Requirements for Implementing Mappings Adaptation Systems

Julio Cesar dos Reis, Marcos da Silveira, Duy Dinh, Cédric Pruski, Chantal Reynaud-Delaître

Requirements for Implementing Mapping Adaptation Systems

Julio Cesar Dos Reis,
Public Research Centre Henri Tudor, Luxembourg
LRI, University of Paris-Sud XI, France
julio.dosreis@tudor.lu

Marcos Da Silveira, Duy Dinh, Cédric Pruski,
Public Research Centre Henri Tudor, Luxembourg
[marcos.dasilveira, duy.dinh, cedric.pruski]@tudor.lu

Chantal Reynaud-Delaître,
LRI, University of Paris-Sud XI, France
chantal.reynaud@lri.fr

***Abstract*— Ontologies, or more generally speaking, Knowledge Organization Systems (KOS) have been developed to support the correct interpretation of shared data in collaborative applications. The quantity and the heterogeneity of domain knowledge often require several KOS to describe their content. In order to assure unambiguous interpretation, overlapped concepts of different, but domain-related KOS are semantically connected via *mappings*. However, in various domains, KOS periodically evolve creating the necessity of reviewing the validity of associated mappings. The size of KOS remains a barrier for a manual review of mappings, and rather requires the support of (semi-) automatic solutions. This article describes our experiences in understanding how KOS evolution affects mappings. We present our lessons learned from various empirical experiments, and we derive primary elements and requirements for improving the automation of mapping maintenance.**

Keywords—Ontology Alignment; Mapping Maintenance; Mapping Adaptation; Ontology Evolution; Semantic Interoperability; Knowledge Engineering

I. INTRODUCTION

The growing number of integrated and collaborative Web environments demands knowledge-intensive software applications and semantic technologies to improve information retrieval, management, reasoning and sharing. Knowledge Organization Systems (KOS)[1] encompass all types of conceptual models for organizing knowledge (*e.g.*, taxonomies, thesauri and ontologies). They play a key role for Web-based collaborative applications, making the semantic of information explicit at different degrees of expressivity. However, the knowledge described by one KOS is often limited to specific sub-domains. Thus, Web-based collaborative systems need to rely on different KOS to cover the scope of their applications, resulting in semantic interoperability problems. This occurs for example, when the knowledge described by two or more KOS overlaps.

The necessity of semantically correlating overlapped computer-interpretable knowledge through mappings have been largely discussed in the literature [2]. However, few approaches consider the dynamic aspects of knowledge and their impact on mappings. Considering the increasing size of KOS (*e.g.,* SNOMED-CT, ICD, *etc*. for the biomedical domain) and the existing efforts to correlate data published on the Web (*e.g.,* Linked Open Data), automatic mapping adaptation has become a necessity to assure a certain level of reliability in dynamic environments. Mapping adaptation addresses the maintenance of mappings referring to the task of modifying existing mappings according to changes affecting the KOS, to keep them semantically valid and complete over time [3].

This article presents the lessons learned from the *DynaMO* project that investigates the impact of KOS evolution on the adaptation of associated mappings. We recall the outcomes of various experiments conducted to (i) identify KOS changes [4, 14], (ii) mapping changes [3], and (iii) the potential correlation between them [6]. On this ground, our contributions are twofold. First, we derive the principal requirements to develop an automatic system for mapping adaptation. Second, we describe and model the processes for managing the evolution of dynamic KOS and mappings between them.

We structure the remainder of this article as follows: Section II presents the related work; Section III reports on the methods adopted in our experiments; Section IV presents the achieved results while Section V gives a discussion with the lessons learned and some requirements for handling the mapping adaptation problem; Section VI wraps up the conclusions and provides an outlook on future work.

II. RELATED WORK

Mappings express the semantic interrelations of concepts issued from different KOS. We consider a mapping *m* as a triple (*s, t, r*) [2], where *s* refers to a concept from the source KOS_S, *t* is a concept from the target KOS_T, ($KOS_S \neq KOS_T$), and *r* stands for the type of semantic relation between *s* and *t* (*i.e., equivalence* ($\equiv$), *more general than* ($\geq$), *less general than* ($\leq$), and *partially matched* ($\approx$)).

