Received: 8 November 2016 | Revised: 5 July 2017 | Accepted: 12 July 2017

DOI: 10.1002/lrh2.10037

Learning Health Systems

EXPERIENCE REPORT

The TRANSFoRm project: Experience and lessons learned regarding functional and interoperability requirements to support primary care

Jean-François Ethier[1,2] | Mark McGilchrist[3] | Adrien Barton[1] | Anne-Marie Cloutier[1] | Vasa Curcin[4] | Brendan C. Delaney[5] | Anita Burgun[2]

[1] Department of Medicine, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Canada

[2] INSERM UMR 1138 team 22 Centre de Recherche des Cordeliers, Faculté de médecine, Université Paris Descartes—Sorbonne Paris Cité, Paris, France

[3] Division of Population Health Sciences, University of Dundee, Dundee, UK

[4] Division of Health and Social Care Research, Faculty of Life Sciences and Medicine, King's College London, London, UK

[5] Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK

Correspondence
Jean-François Ethier, Faculty of Medicine and Health Sciences, Université de Sherbrooke, 3001, 12th Avenue North, Sherbrooke (Quebec), J1H 5N4, Canada.
Email: ethierj@gmail.com

Abstract

Introduction: The current model of medical knowledge production, transfer, and application suffers from serious shortcomings. Learning health systems (LHS) have recently emerged as a potential solution—systems in which health information generated from patients is continuously analyzed to improve knowledge that will be transferred to patient care.

Method: Various approaches of data integration already exist and could be considered for the implementation of a LHS. We discuss what are the possible informatics approaches to address the functional requirements of LHS, in the specific context of primary care, and present the experience and lessons learned from the TRANSFoRm project.

Result: Implemented in 4 countries around 5 systems, TRANSFoRm is based on a local-as-view data mediation approach integrating the structural and terminological models in the same framework. It clearly demonstrated that it has the potential to address the requirements for a LHS in primary care, by dealing with data fragmented across multiple points of service. Also, it has the potential to support the generation of hypotheses from the context of clinical care, retrospective and prospective research, and decision support systems that improve the relevance of medical decisions.

Conclusion: The LHS approach embodies a shift from an institution-centered to a patient-centered perspective in knowledge production and transfer and can address important challenges in the primary care setting.

KEYWORDS

interoperability, learning health system, primary care, requirements, TRANSFoRm project

1 | INTRODUCTION

The traditional model of medical knowledge production, transfer, and application can be schematically described as follows. Research hypotheses are often generated from fundamental research (the "bench"). To test those hypotheses, research cohorts and randomized controlled trials (RCTs) are used to generate research data that can lead to new (or updated) knowledge. This knowledge can then be transferred to health care (the "bedside"), eg, through clinical decision support tools.

However, this model suffers from serious shortcomings. First, it would be valuable to generate hypotheses directly from the context of clinical care—as illustrated, by a study that showed from clinical care data that metformin is associated with an improved survival rate for diabetic patients.[1]

List of acronyms: CDISC, Clinical Data Interchange Standards Consortium; EHR, Electronic Health Record; HL7, Health Level 7; ICPC, International Classification of Primary Care; LHS, Learning Health System; PC, Primary Care; RCT, Randomized Controlled Trial

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2017 The Authors. Learning Health Systems published by Wiley Periodicals, Inc. on behalf of the University of Michigan

Second, there are pragmatic issues in executing research projects based on cohorts and RCTs, which have been in a long-term crisis. Both cost and the difficulty of recruiting patients to participate are issues[2]—as well as the well-described attrition risk for research cohorts. Additionally, RCTs are prone to bias in the selection of eligible subjects, controls, and outcomes measures. Moreover, they tend to have a low external validity, including in primary care (Fortin et al, [3]); as a matter of fact, a correct estimation of safety and effectiveness requires clinical trials in real-world settings, as diagnostic and therapeutic features are not necessarily transferable across populations with different prevalence and spectrums of disease.[4] However, the vast majority of research, be it diagnostic or intervention based, takes place in specialized centers and involves highly selected populations; consequently, patients in RCTs tend to be younger and healthier than real-world populations typically seen in primary care.[5,6]

