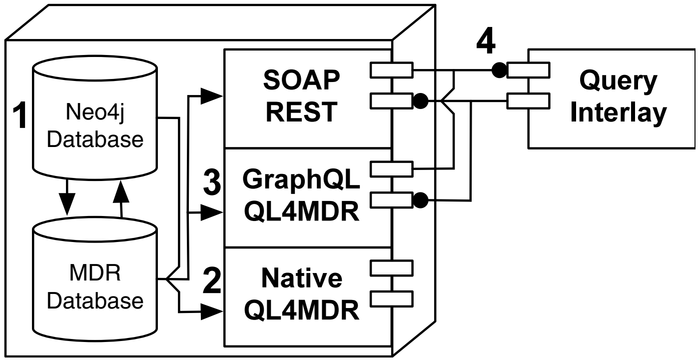

Figure 2 The different integration points to enable an MDR for the use of QL⁴MDR. The numerals are corresponding to the following integration scenarios.

3.1. Option 1: Data Integration

Fortunately, three out of six considered MDR systems shared their underlying source code, so an integration at data-level could be analysed. The integration can be subdivided with three different approaches.

Transformation on data level (1)
If the existing data is not ISO 11179-3 or only partly conform, a data transformation can be applied, and thereby database will be reformatted and transformed into a new reporting database. A suitable approach is to use Neo4j [13], which is a graph-based system with an optional GraphQL plugin. The data transformation should be based on QL^4MDR schema; thus its ISO 11179-3 conformance. Taking the MDM-Portal as an example: the portal provides clinical trials forms in various format, but it is focused on the CDISC ODM format. Stausberg et al. already have stated a transformation of ODM into the ISO 11179 [3], which can quickly be established and the information can be provided using the GraphQL endpoint provided by Neo4j.

Data Manipulation (2)
In a prior study, the integration of QL^4MDR was proven feasible using the existing Samply.MDR [14]. The implementation requires the usage of GraphQL data fetchers to retrieve the data, and deploy it via the interface, and the functional range of the HTTP-based query endpoint. This combination of the GraphQL interface and QL^4MDR is open to be used in any other ISO 11179 based MDRs.

Invasive Interface manipulation (3)
If the MDR already features an existing GraphQL interface, it can be extended with the query format of QL^4MDR. The so-called *schema stitching* allows combining different schemata to fetch all information with one single query [14]. The resource resolver has to be integrated into the existing system, but the pre-existing infrastructure (e.g., authentication or database access layer) can be re-used. However, this means to change the source code of the existing GraphQL interface, which is a suitable scenario for the Aristotle Metadata Registry [9].

3.2. Option 2: Query Integration

If a modification of the MDR implementation's source code is impossible (e.g., because of its license), the integration needs to take place at the query level. An additional external component is needed to transform and translate the incoming requests.

Query Interlay (4)
A query interlay is a component akin to a locally installed communication server and works as a query proxy [15]. It receives the query from a client in a consented format, e.g., QL^4MDR, translates it to the MDR's specific query format (e.g., REST or even another GraphQL schema), and forwards it to the MDRs interface. Accordingly, the query interlay receives the results, aggregates them if needed, transforms them into the consented result format, and then replies to the client. This solution does not change the interface of the MDR itself and depending on the MDR might not be able to provide all functionalities, depending on how well the MDR conforms to the ISO 11179-3 standard. Suitable candidates are ART-DECOR, the Common Data Element Browser, and METeOR, which provides (only) a Restful interface.

4. Discussion

The presented integration scenarios are all valid approaches for addressing the integration of QL4MDR into existing MDRs. They differ in prerequisites and complexity. We identified differences in four main aspects that should be considered.

Regarding the *Development Effort* involved in implementing a given solution, both changes to the source code and the configuration of a given interface have to be considered. Building a query interlay yields considerably more effort than modifying a given interface, as the entire communication to the MDR has to be duplicated. Depending on the existing data structure of the MDR, the data integration scenarios with the transformation on data level might be challenging. The data manipulation as it has been done in proof-of-concept implementation of QL^4MDR [5] appears to require the least development effort due to the direct data access.

Change Management and Data Synchronisation are unavoidable since QL^4MDR provides methods to alter information, thus requiring to synchronise the changed data elements with the source systems. The data integration scenarios seem the most challenging in this aspect. Implementing the transformation on data level requires a second, graph-based database kept in sync with the source data whenever there is a transaction in the primary MDR database. As seen in computer science, avoiding inconsistencies is mandatory. Therefore, it is recommended to either allow data altering via only one interface or trigger a blocking synchronisation. The data manipulation and interface manipulation are easier to handle in this aspect since they are linked to changes of the metadata, not the underlying data structures. The query integration will only be influenced if the system supports data changes via the query endpoint.