Mappings can be manually or automatically created by knowledge engineers or through alignment tools respectively. However, if KOS evolve, modifications affecting one of the concepts (*s* or *t*) may invalidate the relation *r*. Examples of *Mapping Maintenance* tasks include the identification of invalid mappings, the interpretation of existing mappings and their adaptation.

Mapping maintenance refers to the process aiming to keep existing mappings in an updated state, reflecting changes affecting KOS entities at evolution time[4].

In this article, we focus on a part of the mapping maintenance which deals with the mapping adaptation problem. Some research work related to mapping adaptation has been previously proposed. Yu & Popa [5] studied the evolution of database schemas and they proposed to isolate the KOS entities that were modified between two versions of the same KOS, and to identify high level types of changes (*e.g.*, deletion of a table column, *etc.*). According to the change types, they applied a strategy of mapping adaptation. However, the accurate identification of these change types can be as complex as creating new mappings between KOS. Groß *et al.* [6] have extended this technique to the context of ontologies. Additionally, they proposed to complete the resulting set of mappings by using matching techniques over newly added ontology concepts. This work improves the mapping adaptation strategies using ontology complex changes. Groß *et al.* [7] empirically investigate the correlation between mapping evolution and ontology evolution in the domain of life sciences. They proposed a measure of impact ratio for the impact of ontology changes on mapping evolution with respect to three general ontology change types. Their work gives a good overview of the mapping adaptation needs, and motivates deeper investigations to refine the change types to automatically adapt mappings.

An *et al.* [8] consider KOS with two different models (database schemas and ontologies). They define mappings as links between columns of relational tables and properties of concepts in ontology. They represent semantic mappings as formulas relating tables in a schema with a subset of conjunctive formulas encoding a sub-tree in the ontology graph (s-tree). They characterize the validity of mappings through these formulas. The approach requests domain experts to declare the similarities between the old version and the new version of KOS, and to select the appropriate adaptation strategies.

Martins & Silva [9] proposed to adapt mappings impacted by changes in ontologies according to a pre-selected adaptation strategy. The authors assume that all mappings are instances of a pre-defined ontology (SBO) and changes should occur in mappings only if concepts of one of the connected ontologies are removed. Although promising, such approaches lack flexibility in terms of KOS changes considered, an interpretation of the definition of established mappings and the availability of complex behaviours for adapting mapping.

The findings of the aforementioned studies highlight the relevance of gaining a more in-depth understanding of how an advanced and fine-grained classification of KOS changes would affect mapping evolution. Evaluating the accuracy of detection methods for KOS changes and the correlations with mapping changes in real-world datasets may significantly contribute to the knowledge engineering research field. This article highlights findings from our experiments to understand KOS changes and evaluate their effects on mappings.

III. METHODS

In the experiments, we consider a KOS a set of Concepts, Attributes and Relationships. *Relationships* represent the ways in which concepts of the same KOS are related to one another while *Attributes* characterize *Concepts'* attributes. We use two different notations namely *relation* and *relationship* to distinguish a semantic link (*e.g.*, 'equivalent', 'less specific', 'more specific', *etc.*) between two concepts in a mapping from a semantic link (*e.g.*, 'is-a', 'part-of', 'related-to', *etc.*) between them within a KOS, respectively. Our experiments only consider the relationship "*IS-A*" in KOS.

We conducted the experiments with data collected from the biomedical domain. We made this choice because of various aspects: (1) the dynamic characteristics of the biomedical domain and (2)the availability of KOS (and mappings) created and validated by domain experts. We have analysed six different KOS: SNOMED CT (SCT), ICD-9-CM (ICD9), NCI Thesaurus (NCI), MedDRA, MeSH, and ICD-10-CM (ICD10) and their related mappings during a period of three years (from 2009 to 2012). These KOS are characterized by the absence of both concept's instances and by attributes' values containing only textual statements [10].

We defined the experiments to investigate real-world cases of the evolution of KOS and mappings. To this end, we organized the experiments aiming to (1) understand KOS evolution (*cf.* Section A), (2) understand mappings evolution (*cf.* Section B) and (3) understand the impact of KOS evolution on mappings (*cf.* Section C).