Third, knowledge transfer is a slow process: even in countries like the UK, where a national agency (the National Institute for Health and Care Excellence) is funded to carry out this process, guidelines might be updated only once in a decade.[6] Moreover, clinicians encounter an increasing problem of memorization and prioritization of the potential applicable guidelines to a given patient at a given point on his health care trajectory.[7]

Finally, the 3 problems identified above are seriously compounded by a common technological issue: health and research data can be expressed according to various semantics, which are often poorly interoperable. To understand the complexity of the issue, note that the semantics of a piece of data constituted by a terminological code in a database field is encapsulated in 2 elements: the terminological code itself, and the structure of the database—this is the so-called "binding" of structural and terminological information.[8] For example, a code such as "ICPC-X76" (from the International Classification of Primary Care) refers to breast cancer, but depending on the structure of the database in which it is located, an instance of this terminological code can denote various diagnoses of breast cancer, such as the diagnosis of a current condition of the patient, of a past condition of the patient, or of the current condition of a family member.

A potential solution to those problems has emerged with the concept of learning health system (LHS)—a system in which health information generated from patients within that system is continuously analyzed to improve knowledge that will be transferred to patient care (Figure 1). Various sources characterize LHS differently, but the IOM[9,10] defines the LHS as a vision for an integrated health system "... in which progress in science, informatics, and care culture align to generate new knowledge as an ongoing, natural by-product of the care experience, and seamlessly refine and deliver best practices for continuous improvement in health and healthcare."

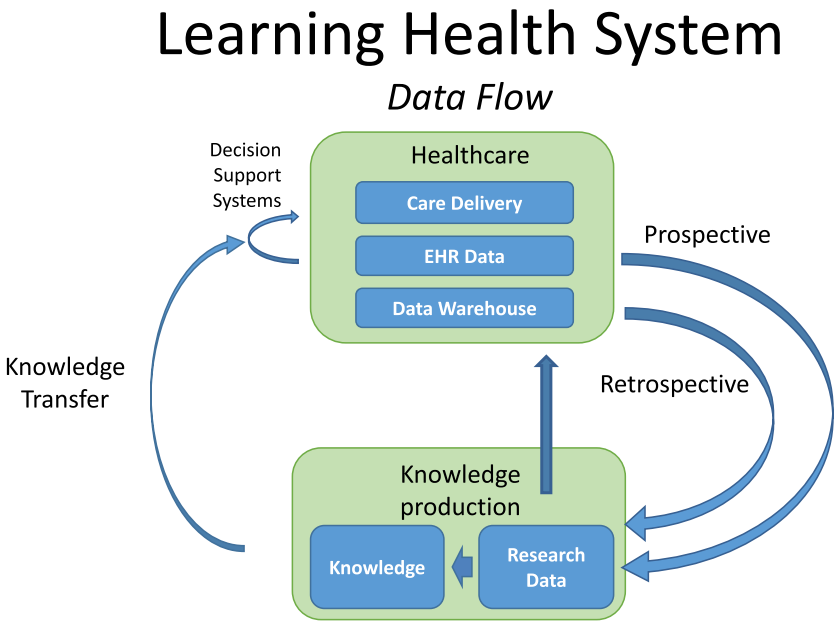
We will present in this paper the lessons learned from the TRANSFoRm project, a recent EU FP7 (7th Framework Programme for Research and Technological Development) project that aimed at comprehensively supporting the integration of clinical and translational research data in the primary care (PC) domain as part of a learning health care system to enhance patient safety.[7,11,12] To our knowledge, it was the first international LHS—being implemented in 5 countries (UK, Netherlands, Greece, Belgium, and Poland) around 6 systems—and the first LHS including PC.

This paper is structured as follows. First, we expose the background requirements for data integration in a LHS supporting primary care. Second, we present the various approaches of data integration that already exist, and how data mediation constitutes the most promising approach in a primary care context. Third, we show how the TRANSFoRM project was designed as a proof of concept for data mediation in a LHS. A discussion and conclusion follow.

2 | BACKGROUND

2.1 | The requirements of a LHS supporting primary care

A LHS has to satisfy a number of requirements,[13] which include the following functional requirements. First, it should support prospective and retrospective research. Second, it should minimize the resources



Colour online, B&W in print

FIGURE 1 Learning health system—data flow

required from an institution to participate in research projects. Third, it should enable decision support systems, by being able to provide relevant pieces of information, feedbacks, and alerts depending on both patient and population data. Fourth, it should be flexible enough to accommodate the addition of new data sources—or deletion or change of current ones. Fifth, it should enable data integration from multiple sources belonging to various fields, such as care, research, and knowledge data, by ensuring their interoperability.

