7. Cooper WW, Seiford LM, Tone K. *Data Envelopment Analysis: a Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2nd ed. New York, NY: Springer; 2007.
8. Hollingsworth B, Peacock SJ. *Efficiency Measurement in Health and Health Care*. Oxon: Routledge; 2008.
9. American Medical Association. *Medicare RBRVS 2013: the Physician's Guide*. Chicago, IL: American Medical Association; 2013.
10. Macario A. What does one minute of operating room time cost? *J Clin Anesth*. 2010;22:233–236.
11. Ozcan YA. *Health Care Benchmarking and Performance Evaluation: an Assessment Using Data Envelopment Analysis*. New York, NY: Springer; 2008.

Author Biographies

Yoshinori Nakata, MD, MBA, has been a professor of healthcare management at Teikyo University Graduate School of Public Health, Tokyo, Japan, since 2011. He received his BA in economics, MD, and PhD in medical science from the University of Tokyo and his MBA from Yale University. A former professor and chair of anesthesiology and vice-president at Teikyo University School of Medicine Ichihara Hospital, he focuses on clinical aspects of health economics and management. He is especially interested in efficiency and productivity of surgery and operating room management.

Yuichi Watanabe, MS, is an MPH candidate at Teikyo University Graduate School of Public Health, Tokyo, Japan. He received his M.S. degree in economics from University of North Carolina at Chapel Hill. His research interest focuses on clinical application of microeconomic models.

Hiroshi Otake, MD, MBA, is a professor in and chair of the Department of Anesthesia, Showa University School of Medicine, Tokyo, Japan. He received his MD from the University of Tokyo and his MBA from the University of Chicago. He was formerly a management consultant at McKinsey & Co. His research interest focuses on operating room management at a large university hospital in urban settings under a universal insurance coverage system.

Toshihito Nakamura, MD, PhD, is an associate professor in the Department of Health Policy for Aged Society, Chiba University Hospital, Chiba, Japan. He was formerly a practicing surgeon at numerous hospitals. His research interest focuses on efficiency and productivity of surgeons at various settings.

Giichiro Oiso, MD, JD, is a professor of medical law at the Hamamatsu University School of Medicine, Hamamatsu, Japan. He received his MD from

Nippon Medical University and his JD from Waseda University. Formerly a practicing gastroenterologist, he is an expert in legal aspects of medical practice in Japan.

Tomohiro Sawa, MD, PhD, is a professor of medical informatics, Teikyo University Medical Information and System Research Center, Tokyo, Japan. He is an anesthesiologist and an expert in medical informatics and clinical data mining. He designed electronic medical record systems for numerous hospitals in Japan.

Original Paper

A Querying Method over RDF-ized Health Level Seven v2.5 Messages Using Life Science Knowledge Resources

Yoshimasa Kawazoe[1*], MD, PhD; Takeshi Imai[2*], PhD; Kazuhiko Ohe[1,3*], MD, PhD

[1]Department of Healthcare Information Management, The University of Tokyo Hospital, Tokyo, Japan

[2]Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[3]Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[*]all authors contributed equally

Corresponding Author:
Yoshimasa Kawazoe, MD, PhD
Department of Healthcare Information Management
The University of Tokyo Hospital
7-3-1, Hongo, Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 5800 8685
Fax: 81 3 3813 7238
Email: kawazoe@hcc.h.u-tokyo.ac.jp

Abstract

Background: Health level seven version 2.5 (HL7 v2.5) is a widespread messaging standard for information exchange between clinical information systems. By applying Semantic Web technologies for handling HL7 v2.5 messages, it is possible to integrate large-scale clinical data with life science knowledge resources.

Objective: Showing feasibility of a querying method over large-scale resource description framework (RDF)-ized HL7 v2.5 messages using publicly available drug databases.

Methods: We developed a method to convert HL7 v2.5 messages into the RDF. We also converted five kinds of drug databases into RDF and provided explicit links between the corresponding items among them. With those linked drug data, we then developed a method for query expansion to search the clinical data using semantic information on drug classes along with four types of temporal patterns. For evaluation purpose, medication orders and laboratory test results for a 3-year period at the University of Tokyo Hospital were used, and the query execution times were measured.

