General Legal Entity Identifier Ontology

Robert TRYPUZ [a,b,1], Dominik KUZIŃSKI [a] and Mirek SOPEK [a]

[a] *MakoLab S.A.*

[b] *The Faculty of Philosophy, The John Paul II Catholic University of Lublin*

Abstract. This paper describes some advantages of RDF/OWL representation of Legal Entity Identifier (LEI) data in contrast to the current XML representation. We propose an OWL specification of LEI data named General Legal Entity Identifier Ontology (GLEIO) that is not only compliant with the current XML representation but also allows for representation of changes of LEI and related LEI reference data over time. The paper also presents a web application that uses GLEIO to serve information about LEI and legal entities.

Keywords. Legal Entity Identifier, LEI, General Legal Entity Identifier Ontology

1. Introduction

The Legal Entity Identifier (LEI) is a code based on the ISO 17442. It is assigned to a legal entity to unambiguously and uniquely identify it during financial transactions. The Common Data File Format (henceforward: CDF Format) is a standard by which an LEI and a legal entity reference data are reported by local operating units (LOUs). LOUs' reports are expressed in XML files defined by XML schema compliant with the CDF Format. The schema consists of LEI File Header and LEI Data Record (see figure 1). This paper describes some advantages of RDF/OWL representation of global LEI data in contrast to the current XML representation (section 2.1). We propose an OWL specification of LEI data named General Legal Entity Identifier Ontology (GLEIO) that is not only compliant with CDF Format but also allows for representation

Figure 1.: XML schema for the Common Data File Format

of changes of LEI and related LEI reference data over time (section 2.2). We focus on benefits that can be gained by using semantic representation. The paper also presents a web portal that uses GLEIO to serve information about LEI and legal entities (section 3).

---

[1] Corresponding Author: Robert Trypuz, MakoLab, ul. Rzgowska 30, 93-172 Łódź, Poland; E-mail: robert.trypuz@makolab.pl.

2. GLEIO – an RDF/OWL model of the CDF Format

2.1. Why use RDF/OWL to represent and share LEIs and related LEI reference data

In the White Paper[2] presented to Global LEI Foundation (GLEIF)[3] we pointed out drawbacks and limitations of the current XML representation of LEI data. Among them are: lack of precise semantics, difficult extensibility, lack of global persistent identifiers and lack of inference. We also showed that OWL ontologies solve all those challenges.

Ontologies allow to classify entities and have enough expressive power to introduce constraints on the level of things. For example, we can formally express that "Each global LEI identifies exactly one legal entity". Semantic relations can help in linking collected information. For instance, in the CDF Format we find references to Countries and Subdivisions. In a semantic representation they could be linked by parthood relation (e.g. Nordrhein-Westfalen *is part of* Germany). As a result the data may be queried in a new way. For instance, having expressed in a semantic LEI data representation of parthood relations between Germany and its states, it is easy to issue a request in the SPARQL query language: "Find a number of LEI codes for each German state."

Since ontologies link things (rather than describing the structure of documents), they are flexible enough to allow for extensions. New data based on terminology from external resources can be added to an RDF graph. It is possible because RDF/OWL ontologies use URIs, which are universal global unique identifiers. By using URIs, browsers and other applications can retrieve the data of the resource identified by that URI.

Browsers render the retrieved representation so that it can be perceived by a user. Other applications can use the content negotiation mechanism to access data in the intended format. URIs of different RDF data can be easily linked. The large-scale, influential initiative known as "Linked Open Data" provides methodology of publishing and interlinking data. By linking data we can reuse external resources which we need for our representation (e.g. ontology of ISO 3166 codes for countries). But linked data also allows for gaining new information and formulating interesting queries. For instance, having semantic representation LEI data and connecting its class of countries with `http://www.geonames.org/countries/` automatically gives us information about capital, area or population of a country. So for instance the following request can be formed: "Find a number of legal entities that possess LEI codes for each Country. Order answers by area of a country."

Ontologies also allow for inference. Reasoners for OWL ontologies can check consistency of ontology, i.e. whether the set of ontological restrictions (axioms) has a model. One can also verify whether the facts explicitly expressed in an OWL ontology do not break the restrictions.