A. Understanding KOS evolution

Historically, users have paid minimal attention to lists of KOS changes and knowledge engineers have often not provided comprehensive and detailed documentation of changes in computer-interpretable formats [11]. Therefore, we defined a set of experiments that aim to observe the evolution of KOS and search for possible KOS "change operation patterns" and supplementary elements that can contribute to adequately adapt mappings. Change operations stand for sequences of changes semi-automatically implemented with the help of knowledge engineers to transform the current KOS version into a new one [12]. For this purpose, we compare two versions of the same KOS to identify the differences between them, and have characterized them as KOS change operations (KCOs). We used the *COnto-Diff* tool [13] to support this activity, which returns a set of KCOs. By observing these results, we can determine the most common and some rare evolution cases. We use this output to thoroughly investigate cases that result in complex mappings adaptation processes (*e.g.*, split of concepts) and to define further methods to identify KCOs [14].

The set of experiments implemented in this phase includes:

- **KOS Overview** [15]. A quantitative and qualitative analysis of KOS was performed, targeting the identification of basic correlations with mappings evolution. Our study consisted in observing all KOS changes that occurred in NCI from version 10.01 (March 2010) to version 11.09 (October 2011) and in identifying the amount of mappings between NCI and ICD9 (v. 2010) and between NCI and MedDRA (v. 12.0) affected by these changes.
- **KOS Change Pattern Operations** [14]. To refine the KOS overview experiments, we searched for specific and recurring KOS changes that we considered as change patterns. We described them according to KCOs

performed in KOS. KCOs include the revision, deletion and addition of KOS entities (*cf.* Table III), which directly influence the reliability of associated mappings. These experiments analysed three KOS (NCI, SCT, ICD9) during the period of 2009 to 2012.

B. *Understanding Mapping evolution*

As observed for KOS changes, a better insight into the concrete reasons for mapping changes would give more accurate understanding for implementing mapping adaptation. However, tracking of mapping changes was unavailable for the analysed sets of mappings. Hence, we implemented several experiments aiming to observe the evolution of mappings and to search for well-defined actions expressing mappings change and/or supplementary elements that might describe the way mappings evolve.

We performed the following set of experiments:

- **Mapping Evolution** [4, 14]. The goal is to observe and analyse the most recurrent behaviours of modifications in mappings.
- **Mapping Analysis** [16]. These experiments support the identification of a "sufficient" subset of attributes from a concept that is relevant for explaining existing mappings. In this investigation, we conducted a set of experiments to assess the quality of the identified attributes using two versions of two biomedical ontologies (SCT 2010 and 2012, and ICD9 2009 and 2011) and their mappings.
- **Mapping Adaptation Actions** [3].According to the outcomes of our observations on mapping evolution, we formally describe specific changes to apply in mappings as a set of mapping adaptation actions.

C. *Understanding the impact of KOS evolution on mappings*

This set of experiments aim to detect potential correlations between changes in KOS and changes in mappings. To this end, we assume that only modified concepts of KOS associated to a mapping can impact on the evolution of mappings (the associated one or other ones).

We performed the following experiments:

- **Map-KOS Correlation** [4]. Given the occurrence of a mapping adaptation action, we searched in KOS for any change involving the source concept (of the mapping) or at least one of its close neighbourhoods (*i.e.*, concepts that are directly linked to the source concept). For each type of mapping adaptation actions (*cf.* Table IV), we selected and analysed the most frequent set of KOS changes to understand and classify the conditions that lead (or not) to the observed changes in mapping. These experiments aim to identify changes in the KOS that may potentially lead to a change in mappings.
- **KOS-Map Correlation** [4]. To complement the first analysis (Map-KOS Correlation), we applied the same approach in the opposite sense. Given a KOS change operation, where one or several concepts are involved, we searched for changes in mappings that have one of the modified concepts as its source concept in mapping. For each KOS change operation, we selected a subset of mappings to manually perform a qualitative analysis.
- **Selection of Mapping Adaptation Strategy** [3]. The outcome from these two previous empirical analyses led to the specification of criteria (based on the KOS change operations) that can "activate" one mapping adaptation action or a sequence of these. Human intervention is required in cases the system cannot decide.

IV. RESULTS

Table I shows an overview of the datasets with the number of changed and unchanged mappings. D_i corresponds to the date of mapping releases (every six months, selected from January 2010). In total, we analysed more than 300.000 mappings.