Results: Approximately 650 million RDF triples for medication orders and 790 million RDF triples for laboratory test results were converted. Taking three types of query in use cases for detecting adverse events of drugs as an example, we confirmed these queries were represented in SPARQL Protocol and RDF Query Language (SPARQL) using our methods and comparison with conventional query expressions were performed. The measurement results confirm that the query time is feasible and increases logarithmically or linearly with the amount of data and without diverging.

Conclusions: The proposed methods enabled query expressions that separate knowledge resources and clinical data, thereby suggesting the feasibility for improving the usability of clinical data by enhancing the knowledge resources. We also demonstrate that when HL7 v2.5 messages are automatically converted into RDF, searches are still possible through SPARQL without modifying the structure. As such, the proposed method benefits not only our hospitals, but also numerous hospitals that handle HL7 v2.5 messages. Our approach highlights a potential of large-scale data federation techniques to retrieve clinical information, which could be applied as applications of clinical intelligence to improve clinical practices, such as adverse drug event monitoring and cohort selection for a clinical study as well as discovering new knowledge from clinical information.

(JMIR Med Inform 2016;4(2):e12) doi:10.2196/medinform.5275

KEYWORDS

electronic health records; health level seven; information storage and retrieval; Semantic Web; linked open data

Introduction

Clinical Data Searches Through Knowledge Level Queries

While secondary use of electronic medical records (EMRs) are widely expected [1,2], medical data in general do not contain adequate amounts of information or knowledge in their original format, making it difficult to retrieve the desired data based on the knowledge in the clinical domain. For example, when we try to screen patients with medication history of "renin angiotensin inhibitors" as possible candidates for a clinical study, it is common for us to prepare a list of drug codes for such drug classes and query a database with the prepared list. If such a query is performed simply using an expression such as "drugs classified as renin angiotensin inhibitors," it will facilitate our use of the database. As a similar example, when we try to screen patients with medication history of "drugs that cause leucopenia," rather than having to list in a query hundreds of codes for drugs showing the adverse events, if drugs that cause leukopenia are identified using external knowledge resources, and if a search is performed over medication data based on the identified drugs, it would facilitate the research use of EMRs.

Clinical Data Searches Using Life Science Knowledge Resources

The Linked Open Data project [3] is an attempt to facilitate data usage via the Internet by making data available in a standard format based on the resource description framework (RDF). In the field of life science, attempts are being made to further increase the value of data sets by linking and integrating them as Linked Data. The Bio2RDF project [4] aims at linking and using over 20 types of data sets including the Kyoto Encyclopedia of Genes and Genomes (KEGG) [5,6], the Open Biological and Biomedical Ontologies [7], the Universal Protein Resource [8], and the Gene Ontology [9]. In addition, the National Bioscience Database Center and the Database Center for Life Science in Japan act as primary driving forces and conduct various activities to promote the use of life science data resources and abroad as Linked Data [10,11].

Applying RDF to build clinical databases for secondary use facilitates integration of external knowledge resources expressed in RDF. Teodoro et al. [12] developed a Web-based antimicrobial resistance monitoring system that uses a Semantic Web-based approach to promote the integration of heterogeneous data sources. Assélé et al. [13] developed a framework to perform SPARQL Protocol and RDF Query Language (SPARQL) queries on clinical databases to obtain results about antibiotic resistance and compared their approach with existing business intelligence approaches in terms of usability and functionality. Riazanov et al. [14] developed an ontology for the clinical domain and reported that SPARQL queries can be expressed and executed in an ad hoc manner by mapping the developed clinical domain ontology and clinical data. Pathak et al. [15,16] used publicly available life science data resources as Linked Data and searched over EMR databases integrated with these resources through SPARQL federation queries. The above studies attempt to improve search usability and functionality by applying Semantic Web technologies to supplement information lacking in the clinical data with knowledge from external resources. However, these studies dealt with only institution-specific EMR databases, and it is not easy to apply their methods at other hospitals because schemas of EMR databases generally differ between hospitals; thus, the RDF data structures constructed from these schemas also differ. To avoid these problems and make these technologies widely available, we use health level seven version 2.5 (HL7 v2.5) [17] messages as clinical data. HL7 v2.5 is a messaging standard for information exchange between clinical information systems and the most widely implemented standard for health care in the world. It specifies a number of standards, guidelines, and methodologies by which various clinical information systems can communicate with each other. HL7 messages, although not comprehensive, contain several important types of data for clinical research, such as patient demographics and diagnostic disease.