2.2. General Legal Entity Identifier Ontology

GLEIO[4] is an OWL ontology compliant with the CDF Format. Its canonical URI is `http://lei.info/gleio/`. Its process of creation involved careful translation of the

---

[2]`https://makolab.com/en/innovation/gleio`

[3]`https://www.gleif.org`

[4]GLEIO has SHIQ(D) expressivity, i.e., it allows complement of any concept, concept intersection, universal restrictions, unlimited existential quantification, role (binary predicate) hierarchy, transitive and inverse relations, qualified cardinality restrictions and use of datatype properties, data values and data types. Hyperlinks to

format into OWL structure, i.e. introducing: *classes of objects* (e.g. Legal Entity, Business Register); *individuals* (e.g. ISO 3166 codes for countries and subdivisions, particular values of the LEI registration status such as "Annulled" or "Duplicate"); *object properties* that relate objects with objects (e.g. hasAssociatedEntity linking concrete legal entities with other legal entities); *data properties* relating objects with literals such as strings, integers or dates (e.g. hasGLEICode linking general LEI registration with its code); *constraints* that specify the conceptualization expressed by CDF Format (e.g. that general LEI registration has at most one LEI code and that each general LEI code identifies exactly one legal entity). In the end each element from CDF Format found its counterpart in GLEIO, so it can be concluded that GLEIO accurately represent its domain (fidelity). To improve intelligibility, each GLEIO class, individual and property that corresponds to an element in CDF Format was also described by annotations: rdfs:label, skos:definition, rdfs:comment and rdfs:isDefinedBy. The last two annotations were used to point out an element from CDF Format that corresponds to a given GLEIO resource. We have also used skos:closeMatch and skos:narrowMatch to link classes and properties of GLEIO with Financial Industry Business Ontology (FIBO) [Bennett, 2013].

Requirements for representing change   Temporal aspect plays important role in the CDF Format. A general LEI registration status can change in time. The same with data describing legal entities, e.g., the headquarters address of a legal entity can change. GLEIF publishes daily an updated concatenated file which includes all LEIs issued globally and related LEIs reference data. So everyday we receive a fresh, up-to-date picture but it is unknown what have changed. In GLEIO we decided to allow for explicit representation of change. Following [Fine, 1999,Moltmann, 2016] we divided all entities into[5]: *variable entities*, i.e. the entities that have different manifestations at different times, and *non-variable entities* that do not have different manifestations as different objects at different times. Among non-variable entities are *manifestations of variable entities*. Each such manifestation has its time stamp (being its beginning of existence). The following assumptions introduced in [Moltmann, 2016] have been adopted in GLEIO:

A1 Each variable entity has its last (or current if the entity exists) manifestation.

A2 If a variable entity is manifested at time t by an entity manifestation m, then it exists at t, takes a geographical location of m and has all properties possessed by m in t.

A3 If all manifestations of a variable entity have some property, then the variable entity possesses the property necessarily (technically speaking the property can be directly attributed to the variable entity).

A1 is an axiom of GLEIO. **A2** and **A3** are rules implemented on the level of software and are used when our application (described in section 3 below) serves a current data about a variable object (e.g. a legal entity or an LEI) – it takes properties of its current manifestation and the necessary properties of the variable object.

DL axioms of GLEIO, its the turtle serialization, OWL Doc representation and class taxonomy picture can be found here: `https://makolab.com/en/innovation/gleio`.

[5]It is a topic for a separate paper to discuss why we have finally chosen this representation and not reification or 4D approaches compared and discussed in the OWL context e.g. in [Welty and Fikes, 2006].

This representation allows for change tracking across time stamps. Every day our LEI application reads a complete concatenated XML file and creates new manifestation of a legal entity or global LEI, if a change was found (e.g., LEI status was changed or the headquarters address of the legal entity changed). Manifestations are linked by temporal precedence relation "hasPredecessorEntityManifestation" (see figure 2). The changes in LEI registry is also represented in

Figure 2.: Variable entity and its manifestations

GLEIO by manifestations. For instance, the status of LEI registration can change from "Issued" to "Lapsed" (here we still have to do with the same LEI). In that case, we create a new manifestation of the LEI. In some other cases, for instance when the LEI status is "Duplicate", it will have its successor. In some sense it ceases to exist and stops being updated. Its successor becomes the 'issued' LEI.