TABLE I. ANALYSIS OF SCT-ICD9 MAPPINGS FROM JAN/2010 TO JUL/2011

Dataset	Total	Unchanged	Changed
D_1	100.875	100.394	481
D_2	101.254	100.281	973
D_3	102.601	101.076	1.525

Table II presents a deeper overview, where we consider four KOS entities (concept, attribute, relationship, and the neighbours) as well as the following mapping changes: *Unchanged* (U), *Added* (A), *Removed* (R), *Modification of target concept* (M_t), and *Modification of the relation* (M_r) or both *Modifications in target concept and relation* (M_{t_r}). The results of this analysis showed some expected situations. For example, the addition (removal) of mappings is mainly associated with addition (removal) of concepts. However, some cases do not follow this correlation. For instance, in D_3 (*cf.* A in Table II), only 66% of the added mappings result from the addition of newly added concepts. Another interesting observation indicates that changes in KOS are not always associated with changes in mappings. For instance, unchanged mappings (U) are associated with 0.06% of changed concepts. This indicates that changes in concepts do not always trigger mapping adaptation actions. This raises the following questions: What types of KOS changes can impact the mapping evolution? Which attributes of interrelated concepts impact the underlying mappings and why? How to identify these attributes?

TABLE II. CHANGES IN SCT ENTITIES CORRELATED WITH MAPPING CHANGES.

MCO	DS	Concept (%)	Attribute (%)	Relationship (%)	Neighbour (%)
U	D_1	0.06	0.25	6.75	1.71
	D_2	0.02	0.12	3.00	1.17
	D_3	0.03	0.11	3.34	1.38
A	D_1	100	100	100	12.93
	D_2	100	100	100	20.87
	D_3	66.18	66.18	66.5	28
R	D_1	98.22	98.18	100	39.2
	D_2	100	100	100	43.9
	D_3	83.64	83.64	100	47.27
$\mathbf{M_t}$	D_2	1.4	1.4	4.7	0.47
	D_3	0	0	10	5
$\mathbf{M_r}$	D_2	0.5	0.5	10	3.51
	D_3	0	0	0	0
$\mathbf{M_{t\_r}}$	D_2	1.2	1.2	4.7	2.38
	D_3	0	0	3.45	3.45

To address these questions, we investigated the refinement of KOS change operations (KCOs). By analysing two versions of the same KOS, we identified a set of operations used to perform the KOS changes. Table III [13] shows a non-exhaustive set of KCOs. Column "Nr. of changes" presents the absolute number of each change operation observed in our experiments with SCT and ICD9 [13]. A KCO with a frequency equal to zero was added based on our assumption that all KCOs observed have an inverse KCO. For instance, *merge* stands for the inverse KCO of *split*.

TABLE III. KOS CHANGE OPERATIONS

	Change Operation	Description	Nr. of changes SCT	Nr. of changes ICD9
Atomic	addC(c)	Add a new concept	-	-
	delC(c)	Delete an existing concept	-	-
	addA(a,c)	Add a new attribute	7.720	79
	delA(a,c)	Delete an existing attribute	4.003	25
	addR(r,c_1,c_2)	Add a relationship between c_1 and c_2	4.327	110
	delR(r,c_1,c_2)	Delete a relationship between c_1 and c_2	7.210	39
Complex	chgA(c,a,v)	Change the value of an attribute	950	155
	moveC(c,p_1,p_2)	Move a concept (and sub-tree) from the parent p_1 to p_2	7.140	31
	subst(c_1,c_2)	Replace c_1 by c_2	0	0
	merge(C_k,c_1)	Fusion of a set of concepts C_k into one concept c_1	0	0
	split(c_1,C_k)	Divide c_1 into a set of concepts C_k	134	45
	toObsolet(c)	Set c as obsolete	794	0
	addIn(c_1,p_j)	Add a concept in the middle of a sub-tree	1.348	26
	delIn(c_1,p_j)	Delete a concept from the middle of a sub-tree	0	0
	addLeaf(c_1,p_j)	Add a concept in the bottom of a sub-tree	3.649	140
	delLeaf(c_1,p_j)	Delete a concept from the bottom of a sub-tree	2	0
	rvkObsolet(c)	Revoke the status of obsolete	20	0
		Total	**37.297**	**650**

We also designed mapping adaptation actions (MAA) composed of actions used to perform changes in mappings. Table IV presents the formalization of MAA and Fig. 1 defines the notations used in Table IV. This also implies knowing that *CT(c)* refers to the context of a concept *c* composed of all super, sub and siblings concepts.