2.2 | Specificities of primary care

Primary care services are the cornerstone of the health care system. Nevertheless, implementing a LHS in this context is ambitious and faces many obstacles. First, to provide access to community-based care for a greater population, PC is fractionated across multiple points of services; patient data are accordingly fragmented, making it difficult to get a complete picture for each patient. Another challenge results from the variability of numbers of patients and their clinical or demographic characteristics across primary care clinics, even within the same neighborhood. Those parameters are often unknown, and this complexifies the selection of the optimal set of clinics to answer a given research question.

Besides, various clinical and administrative activities have different requirements and often use their own information systems, with specific underlying processes and data models to support them. As a result, multiple electronic health records (EHR) systems are used in PC. To address this, a LHS cannot force EHR vendors to structure data differently; thus, a unique structural model cannot be imposed. Moreover, those multiple PC institutions have different mandates and legal frameworks,[14] which makes it impossible to copy all data into a central location.

Typically, PC clinicians are the first contact and main anchor for chronic disease long-term care. Consequently, reasons for consultation may vary significantly and are often not as clear as in specialized care, especially for undifferentiated symptoms or diseases. The scope of data generated and its granularity, as well as the way data are collected and stored, do vary and have an impact on their usability to answer a specific question. Because they are not restricted to 1 physiological system, PC institutions are greatly solicited to participate to various research projects which require significant resources. Because PC resources are limited and vary among sites, a successful implementation of a LHS in a PC context must minimize resources required from institutions to get them participating, and, often overlooked, maintain their participation over time.

Despite those difficulties, the implementation of a LHS in PC is a project of prime importance. As a matter of fact, PC can significantly benefit from the diagnostic decision support in a LHS, as it includes many patients presenting with undifferentiated problems, requiring timely and safe diagnostic activities. We will now present the possible informatics approaches that have been evaluated in the context of the TRANSFoRm project to address the functional requirements for a LHS in PC.

3 | METHODS

Various approaches of data integration already exist and could be considered for the implementation of a LHS (see Table 1 for a summary). A first approach is "data warehousing" (see[15] for a recent review): integrating various data sources into a common data warehouse—typically using an extract-transform-load (ETL) process.[16] This approach is typically used to integrate various sources within an institution,[17] where they can be leveraged to facilitate retrospective analysis by gathering patients with similar characteristics in retrospective patient cohorts. They could be considered for application on multiple sources scattered across institutions. However, as previously mentioned, institutions

TABLE 1 Summary of the various methods, their advantages, and limits

Method	Principle	Advantages	Limits
Data warehousing	Various data sources are integrated into a common data warehouse.	Facilitates retrospective analysis by gathering patients with similar characteristics in retrospective patient cohorts. Generally good performance.	- Distinct mandates, regulatory, and legal frameworks for PC institutions across various countries constitute an obstacle to deliver data. - Its use is limited to feed decision support systems or optimally recruit patients while in clinic.
Data federation	All data sources are structured identically.	- The same query could be run at each site, and data could be easily aggregated. - Data are transmitted only when needed and allowed by the local data source curator.	- It is unrealistic to require that PC institutions would coordinate and agree to use the same data structures and terminologies.
Data mediation…	Local data source models are mapped to a central model that supports query expression.	- Each data source can keep its own structure and terminology. - Data are transmitted only when needed and allowed by the local data source curator.	
… global-as-view	The central model is a view of the sum or union of each local model.	- Efficient.	- Risks of asynchrony and incoherence in a context in which the sources are not predetermined.
… local-as-view	The central model is designed independently of any local source.	- More coherent and stable than global-as-view.	- Mapping each data source model to a central model is time consuming. - Somewhat less efficient than global-as-view.

have distinct mandates or regulatory and legal frameworks, and most of them would not accept to deliver data in a bulk. The problem is compounded in the context of a LHS including PC, with data fragmented across multiple organizations, possibly in multiple countries.[18] Moreover, there is a delay between data generation in the source system and its transfer into the data warehouses. While this delay has been diminished in more recent implementations, it is significantly limits data warehouse use to feed decision support systems or optimally recruit patients as they are seen in clinic for new problems.