RDF for Developing Clinical Databases

Applying RDF in developing clinical databases for secondary use provides the following benefits. First, because the RDF data structure is simple, they can express highly heterogeneous data sets including clinical data, disease concepts, drugs, clinical tests, and genome information using a single data model, making it possible to integrate and handle them in a coherent manner. Second, the inference mechanism supports data sets with hierarchical relationships, such as those containing disease and drug information, through an RDF schema (RDFS) [18] vocabularies. With the relational databases typically used in clinical databases, special measures are required to express the hierarchical structures that exist in data. With RDF, however, this can be accomplished simply by adding the rdfs:subclassOf relationship between the resources. Third, RDF identifies resources through uniform resource identifiers (URIs); therefore, data can be shared via HTTP between different network locations. SPARQL federation query integrates publicly available data sets and allows different network locations to refer to and search over these integrated data sets, maintaining high confidentiality of EMRs. This is expected to be useful when developing clinical databases.

Aim of the Study

Using RDF as the format for HL7 messages, it is possible to integrate large-scale clinical data and life science knowledge resources. In this study, we implement the following measures to verify this approach. We develop a method for converting HL7 messages into RDF data. Noting that publicly available drug databases constitute useful resources for query expansion in clinical data searches, we show how SPARQL describes adverse drug events (ADEs) and perform searches using such SPARQL expressions. We also examine the search performance and discuss the applicability of the proposed approach to the searches over large-scale data.

Methods

RDF and SPARQL

Semantic Web technologies use simple data structures to integrate and use data on a Web-level scale. RDF is the most basic technology for standardizing data expressions, and it consists of a set of URI references (U), a set of blank nodes (B), and a set of literals (L). An RDF triple is a tuple of three elements, that is, a subject (s), a predicate (p), and an object (o), that satisfy $s \in (U \cup B)$, $p \in U$, and $o \in (U \cup B \cup L)$, respectively. The RDF graph is a directed graph of RDF triples. A data schema in RDF is defined by the vocabulary and semantics of the RDFS. The RDFS is a set of vocabulary and inference rules defined for the vocabulary, and the RDF processor executes these inference rules to derive new RDF triples, which are then added to the RDF graph. For example, rdfs:subclassOf is a vocabulary that defines the class–subclass relationship, and this vocabulary is defined by two rules (ie, a transitive rule and a rule to express a lower class instance being also an upper class instance). Through this inference rule, a search over a lower class and its instances becomes possible by using a higher level abstraction as the search terminology.

SPARQL is an RDF query language. It describes, in the query condition, variables of a pattern to match and their values to use for filtering and extracts the subgraphs that match the given pattern from an entire RDF graph so that the corresponding values of the specified variables are obtained. Filtering of values is performed by using FILTER keywords and by computing a boolean value using the values bound to the variables. Examples of typical functions include a function that performs matching of text strings in their regular expressions and functions that perform logic operations. One beneficial feature of SPARQL is that it can handle multiple RDF graphs as a single graph. SPARQL 1.1 further enhances this feature, making it possible for a single federated query [19] to inquire multiple RDF graphs at different network locations. A federated query expression first designates the SPARQL endpoint with a SERVICE keyword and then describes variables of a pattern to match, similar to a regular SPARQL query, in a clause that follows the endpoint. Consequently, using variables, a federation query can describe a query that can search local or remote RDF graphs.