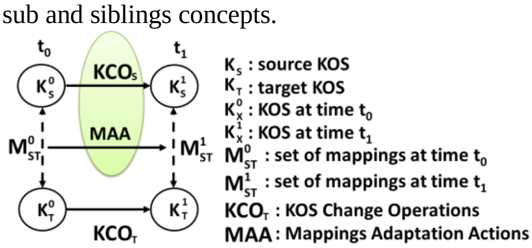

Fig. 1 Mapping adaptation based on KOS changes

The relation r_{st} between interrelated concepts is among the following types: *unmappable(¬)*, *equivalent(≡)*, *less specific than (≤)*, *more specific than (≥)*, and *partially matched(≈)*.The function *sim*(c^0_s, c^1_k) represents the similarity measure adopted, where the outcome refers to a value between 0 and 1 (higher is the value, more similar are the concepts). The symbol σ defines the threshold for considering two concepts as semantically similar. Details about the studied similarity measures can be found in [16].The set of concepts of a given KOS at time t_0 is given by the function *Concepts*(K^0_x).

TABLE IV. MAPPING ADAPTATION ACTIONS. SEE FIG 1 FOR NOTATIONS

MAA	Description	Definition
Remove	This is an atomic action through which a mapping m^0_{st} is deleted from M^0_{ST}	$rmvM(m_{st}) \rightarrow m^0_{st} \in M^0_{ST} \wedge m^1_{st} \notin M^1_{ST}$
Addition	This is an atomic action through which a new mapping m^1_{st} is added to M^1_{ST}	$addM(m_{st}) \rightarrow m^0_{st} \notin M^0_{ST} \wedge m^1_{st} \in M^1_{ST}$
Move	This is a composed action for which an existing mapping from M^0_{ST} is re-allocated in M^1_{ST}, thus the source concept is different.	$moveM(m_{st},c^1_k) \rightarrow m^0_{st} \in M^0_{ST} \wedge m^1_{st} \notin M^1_{ST} \wedge c^1_k \in Concepts(K^1_S)$ $(\exists c^1_k \in CT(c^1_s), m^1_{kt} \in M^1_{ST} \wedge sim(c^0_s, c^1_k) \geq \sigma)$
Derivation	This is a composed action for which an existing mapping in M^0_{ST} is copied in M^1_{ST} with a different source concept.	$DeriveM(m_{st},c^1_k) \rightarrow m^0_{st} \in M^0_{ST} \wedge m^1_{st} \in M^1_{ST} \wedge c^1_k \in Concepts(K^1_S)$ $(\exists c^1_k \in CT(c^1_s), m^1_{kt} \in M^1_{ST} \wedge sim(c^0_s, c^1_k) \geq \sigma)$
Change Relation	This is a composed action in which the type of the semantic relation of a given mapping is modified.	$chgR(m_{st},new\_r_{st}) \rightarrow m^1_{st} \in M^1_{ST} \wedge new\_r^1_{st} \in \{\neg,\equiv,\geq,\leq,,\approx\}$ $(\exists m^1_{st} \in M^1_{ST}, new\_r^1_{st} = r^0_{st})$
No-Action	In this case, any modification is observed in the mapping	$no\text{-}action(m_{st}) \rightarrow m^0_{st} \in M^0_{ST} \wedge m^1_{st} \in M^1_{ST}$

At this stage, we aim to understand the correlations between the list of KCOs and the list of MAAs. We observe that for the studied KOS where concepts are described by textual attributes. We investigated how these attributes can be correlated to established mappings. To this end, we developed the *TopA* algorithm [16], which gives the *N* most relevant attributes for a given mapping. *TopA* relies on the adaptation of different semantic similarity measures targeting the lexical level [17], the syntactic level [18] and the semantic level [19].