A second conceivable approach is called data federation, in which all data sources are structured identically. In such a system, the same query could be run at each site, and data could be easily aggregated. By contrast to a hypothetical data warehousing system that would span various institutions, data could reside in each institution and does not need to be stored in a central location: the relevant pieces of data could be transmitted only when needed and allowed by the data source curator. However, this approach is also not viable in the context of a LHS that would span from primary to tertiary care: it is unrealistic to require that PC institutions would coordinate and agree to use the same data structures and terminologies.

As a matter of fact, data federation can be seen as a specific case of the so-called "data mediation" approach,[19] in which a central model is designed to support query expression sent to the system. Data mediation[20-22] can be implemented without imposing the same data structure and terminologies to the participating data sources. Local models can be produced to represent the structure of each data sources and then mapped to the central model. Queries can then be formulated based on the central model and translated in each data source into a query that can be executed locally; data are then returned centrally and aggregated. This system present several advantages: each data source can keep its own structure and terminology, and data are transmitted only when needed and allowed by the local data source curator. For those reasons, a data mediation system is the preferred choice for a LHS involved in PC, which needs to integrate data from multiple sources, when centralization or change of data source structures cannot be mandated.

A data mediation system requires a mapping between the central model and each local model.[23] This can be implemented in 2 different ways. In the global-as-view approach,[24] the central model is a view of the sum or union of each local model—and therefore a direct reflection of the available sources at a specific time. However, any change in the local sources may induce a change in the central model, which raises serious risks of asynchrony and incoherence with the platform applications using it. This is a problem, as all sources that will ultimately be part of the system cannot be known at the start, the contact with some sources can be lost during the implementation of the LHS, and sources will evolve over time. Therefore, an alternative approach called "local-as-view" is often preferred, in which the central model is designed independently of any local source. Such a system may present some performance impact in specific use-cases, but it is significantly more stable: if a local source is modified, only the mapping between this source model and the central model need to be updated, whereas the mappings involving other local sources are unaffected.

Note that the central model cannot rely on query requirements elicited by a focus group of users, as many queries to be executed in the LHS cannot be known ahead of time. Therefore, such a system requires its central model to evolve cumulatively: any query that can be formulated with a former version of the central model should still be expressible with a newer version, and the gain in expressivity of the newer version should enable to formulate new queries. To devise such central models, "ontologies" are a natural tool: they are structured terminological framework represented in a computerized form, which formalize and explicit the logical relations between the entities of a domain. For example, the ontology FMA (Foundational Model of Anatomy[25]) defines the entities *Organ*, *Heart* or *Mitral_valve*, and formalize relations such as *Heart is_a Organ* or *Mitral_valve part_of Heart*. Well-designed ontologies—especially those who are built according to the so-called realist methodology—can provide a central model that can evolve cumulatively.[11] To sum up, a data mediation system in the local-as-view approach whose central model is built as an ontology can enable a dynamic—rather than static—approach of interoperability, as it can accommodate both new queries and changes of available data sources.

We will now present how this local-as-view data mediation system has been used in the context of the TRANSFoRm project.

4 | THE TRANSFoRm EXPERIENCE

As mentioned earlier, the TRANSFoRm project was a proof of concept of an international LHS based on PC, supporting 3 different kinds of applications: recruitment for a prospective study, analysis for a retrospective genomic-clinical study, and decision support.

TRANSFoRm involved a variety of data sources in different computer formats (such as relational databases and XML data extract files), with different structures, and using different terminologies. Given this heterogeneity, TRANSFoRm was based on the data mediation with a local-as-view approach, as described earlier. Developing this LHS required to develop a unified structural/terminological interoperability framework to enable data integration from various sources. On the structural side, data source models had to be mapped with a central model. On the terminological side, terminological codes pertaining to different terminologies but referring to the same real-world entity (eg, the same disease) had to be represented as synonyms.

In this part, we will first explain how the central model was represented as an ontology; second, we will explain how LexEVS was used both to map the data source models to the central model, and to represent synonymy relations between terminological codes; third, we will present the applications of TRANSFoRm.