SS-MIX2: HL7 Message-Based Clinical Data Storage in Japan

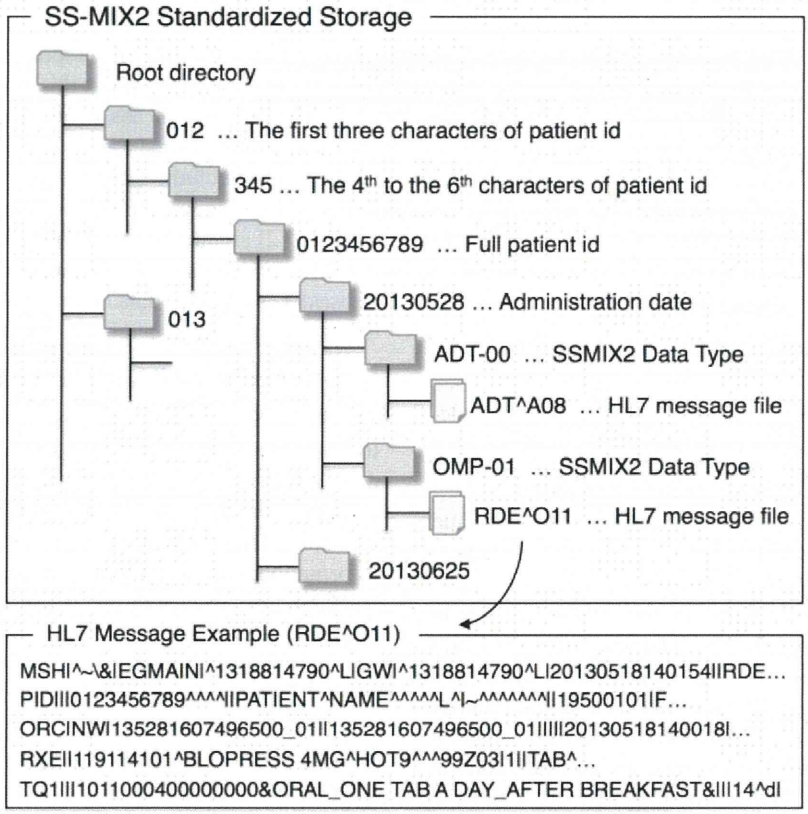
We used HL7 messages stored in the Standardized Structured Medical Record Information Exchange version 2 (SS-MIX2) that has been developed to facilitate secondary use of EMRs as a Ministry project in Japan [20,21]. SS-MIX2 defines the specification of a container for storing EMRs, and the main body of the EMRs is the HL7 v2.5 message. It consists of the standardized storage and the annex storage. The standardized storage contains structured clinical data in the form of an HL7 v2.5 message, such as patient demographics, diagnostic disease, medication orders, laboratory test results, and several kinds of examination orders. The annex storage contains nonstructured clinical data, such as clinical reports, examination reports, and imaging data in arbitrary format. Earlier than the development of the SS-MIX2, standardized terminology for drugs, laboratory tests, procedures, and diagnostic disease has also been developed by the Medical Information System Development Center (MEDIS-DC) [22], and exchange rules for clinical information to be conformed with HL7 have also been developed by the Japanese Association of Healthcare Information System Industry [23]. In 2011, the Ministry of Health, Labor, and Welfare adopted these terminologies and exchange rules as the standard specifications for the health and medical care information field, thereby facilitating the development of standardized medical information systems. Against this background, as of July, 2015, the SS-MIX2 storage has been deployed at 518 hospitals in various regions of Japan [24]. Examples of SS-MIX2 storage applications include (1) an intermediate storage linking multivendor systems and electronic medical record/order entry systems, (2) an intermediate storage for linking regional health care systems, (3) a backup data storage for use in the event of a disaster, and (4) a data source for postmarketing survey of drugs and clinical research.

Structure of SS-MIX2 Storage and HL7 Message

The SS-MIX2 stores HL7 messages below the ordinary directory trees. Under the root directory, patient identifier, administration date, and SS-MIX2 data type are hierarchically located, and corresponding HL7 messages are placed under the bottom directory. The SS-MIX2 data types identify types of clinical information, such as patient demographics, medication orders, and laboratory test results, and these data types are semantically mapped on HL7 message types. For example, HL7 message types to update or delete patient demographics are ADT^A08 and ADT^A23, respectively. SS-MIX2 uses a single data type (ie, ADT-00, for these two HL7 message types). In an HL7 message, each line is called a segment and contains a specific category of information, such as patient identification (PID), order-related information (ORC), and pharmacy (RXE). Each segment consists of a field delimited by a pipe symbol, and the field consists of a field's element delimited by a hat symbol. For example, a patient identifier is located in the third field of the PID segment and a drug code is located in the first field's element in the second field of the RXE segment. Two or more segments may be organized as a logical unit called a segment group, which might or might not repeat. The boundary of the segment group is not identical in a standard form of the HL7 message itself, but it appears in an extensible markup language (XML)-encoded HL7 message described in the next section. Some fields or a field's element may contain a code defined by a certain terminology. In the SS-MIX2, terminologies are used, such as MEDIS DRUG [22] for drugs, JLAC10 [25] for laboratory tests and International Classification of Diseases, and 10th Revision (ICD10) for diagnostic diseases, which are all provided by MEDIS-DC as a nationwide standard. Although these terminologies are unique to Japan except for ICD10, the terminology for drugs can be mapped on the Anatomical Therapeutic Chemical Classification System (ATC) and United States Pharmacopeia (USP) [26] using intermediate resources such as KEGG. This mapping information becomes the key-point to supply an HL7 message with external knowledge recourses by matching a code in the message to a class represented in the recourses. Figure 1 shows examples of an SS-MIX2 storage structure and an HL7 message.