We use the detected attributes during our correlation analysis, to verify, for example, whether changing these attributes to a different concept in another KOS version relates to moving the associated mapping. We defined four specific types of change patterns at the level of attributes: **Total transfer** (TT) – when an attribute is deleted from one concept and entirely moved to another one; **Partial Transfer** (PT) – when an attribute is deleted from one concept and a modified version of this attribute is added into another one), **Total Copy** (TC) – when an entire copy of the attribute is added into another concept; **Partial Copy** (PC) - a modified copy of the attribute is added into another concept. The partially copied *PC* or transferred attributes *PT* are identified by the degree of similarity with the original attribute (over a threshold σ). We assume that if the similarity *sim* is close to 1 (*i.e., sim* ≥ σ), we consider the attribute as totally copied *TC* or transferred *TT*.

Table V presents our findings regarding correlations between each MAA and the defined change patterns. We do not consider mapping addition in this analysis because MAA are only applied to modify established mappings. Our achieved results point out the correlation between MAA and the proposed change patterns. However, this demands further studies to explain why some correlations lack in some cases.

TABLE V. MAP-KOS CORRELATIONS

MAA	Nr.	TT	PT	TC	PC	No-change
Move	362	68	1	190	14	223
Derive	583	2	0	133	54	419
Remove	167	3	0	16	7	156
Change Relation	55	0	0	2	1	45
No-action	9024	16	2	176	70	8073

V. LESSONS LEARNED

While KOS maintenance belongs to the lifecycle of KOS, efforts to track changes about their evolution in the biomedical domain to automatically maintain their associated mappings up-to-date remain insufficient. Consequently, to define methods aiming to understand and classify KOS changes over time requires a complex and time consuming work. To overcome this problem, we had mostly to deal with:

- *Related to KOS evolution:*

- **Data quality**. Data sources are often published in a proprietary format or are not fully available in a computer-interpretable format. For instance, ICD9 is published as a MS-Word document that requires a parser to extract the KOS content. This parser is unavailable, which forces each research team to develop their own one (potentially leading to different datasets and errors).
- **Types of change operations.** We have defined a non-exhaustive list of change operations according to the conducted experiments and available tools. This may lead to unknown situations regarding how KOS evolve.
- **Accuracy of change operations identification between KOS versions**. Since we mainly ground our approach to mapping adaptation on changes in concept attributes (we have defined change patterns), and since these attributes stand for textual statements, it remains a very complex task to correctly identify these operations and interpret their consequences on concept attributes values.

- *Related to mappings:*

- **Identification of relevant KOS entities for interpreting mappings**. KOS's statements to justify why and how mappings were created and KOS's entities that impact the alignment of the concepts are unavailable in an explicit way for mappings considered in our studies, which may influence the accuracy of maintenance tasks.
- **Mappings' modification reasons.** Reasons explaining changes affecting mappings from one release to another are unavailable in datasets. We identified at least two main reasons for mapping adaptation in our work: (1) the propagation of KOS changes; and (2) the correction of erroneous mappings (when the validation process fails).
- **Mapping adaptation techniques.** The lessons learned from our conducted experiments have guided us to establish a set of useful heuristics to support automatic mapping maintenance.

Inspired by the investigation of Stojanovic [20] and Noy *et al.* [21] and based on how we managed the difficulties previously listed for the evolution of KOS and mappings, we propose a model for the process of KOS evolution (*cf.* Fig.2B). We partially transpose this model to inherently describe a mapping evolution process (*cf.* Fig.2A).

Following the phases proposed by Stojanovic, the KOS evolution process demands at least the list of KOS's entities requiring modification and of their respective types. We separate this information within two concepts *Elements* and *Goals*. Additionally, we deem useful to find the specific reasons for mapping changes. For instance, the correction of spelling mistakes can indicate a reason. Note that for clarity reasons, Fig.2 hides some relationships and attributes, *e.g.*, the relationship between goals and change patterns.

Considering the types of KOS changes, we assume that atomic changes are implemented alone and that we can associate them to achieve more complex changes [12]. We suggest three types of complex changes. The *Lexical Complex Changes* include modifications in attributes' values of concepts (we only consider textual attributes) (*cf.* Section Methods).