4.1 | The ontology CDIM as the central model

The central model in TRANSFoRM was represented as an ontology named CDIM (Clinical Data Integration Model[11]), which was designed to represent clinical entities relevant to primary care. CDIM was built according to OBO Foundry principles[26]—a set of principles guaranteeing compatibility between modular, open-access, and complementary ontologies.

The OBO Foundry is to date one of the most significant attempts to build interoperable ontologies in the biomedical domain. When building an ontology for a domain not yet covered in the OBO

Foundry, categories from former OBO Foundry ontologies should be reused whenever possible, leading to a largely cumulative development. It therefore provides a natural framework for a central model like CDIM that could be expanded and reused in future projects.

4.2 | Using LexEvs for structural/terminological binding

Several tools such as LexEvs and Bioportal[27] have been developed to facilitate the management of semantic biomedical resources. LexEVS was chosen for 2 main reasons. First, LexEVS can be used locally, whereas Bioportal is based on a central server. This enables higher control over the information, with 2 noticeable advantages: there is no risk of losing access to our data when the central system is updated, and there is no need to store confidential information on an outside system (because of confidentiality concerns, EHR companies would not agree to put their EHR data model in a public server). Second, LexEVS enables more flexibility. While LexEVS was initially designed to accept terminology files, loader helper modules can be created to expand its use to load structural models too. This enables coherent and complete binding within the same system (LexEVS) to subsequently serve the information uniformly.

This project made a 2-fold use of LexEvs (see Figure 2). A first, classical use was to map codes between various terminologies—with relations of synonymy or quasi-synonymy. For example, ICPC-T90.2 and ICD10-E11 can be mapped by a synonym relation, as they both refer to diabetes mellitus type 2. Codes can also be gathered in "value sets", which are groups of terminological codes—such as the value set for diagnosis, the value set for symptoms, or the value set for infection causes.

The second, novel use of LexEVS consists in mapping the central model CDIM with each local source models, relating entities from the central model with combination of fields in the local data source. Thus, it integrated both structural and terminological models in the same framework. Finally, it represented the binding between the structural and terminological framework by relating entities from the central model to value sets—for example, by relating the entities "patient current condition diagnosis" or "patient past condition diagnosis" to the value set for diagnosis. Traditionally, the structural models of the data source have not been available in a standardized manner. This is important to enable coherent and efficient structural/terminological binding.

For example, in Figure 2, a field in a database might be named Dx and contain the value ICPC-T90. Given that Dx represents a patient current diagnosis, and the term T-90 denotes non-insulin dependent diabetes in the International Classification of Primary Care 2 (ICPC-2), we can assert that this represents a current diagnosis of non-insulin dependent diabetes. To express this knowledge in a general fashion, Dx is mapped to the CDIM entity "Patient current diagnosis", and ICPC-T90 is mapped to synonym entities in other terminologies such as ICD10-E10.

Finally, LexEvs supports HL7 CTS2 (common terminology services 2) standards, which are an agreed upon set of methods to interact with a terminology server. Thus, the LHS can serve the structural data source models to any application in a format congruent with those standards. This is important to facilitate reuse of information across systems instead of always recreating new models in different systems.

4.3 | Applications

The LHS was applied on 3 use cases. First, a retrospective diabetes clinical-genomic study, which demonstrated that linked genomic and clinical data, could be used across several countries using the same platform. Second, a diagnostic decision support system in case of chest pain, abdominal pain, and shortness of breath,[28] which demonstrated that giving early prompts of diagnosis to clinicians was more favorable than giving late prompts.[29] Third, a prospective acid reflux clinical study, an international RCT evaluating dosing regimen of proton pumps inhibitors (medication used for gastric reflux) on multiple sites.[2] Note that the LHS supported existing standards such as the CDISC Operational Data Model, enabling prospective RCT to be developed using CDISC research standards (in the form of the Operational Data Model, ODM—cf.[30]) and to be deployed on the TRANSFoRM platform. Study recruited over 600 patients in 4 countries (Poland, Greece, UK, Netherlands) and 5 different EHR Software suites.[31]