This example HL7 message (RDE^O11) contains information on a medication order for a patient identified by 0123456789 administered on May 28, 2013. The message contains the following segments: message header (MSH), patient identification (PID), order-related information (ORC), pharmacy encoded (RXE), and timing and quantity (TQ1).

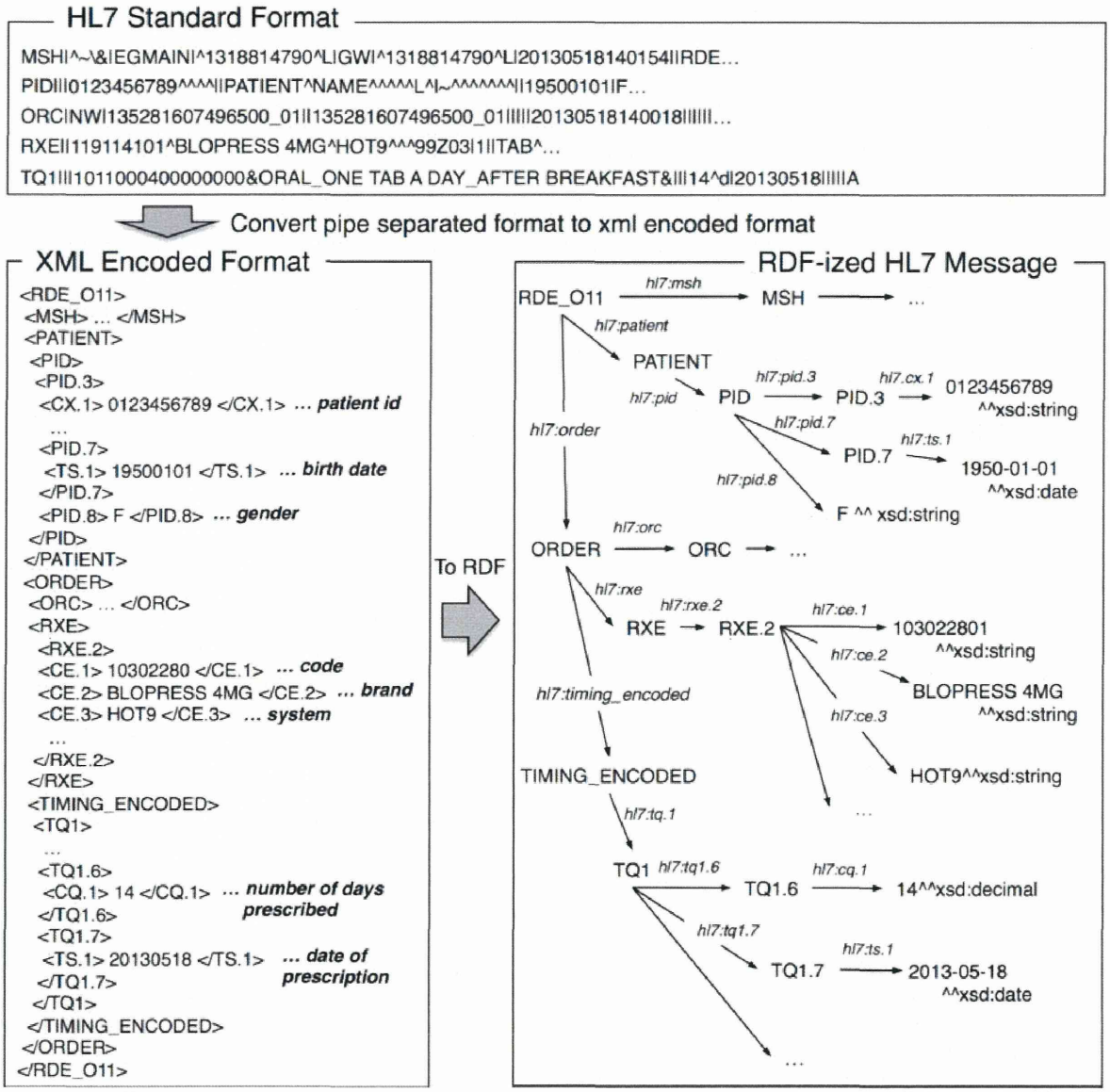
Figure 1. Examples of an SS-MIX2 storage structure and an HL7 message. This example HL7 message (RDE^O11) contains information on a medication order for a patient identified by 0123456789 administered on May 28, 2013. The message contains the following segments: message header (MSH), patient identification (PID), order-related information (ORC), pharmacy encoded (RXE), and timing and quantity (TQ1).