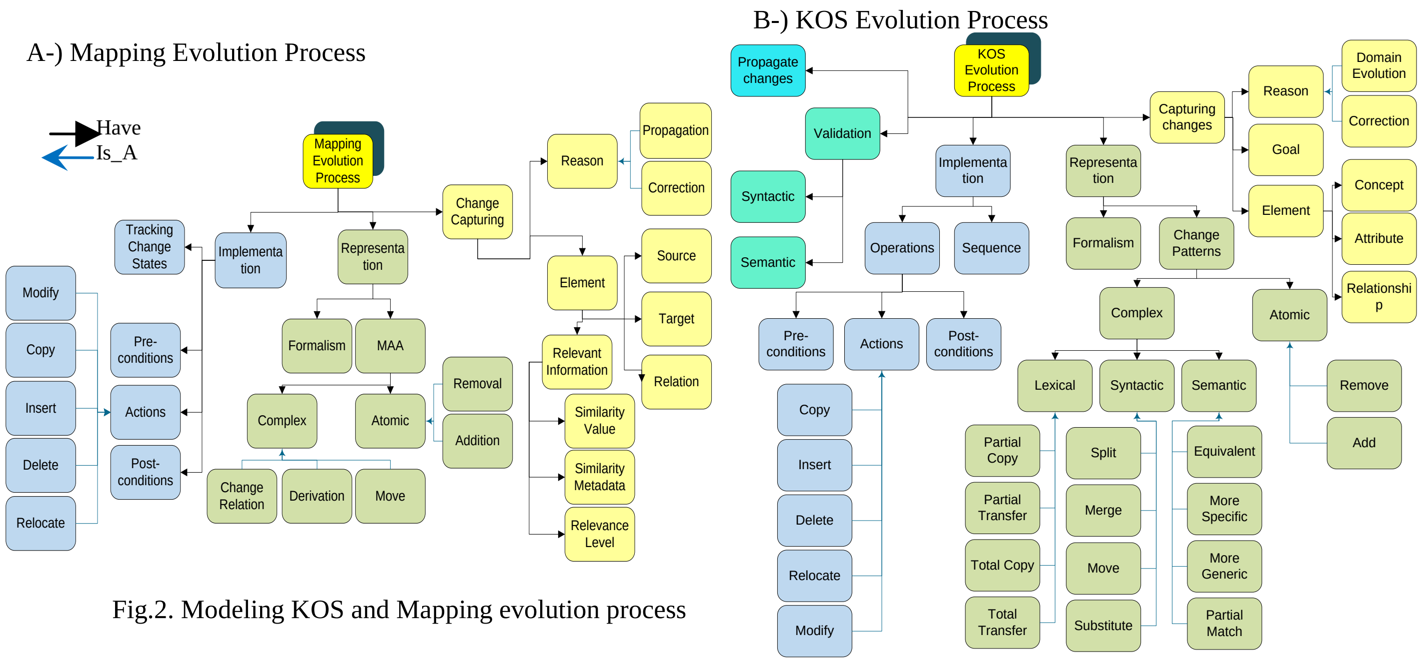

Fig.2. Modeling KOS and Mapping evolution process

These attributes can be partially or totally copied, or transferred to another concept, in another KOS version. This information might help identifying *Syntactic Complex Changes*. For instance, totally transferring an attribute to another concept can represent one of the evidences necessary to characterize a split [14]. The semantic consequences of a *Lexical Complex Change* are described in the *Semantic Complex Change*. For instance, an attribute transferred from one concept to another can have semantic consequences as follows: the former concept becomes more general and the later concept becomes more specific.

The implementation of KOS changes applies one or more change operations according to the pre-conditions and in a temporal order (*Sequence*). The change *Validation* consists of an important and laborious task where KOS engineers can detect and eliminate potential inconsistencies. Finally, the set of changes are propagated to associated artefacts (*e.g.*, mappings, annotations, queries, *etc.*).

The mapping evolution process (*cf.* Fig.2A) mostly differs from the KOS evolution process in the capturing phase, describing complex changes and representing the changes' state. In the capturing phase, we thoroughly detail the elements that best describe how mappings were established (*Relevant Information*), the similarity measures used (*Similarity meta-data*), their values (*Similarity value*), and the level of relevance of a specific element for mapping creation (*Relevance level*). Thus, if KOS evolution affects any relevant attribute (an identified KOS entity), we can measure the impact on mappings with more accuracy.

Complex changes applied to mappings include *Derivation* (a copy of a mapping with a new source concept), *Move* (transferring a mapping into another source concept) and *Change relation* (changing the semantic relation between source and target concepts). The process also tracks the state mappings change, and the status of each change includes valid, under validation, or invalid.