We believe that Moltmann's idea of representing changes adopted in our ontology can also be successfully reused for other domains.
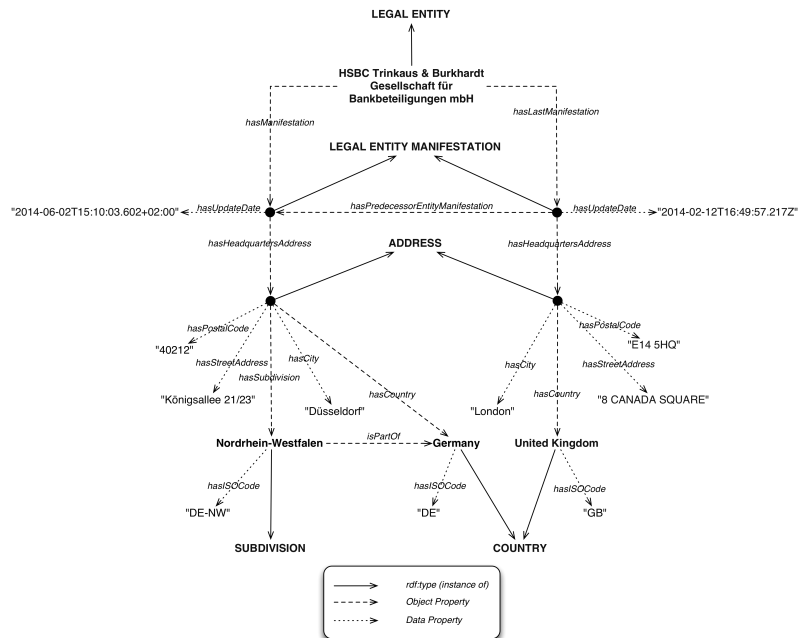
3. Web portal allowing for storing and displaying information about LEI in LOD standard (lei.info)

GLEIO is part of information system – a web portal – allowing for storing and displaying information about LEI. The web portal is complaint with Linked Data principles[6].

In figure 3 we present the components of the web portal. Input: GLEIF's XML concatenated files. A console application (Updater) runs on our server once a day. It downloads a fresh XML data file and checks for newly updated records. For each of them, the updater parses the data, transforms it to an RDF named graph and saves it in a Triplestore (we use AllegroGraph installed on a single virtual machine).

---

[6]https://www.w3.org/DesignIssues/LinkedData.html

The updater is also responsible for tracking changes. GLEIO provides a schema for the RDF graphs. Global LEI data stored in TripleStore can be queried by SPARQL endpoint[7]. The endpoint can be also used by external applications to access knowledge about LEI data. Romantic Web[8] is a MakoLab's open source library for .NET used as object-triple mapping tool. "lei.info" currently is a .NET application that allows searching legal entities only by LEI or by name and serves information in multiple formats (HTML or RDF serializations) by content negotiation and RDF embedded in HTML. In the future it will have a GUI allowing for tracking changes.

At the moment of writing the GLEIO knowledge base contained over 420K entity graphs, which gives over 24 million triples in total. There are many various types of competency questions that can be answered by our representation. Let us mention the following examples of questions: Query returning Legal Entities from France registered this year; Query for entities which renewal date is today; Query counting entities in LOUs; Query returning entities with HQ in VALENCIA, ES; Query for 5 newest LOUs ordered by registration date. SPARQL queries execute in times of milliseconds (1 - 500 depending on query). Response time achieved when requesting lei.info website for a single LEI record is around 250ms.
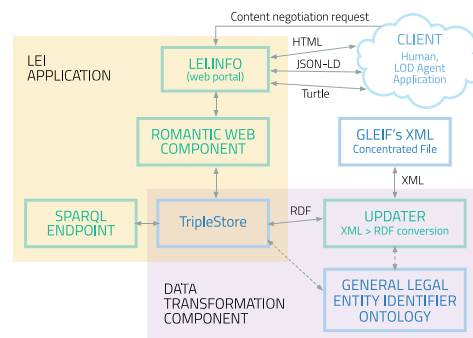

Figure 3.: Variable entity and its manifestations

4. Conclusion

We have presented the main ontological choices behind Global LEI Ontology. The ontology allows for representation of changes of LEI and related LEI reference data over time by introducing variable entities and their manifestations. We have also described the components of a web application using GLEIO.

References

[Bennett, 2013] Bennett, M. (2013). FIBO: Best practice in big data. *J Bank Regul*, 14(3-4):255–268.

[Fine, 1999] Fine, K. (1999). Things and their parts. *Midwest Studies in Philosophy*, 23(1):61–74.

[Moltmann, 2016] Moltmann, F. (2016). Variable Objects and Truth-Making. In Dumitru, M., editor, *Metaphysics, Meaning, and Modality. Themes from Kit Fine*, volume (forthcoming). Oxford Univerity Press.

[Welty and Fikes, 2006] Welty, C. A. and Fikes, R. (2006). A Reusable Ontology for Fluents in OWL. In Bennett, B. and Fellbaum, C., editors, *FOIS*, volume 150 of *Frontiers in Artificial Intelligence and Applications*, pages 226–236. IOS Press.

---

[7] http://lei.info/sparql

[8] http://romanticweb.net/