Converting HL7 Messages Into RDF Data

In the standard form of an HL7 message, metadata for fields or a field's elements are not included. For example, the patient's date of birth is located in the seventh field of the PID segment, although, the message itself does not contain the information. If the name of an RDF resource is determined based on its metadata, HL7 messages are efficiently converted to RDF data. Prasser et al. [27] proposed a method that uses a generic Java-based parser provided by the HL7 Application Programming Interface (HAPI), and that uses the Java class and method names as metadata, traversing Java objects, to convert an HL7 message to RDF data [28]. We also use the HAPI to parse a standard form of the HL7 message, although, we first encode the HL7 message to a form of XML that is also defined in the HL7 specifications. In an XML-encoded HL7 message, segments and segment groups are given in hierarchical XML elements. For example, an XML form of an HL7 message for a medication order starts with an <RDE_O11> tag that describes the type of HL7 message, followed by a tag that describes the segment of a message header <MSH> and segment groups of patient information <PATIENT> and order information <ORDER>. In the segment groups, the corresponding segments are included, such as the PID segment in the PATIENT segment group or the ORC and RXE segments in the ORDER segment group. Similarly, each segment contains a tag for each of its fields to describe either the field or the field's element, such as a time stamp <TS> or a coded character string <CWR>, and text data is marked up with these tags. We then applied a generic method of transformation from XML to RDF [29], in which an RDF resource is generated using the element name of the XML as the name of the resource, creating a subject-predicate-object triple by traversing the hierarchical structure, and mapping the text content to an RDF literal. Note that the mapping needs to be determined in advance because the XML-encoded HL7 message does not contain the data type of the text content. Thus, we sought to map the numerical type of the text content to *xsd:decimal,* the date type to *xsd:date*, the timestamp type to *xsd:dateTime*, and all other types to *xsd:string*. In comparison with the previously mentioned method, there is an advantage to be able to use the names of the segment or field defined by the HL7 specifications, which is not modified depending on the implementation of the Java class and method names. Figure 2 shows a medication order in the standard form of an HL7 message, an XML-encoded HL7 message, and an RDF representation after conversion.

Figure 2. A medication order in the HL7 standard format, XML-encoded format, and after conversion to RDF.



URI Naming

To determine a URI of an RDF resource, we considered two requirements: (1) the name of the URI should preferably contain a structured path to facilitate the application's access to RDF resources [30], (2) the name of the URI should be generated uniquely from the available information for an HL7 message to avoid redundancy of referring to an RDF repository each time when determining it. To satisfy these requirements, we constructed the name of the URI by connecting a directory path to an HL7 message file, which is already unique in SS-MIX2 storage, with a path to an element in XML that is encoded from the HL7 message. Note that as several HL7 segment groups, such as ORDER and RESULT may appear multiple times in the same hierarchy layer in the XML, duplication of the path names should be avoided by counting how many times they appear in the path. As the HL7 message specifications define which segment groups may appear multiple times, the name of the URI can uniquely identify the deepest elements by considering the duplication. This naming method depends on SS-MIX2 in terms of using the directory path to an HL7 message, although, if only the path to an HL7 massage is uniquely determined, any other way can be applied. Figure 3 shows a portion of a serialized RDF representation of a medication order.

Depending on the purpose of use of the HL7 message, it may contain numerous redundant segments, fields and field's elements, and it may not be necessary to convert all content to RDF data. For example, a MSH segment that provides header information for communication between systems, as well as fields other than the patient identifier, date of birth, and gender in a PID segment, is not required in clinical research. Therefore, when converting to RDF, the amount of RDF data to generate is reduced by only using the segments and fields that are needed for the purpose.