Authentication: if a new interface or a new component is introduced into the existing MDR infrastructure it has to implement the already existing authentication processes; otherwise it would risk a breach of security for the MDR. The data integration has little effort in this aspect as the authentication mechanisms of the MDR can be used. However, the query interlay as a new component has to adopt the existing authentication process. This could be a challenging task, depending on the authentication mechanisms in place.

Maintenance and Sustainability: focusing on acceptance and development adoption of QL^4MDR, the integrations have to be maintainable and sustainable. The QL^4MDR component needs to be easily adaptable to adjustments if the MDR interface or the data representation is changing. The data integration contains easier maintainable scenarios, as they are integrated within the system. The query interlay is a new component that has to be implemented in a way that ensures maintenance and sustainability. Due to the effort of a newly implemented component, there will be a trade-off between the development effort and sustainability, since the latter can be addressed by software design.

Given these criteria, it is unavoidable to analyse the MDR that is going to be federated very carefully in order to decide the most suitable method of metadata federation. As a result of our study, we can address two (Art-decor and Samply.MDR) out of four MDRs used within the national medical informatics initiative [16], which is aiming to promote the exchange of medical records in Germany. The missing consortia are establishing its metadata management using a commercial product and therefore did not match our inclusion criterium of free accessibility. The integration study was able to prove that the remaining two established MDRs meet the requirements to integrate QL^4MDR, regardless of the different maintenance effort.

5. Conclusion

Providing technical interoperability to federate metadata repositories is hard and effortful, but attainable. Integrating QL^4MDR into other MDR implementations is possible and is an important first step towards the federation of metadata silos and opening up the carefully curated information for further use, ranging from metadata matching up to data transformation.

6. Acknowledgements

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft) DFG grants IN 50/3-1 and LA 3859/2-1.

References

[1] C. Safran, M. Bloomrosen, W.E. Hammond, S. Labkoff, S. Markel-Fox, P.C. Tang, and D.E. Detmer, Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper, *J. Am. Med. Inform. Assoc.* **14** (2007) 1–9.

[2] K. Khoumbati, M. Themistocleous, and Z. Irani, Integration Technology Adoption in Healthcare Organisations: A Case for Enterprise Application Integration, *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.* (2005) 9.

[3] S.M. Ngouongo, M. Lobe, and J. Stausberg, The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research?, *J. Biomed. Inform.* **46** (2013) 318–327. doi:10.1016/j.jbi.2012.11.008.

[4] CEN TC 251, ISO TC 215/26 Healthcare Informatics Metadata Repository Requirements (MetaRep) Working Draft Stage {Online}, (2017). https://login.cen.eu/portal/.

[5] Ulrich, Jori Kern, Deniz Tas, Ann-Kristin Kock-Schoppenhauer, Frank Ückert, Josef Ingenerf, and Martin Lablans, QL4MDR: A GraphQL Query Language for ISO 11179-based Metadata Repositories, *BMC Med. Inform. Decis. Mak.* (2019).

[6] D. Kadioglu, B. Breil, C. Knell, M. Lablans, S. Mate, D. Schlue, H. Serve, H. Storf, F. Ückert, T. Wagner, P. Weingardt, and H.-U. Prokosch, Samply.MDR - A Metadata Repository and Its Application in Various Research Networks, *Stud. Health Technol. Inform.* **253** (2018) 50–54.

[7] M.D. Wilkinson et al., M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018. doi:10.1038/sdata.2016.18.

[8] P.M. Nadkarni, and C.A. Brandt, The Common Data Elements for cancer research: remarks on functions and structure, *Methods Inf. Med.* **45** (2006) 594–601.

[9] Aristotle Metadata Registry - Home, (n.d.). https://www.aristotlemetadata.com/ (accessed February 20, 2019).

[10] Australien Institute of Health and Welfare, METeOR home, (n.d.). http://meteor.aihw.gov.au/content/index.phtml/itemId/181162 (accessed June 29, 2018).

[11] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland, and J. Varghese, Portal of medical data models: information infrastructure for medical research and healthcare, *Database*. **2016** (2016). doi:10.1093/database/bav121.

[12] art-decor.org, (n.d.). https://art-decor.org/mediawiki/index.php/Main_Page (accessed February 13, 2017).

[13] J. Webber, and I. Robinson, A programmatic introduction to neo4j, Addison-Wesley Professional, 2018.

[14] T. Frisendal, Getting the Structure Right, in: Vis. Des. GraphQL Data, Springer, 2018: pp. 45–60.

[15] M.L. Braunstein, and D. Detmer, Interoperable informatics for health enterprise transformation, *J. Enterp. Transform.* **6** (2016) 110–119.

[16] S.C. Semler, F. Wissing, and R. Heyder, German medical informatics initiative, *Methods Inf. Med.* **57** (2018) e50–e56.