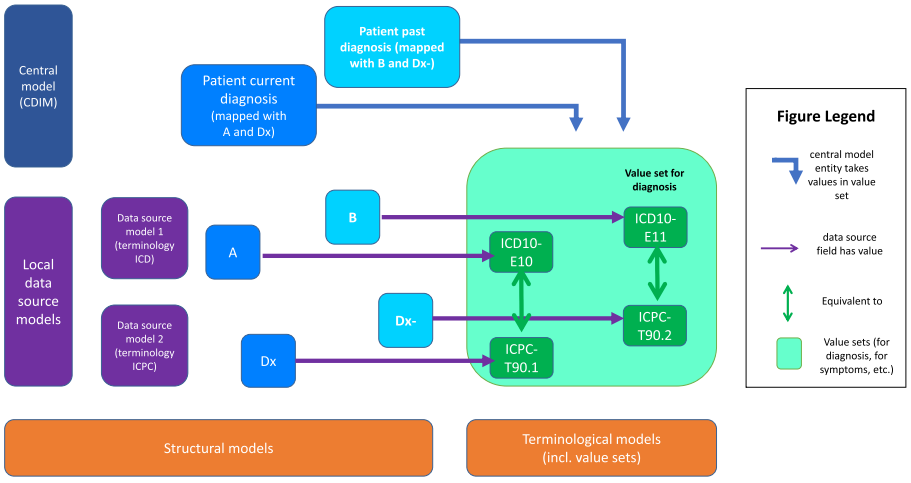


FIGURE 2 Using LexEVS for structural/terminological binding

Colour online, B&W in print

5 | DISCUSSION AND CONCLUSION

The framework described above can address the challenges mentioned earlier in knowledge production, transfer, and application.

The mediation approach can allow a LHS to address the interoperability issue by providing a unified structural-terminological interoperability framework encompassing several kinds of mappings: between the central model ontology and the local data models, between various terminologies, and between the models and value sets of terminological codes. Such a dynamic approach of interoperability can integrate new data sources along the way and remain stable when former sources are changed or lost. Thus, such a system could likely be extended to enable the integration of various styles of data sources (existing cohorts, RCTs, omics, laboratories, etc.) whose structures and supported terminological resources differ.

The mediation approach enables a LHS to address the 3 problems mentioned in the introduction. By integrating health care data from various sources, it has the potential to support a system generating hypotheses from the context of clinical care. The value of hypothesis generation was demonstrated eloquently by the association between Metformin (diabetes drug) and better cancer survival, leading back to fundamental research work to elucidate this potential effect.[32] While having a significant impact on large-scale projects, it can also be quite helpful to help generate quality improvement questions in a clinic, e.g., "which proportion of our diabetic patients have had a lab test (HBA1C) in the last year?" By facilitating audit and feedback activities for health care professionals, it facilitates knowledge transfer, eg, indicating relevant scientific articles for improving follow-up of diabetic patients.

The platform can also support retrospective and prospective research, by assessing the feasibility of research and identifying potential research sites. More specifically, it can support "pragmatic RCTs", which use data generated from health care to recruit patients and pre-load electronic case report forms for RCTs. Not only can it facilitate patient recruitment, it can also pre-populate electronic case-report forms with EHR record data as demonstrated in TRANSFoRm. Moreover, the outcome assessment is facilitated via routine data collection. In a recent controlled effectiveness trial conducted in 75 general practices, Vestbo et al[33] used an integrated primary and secondary care EHR to collect their outcomes measures and report adverse events in real time.

Finally, a LHS can also use routine EHR data to support decision support systems that improve the relevance of medical decisions by contextualizing guidelines upon the characteristics of the target population. In particular, it can support diagnostic decision aids, which are less investigated and more difficult to devise than, eg, therapeutic decision aids,[34] and require the use of data captured during the medical visit.

Such a system has therefore the potential to address the requirements for a LHS in PC, by dealing with data fragmented across multiple points of service, which have populations of patients with various clinical and demographic characteristics, use different data structures, containing data with various scope and granularity, while lowering the investment in time and resources required from those facilities. To summarize, the LHS approach embodies a shift from an institution-centered to a patient-centered perspective in knowledge production and transfer.

The approach presented here could be extended in various directions. First, it could also include specialized care linked with PC. Second, the central model ontology could be extended. Currently, most queries need to be written using both the language of the ontology and some terminological codes; as an example, querying for patients with diabetes type 2 would require to refer to diabetes type 2 using a terminological code such as ICPC-T90.2 or ICD10-E10. However, if the ontology was to encompass all disease entities, then queries could be formulated using only the ontological language. Such disease entities could be modelled in the framework of the Ontology for General Medical Science (OGMS) [3], which provides a general model of disease.