Figure 3. Serialized RDF representation of a medication order in turtle format.

```
@prefix hl7v25: <http://hl7.org/v25#> .
@prefix ssmix2: <http://ssmix.org/v2#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1>
  hl7v25:PATIENT <http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/PATIENT> ;
  hl7v25:ORDER   <http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1> .

<http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1>
  hl7v25:ORC <http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1/ORC> ;
  hl7v25:RXE <http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1/RXE> .

<http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1/RXE>
  hl7v25:RXE.2 <http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1/RXE/RXE.2> .

<http://m.u-tokyo.ac.jp/0123456789/20130518/OMP-01/.../RDE_O11/1/ORDER/1/RXE/RXE.2>
  hl7v25:CE.1 "103022801"^^xsd:string ;
  hl7v25:CE.2 "BLOPRESS 4MG"^^xsd:string ;
  hl7v25:CE.3 "HOT9"^^xsd:string .
```

Query Expansion Using Linked Drug Data

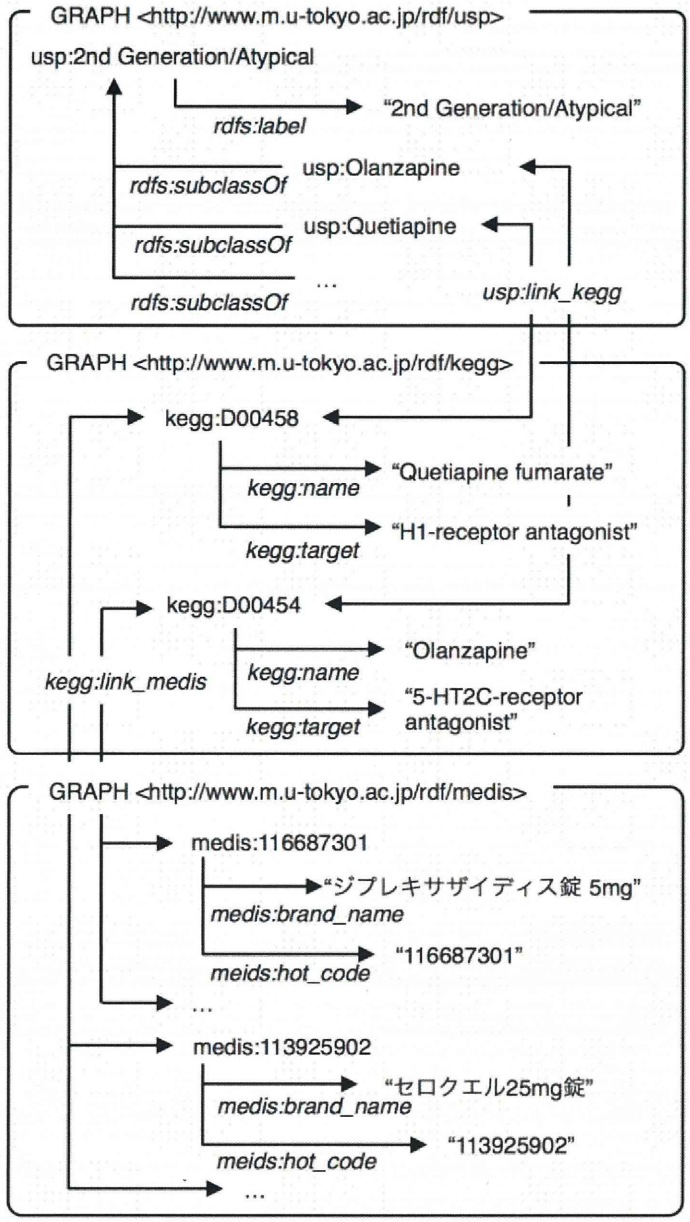
If a type of drug is identified by its detailed information, it is useful for a query to search for ADEs of a drug. By converting drug databases to Linked Data, it is possible to identify drugs through expressions that use their detailed information and to resolve the identified drugs to their codes used in the HL7 message. For example, a medication order search for atypical antipsychotic drugs that have an inhibitory effect on the serotonin 2C (5HT2C) receptor or the histamine H1 (H1) receptor consists of the following steps: (1) use the USP to identify drugs classified as atypical antipsychotic drugs, (2) use a link between the USP and KEGG to identify corresponding KEGG drugs. Then, narrow down the list to those drugs that have an inhibitory effect on the 5HT2C receptor or the H1 receptor, (3) use a link between the KEGG and MEDIS DRUG to identify corresponding drugs on the MEDIS DRUG and to identify the codes of the drugs to use in the HL7 message, and (4) Use the identified drug codes to search for a medication order over HL7 messages. Figure 4 illustrates relationships between USP, KEGG, and MEDIS DRUG used in this search.