Third, a challenge would be to integrate complex temporal reasoning, such as whether a patient was hospitalized on a ward during a period of nosocomial infection such as *Clostridium difficile*. Fourth, it would be conceivable to formulate ethical guidelines in the language of the central model ontology that would determine whether a submitted query is readily ethically acceptable, unacceptable, or requires further evaluation by ethics authorities.

CONFLICT OF INTEREST DISCLOSURE

No conflict of interest to be declared by any co-author.

ORCID

Jean-François Ethier http://orcid.org/0000-0001-9408-0109
Brendan C. Delaney http://orcid.org/0000-0002-3518-0131

REFERENCES

1. Eurich DT, Majumdar SR, McAlister FA, Tsuyuki RT, Johnson JA. Improved clinical outcomes associated with metformin in patients with diabetes and heart failure. *Diabetes Care*. 2005;28:2345-2351.
2. Mastellos N, Andreasson A, Huckvale K, et al. A cluster randomised controlled trial evaluating the effectiveness of eHealth-supported patient recruitment in primary care research: the TRANSFoRm study protocol. *Implement Sci IS*. 2015;10:15. https://doi.org/10.1186/s13012-015-0207-3
3. Fortin M, Dionne J, Pinho G, Gignac J, Almirall J, Lapointe L. Randomized controlled trials: do they have external validity for patients with multiple comorbidities. *Ann Fam Med*. 2006;4(2):104-108. https://doi.org/10.1093/fampra/cmv012
4. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med*. 1997;16:981-991.
5. Louie V, Latt NL, Gharibian D, et al. Real-world experiences with a direct-acting antiviral agent for patients with hepatitis C virus infection. *Perm J*. 2017;21. https://doi.org/10.7812/TPP/16-096
6. Martin K, Bégaud B, Latry P, Miremont-Salamé G, Fourrier A, Moore N. Differences between clinical trials and postmarketing use. *Br J Clin Pharmacol*. 2004;57:86-92. https://doi.org/10.1046/j.1365-2125.2003.01953.x
7. Delaney BC, Curcin V, Andreasson A, et al. Translational medicine and patient safety in Europe: TRANSFoRm—architecture for the learning health system in Europe. *Biomed Res Int*. 2015;2015:961526. https://doi.org/10.1155/2015/961526
8. Rector AL, Qamar R, Marley T. Binding ontologies and coding systems to electronic health records and messages. *Appl Ontol*. 2009;4:51-69.

9. Committee on the Learning Health Care System in America, Institute of Medicine, 2013. Best care at lower cost: the path to continuously learning health care in America. National Academies Press (US), Washington (DC).
10. Institute of Medicine (US), 2011. Digital infrastructure for the learning health system: the foundation for continuous improvement in health and health care: workshop series summary, The National Academies Collection: Reports funded by National Institutes of Health. National Academies Press (US), Washington (DC).
11. Ethier J-F, Curcin V, Barton A, et al. Clinical data integration model. Core interoperability ontology for research using primary care data. *Methods Inf Med*. 2015;54:16-23. https://doi.org/10.3414/ME13-02-0024
12. Ethier J-F, Dameron O, Curcin V, et al. A unified structural/terminological interoperability framework based on LexEVS: application to TRANSFoRm. *J Am Med Inform Assoc JAMIA*. 2013;20:986-994. https://doi.org/10.1136/amiajnl-2012-001312
13. Friedman C, Rubin J, Brown J, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Inform Assoc JAMIA*. 2015;22:43-50. https://doi.org/10.1136/amiajnl-2014-002977
14. Bastião Silva LA, Días C, van der Lei J, Oliveira JL. Architecture to summarize patient-level data across borders and countries. *Stud Health Technol Inform*. 2015;216:687-690.
15. Khnaisser C, Lavoie L, Diab H, Ethier J-F. Data warehouse design methods review: trends, challenges and future directions for the healthcare domain. In: Morzy T, Valduriez P, Bellatreche L, eds. *New Trends in Databases and Information Systems, Communications in Computer and Information Science*. Springer International Publishing; 2015:76-87.
16. Vassiliadis, P., Simitsis, A., Skiadopoulos, S., 2002. Conceptual modeling for ETL processes, in: Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP, DOLAP '02. ACM, New York, NY, USA, pp. 14–21. https://doi.org/10.1145/583890.583893
17. Huser V, Cimino JJ. Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories. *AMIA Annu Symp Proc AMIA Symp AMIA Symp*. 2013;2013:648-656.
18. Weber GM. Federated queries of clinical data repositories: scaling to a national network. *J Biomed Inform*. 2015;55:231-236. https://doi.org/10.1016/j.jbi.2015.04.012
19. Wiederhold G. Mediators in the architecture of future information systems. *Computer*. 1992;25:38-49. https://doi.org/10.1109/2.121508
20. Hernandez T, Kambhampati S. Integration of biological sources: current systems and challenges ahead. *SIGMOD Rec*. 2004;33:51-60. https://doi.org/10.1145/1031570.1031583
21. Lenzerini, M., 2002. Data integration: a theoretical perspective, in: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '02. ACM, New York, NY, USA, pp. 233–246. https://doi.org/10.1145/543613.543644
22. Louie B, Mork P, Martin-Sanchez F, Halevy A, Tarczy-Hornoch P. Data integration and genomic medicine. *J Biomed Inform*. 2007;40:5-16. https://doi.org/10.1016/j.jbi.2006.02.007
23. Shvaiko P, Euzenat J. A survey of schema-based matching approaches. In: Spaccapietra S, ed. *Journal on Data Semantics IV*. Berlin Heidelberg: Lecture Notes in Computer Science. Springer; 2005:146-171.
24. Ashish N, Ambite JL, Muslea M, Turner JA. Neuroscience data integration through mediation: an (F)BIRN case study. *Front Neuroinform*. 2010;4:118. https://doi.org/10.3389/fninf.2010.00118
25. Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003;36:478-500. https://doi.org/10.1016/j.jbi.2003.11.007
26. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol*. 2007;25:1251-1255. https://doi.org/10.1038/nbt1346
27. Noy NF, Shah NH, Whetzel PL, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res*. 2009;37:W170-W173. https://doi.org/10.1093/nar/gkp440
28. Corrigan D. An ontology driven clinical evidence service providing diagnostic decision support in family practice. *AMIA Jt Summits Transl Sci Proc AMIA Jt Summits Transl Sci*. 2015;2015:440-444.
29. Kostopoulou O, Porat T, Corrigan D, Mahmoud S, Delaney BC. Diagnostic accuracy of GPs when using an early-intervention decision support system: a high-fidelity simulation. *Br J Gen Pract*. 2017;67(656):e201-e208. https://doi.org/10.3399/bjgp16X688417. Epub 2017 Jan 30.
30. Bruland P, Breil B, Fritz F, Dugas M. Interoperability in clinical research: from metadata registries to semantically annotated CDISC ODM. *Stud Health Technol Inform*. 2012;180:564-568.
31. Ethier J-F, Curcin V, McGilchrist MM, Lim SN, et al. eSource for clinical trials: Implementation and evaluation of a standards-based approach in a real world tria. *International Journal of Medical Informatics*. 2017;106:17-24. ISSN 1386-5056, https://doi.org/10.1016/j.ijmedinf.2017.06.006
32. Landman GWD, Kleefstra N, van Hateren KJJ, Groenier KH, Gans ROB, Bilo HJG. Metformin associated with lower cancer mortality in type 2 diabetes: ZODIAC-16. *Diabetes Care*. 2010;33:322-326. https://doi.org/10.2337/dc09-1380
33. Vestbo J, Leather D, Diar Bakerly N, et al. Effectiveness of fluticasone furoate–vilanterol for COPD in clinical practice. *N Engl J Med*. 2016;375:1253-1260. https://doi.org/10.1056/NEJMoa1608033
34. Kostopoulou O, Lionis C, Angelaki A, Ayis S, Durbaba S, Delaney BC. Early diagnostic suggestions improve accuracy of family physicians: a randomized controlled trial in Greece. *Fam Pract*. 2015;32:323-328. https://doi.org/10.1093/fampra/cmv012

How to cite this article: Ethier J-F, McGilchrist M, Barton A, et al. The TRANSFoRm project: Experience and lessons learned regarding functional and interoperability requirements to support primary care. *Learn Health Sys*. 2017;e10037. https://doi.org/10.1002/lrh2.10037