To enable this method, we converted publicly available drug databases into RDF and provided explicit links among the corresponding items to obtain linked data. Because there were no data sources publicly available in RDF format, we converted each source individually to RDF. We got the sources of ATC, USP, and KEGG from a website of the KEGG and made the explicit links based on the information obtained from the KEGG. We used rdfs:subclassOf to describe the higher and lower level relationship in the ATC and USP, and inference was executed and materialized in advance. We also got the sources of SIDER 2 (SIDe Effect Resource) [31] and MEDIS DRUG from each website. In the SIDER 2 dataset, drug classes are coded in STITCH [32] identifiers and names of ADEs are coded in MedDRA along with upper and lower bound of the frequency. The information to link between the SIDER 2 and ATC were obtained from website of STITCH. We used the MEDIS DRUG to match the drug concept in the KEGG to the drug code used in the HL7 message, and the information to link between them were obtained from the the KEGG source. This linked drug data set is hereafter referred to as Linked Drug Data. A summary of the Linked Drug Data is shown in Table 1. The Linked Drug Data is available from our project repository [33].

Table 1. A summary of the linked drug data.

Original drug databases	Descriptions	Link to the other databases	Number of drug classes (triples)
Anatomical Therapeutic Chemical Classification System (ATC)	A drug classification system developed by World Health Organization. It divides drugs into different classes according to the organ or system on which they act or their therapeutic and chemical characteristics, such as antihypertensives and the cardiovascular system. In converting to RDF, we used rdfs:subclassOf to represent the hierarchical relationships and added links to the drug classes of KEGG and SIDER 2 at the chemical substance subgroup level.	KEGG, SIDER 2	5770 (48,504)
United States Pharmacopeia Classification (USP)	A drug classification system developed by the US Pharmacopeial Convention. It contains approximately 50 categories, which are typically based on diseases or symptoms that drugs are used to treat, such as pain and psychosis. In the same way as ATC, the hierarchical relationships were represented by rdfs:subclassOf.	KEGG	1459 (7567)
SIDER 2	A resource that contains ADEs and their frequency, which are extracted from package inserts and publicly available documents. The drugs are coded by STITCH compound identifiers, and the ADEs are described in the preferred terms of MedDra.	ATC	997 (7,848,862)
KEGG	A resource that consolidates drug data from Japan, the Unites States, and Europe. It organizes drug data based on their chemical structures and ingredients and adds information on their molecular interactions including chemical drug targets and metabolic enzymes. Many entries also include their mapping to other drug databases, and we use the mapping information to establish links to ATC, USP, and MEDIS DRUG.	ATC, USP, MEDIS DRUG	5780 (109,976)
MEDIS DRUG	A standard drug terminology that maps various drug terminologies used in Japan. We used MEDIS DRUG to match the drug code in KEGG to the drug code used in the HL7 message.	KEGG	26,126 (387,319)

Figure 4. Relationships between USP, KEGG, and MEDIS DRUG used in search for atypical antipsychotic drugs that have an inhibitory effect on the 5HT2C receptor or the H1 receptor.



Temporal Patterns to Determine Adverse Drug Events

To identify adverse events, a query condition needs to describe the temporal relationship between the administration of a drug and the adverse events that were assumed to be caused. We classify the temporal relationships into the following four types of basic temporal patterns and explain query expressions using these patterns to identify adverse events.

Temporal Pattern 1: Searching for all Medication Orders

This pattern is used to retrieve all medication orders of a specific drug without considering their temporal relationships with other events. This is the most basic pattern of clinical data searches.

Temporal Pattern 2: Searching for Adverse Events During Each Medication Period

This pattern estimates the medication period as beginning on the day that a drug medication order was issued and continuing for the number of days prescribed, and it searches for the adverse events during the estimated medication period. Although the medication period estimated in this pattern is likely to be close to the actual drug administration period, irregular medication orders, when issued, could make a period when a drug has been administered appear as it had not been, and the estimated medication period could erroneously exclude such periods. Consequently, it is possible to overlook adverse events during such excluded periods.