VI. CONCLUSION

This article performed a meta-analysis on a series of previously conducted empirical experiments that allowed on this basis a discussion of lessons learned. We highlighted the complexity of analysing KOS evolution when failing to report KOS changes in a computer-interpretable format. We focused our analyses on issues related to the propagation of KOS changes on mappings. Empirical experiments analysed were performed using a set of biomedical KOS and their associated mappings to observe and classify correlations between their changes. We originally introduced models for representing the KOS evolution process and the mapping evolution process. We designed these models to support mapping maintenance tasks.

Having methods that semi-automatically update KOS and maintain the validity of mappings might enable interoperable systems to follow logical reasoning and explain its implications to authoring groups in charge of the maintenance of KOS and mappings. Developers whose software depends on these semantic artefacts will allocate less time and costs for maintenance tasks if they can reduce the manual work required to handle different techniques used by KOS constructors for representing the domain and to report its changes over time.

In our future work, we will particularly improve the proposed methods and tools to automatically identify semantic complex changes between KOS versions. We also plan to generalize our approach evaluating in other domains, and to study new similarity measures to apply for KOS with instances and non-textual attributes.

ACKNOWLEDGMENT

The National Research Fund (FNR) of Luxembourg (Grant #C10/IS/786147) entirely supports this work under the *DynaMO* research project.

REFERENCES

[1] G. Hodge, *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington: Council on Library and Information Resources, 2000.

[2] J. Euzenat and P. Shvaiko, *Ontology Matching*: Springer, 2007.

[3] J. C. Dos Reis, *et al.*, "Mapping adaptation actions for the automatic reconciliation of dynamic ontologies". In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, USA, 2013, pp. 599-608.

[4] J. C. Dos Reis, *et al.*, "Understanding Semantic Mapping Evolution by Observing Changes in Biomedical Ontologies". *Journal of Biomedical Informatics,* 2014, vol. 47, pp. 71-82.

[5] C. Yu and L. Popa, "Semantic Adaptation of Schema Mappings when Schemas Evolve" In *Proceedings of the 31st international conference on Very large data bases*, 2005, pp. 1006-1017.

[6] A. Gross, *et al.*, "Semi-Automatic Adaptation of Mappings between Life Science Ontologies," In *Proceedings of the 9th International Conference on Data Integration in the Life Sciences*, Montreal, Canada, 2013, pp. 90-104.

[7] A. Groß, *et al.*, "How do computed ontology mappings evolve? - A case study for life science ontologies" In Joint Workshop on Knowledge Evolution and Ontology Dynamics at ISWC, 2012.

[8] Y. An, *et al.*, "Maintaining Mappings between Conceptual Models and Relational Schemas" *Journal of Database Management,* vol. 21, pp. 36-68, 2010.

[9] N. Martins and N. Silva "A User-driven and a Semantic-based Ontology Mapping Evolution Approach" In *11th International Conference on Enterprise Information System*, Milano, Italy, 2009, pp. 214-221.

[10] O. Bodenreider, "Comparing SNOMED CT and the NCI Thesaurus through Semantic Web Technologies," in *KR-MED*, 2008.

[11] D. E. Oliver, *et al.*, "Representation of change in controlled medical terminologies," *Artificial Intelligence in Medicine,* vol. 15, pp. 53-76, 1999.

[12] M. Klein, *et al.*, "Ontology versioning and change detection on the web," in *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*: Springer, 2002, pp. 197-212.

[13] M. Hartung, *et al.*, "COnto–Diff: generation of complex evolution mappings for life science ontologies" *Journal of Biomedical Informatics,* vol. 46, pp. 15-32, 2013.

[14] J. Dos Reis, *et al.*, "Characterizing Semantic Mappings Adaptation via Biomedical KOS Evolution: A Case Study Investigating SNOMED CT and ICD" In *Proceedings of the AMIA Symposium*, USA, 2013, pp. 333-342.

[15] J. C. Dos Reis, *et al.*, "Analyzing and supporting the mapping maintenance problem in biomedical knowledge organization systems" *In Proceedings of SIMI Workshop at ESWC,* Heraklion, Grece, 2012, pp. 25-36.

[16] J. C. Dos Reis, *et al.*, "The influence of similarity between concepts in biomedical ontology evolution for mapping adaptation". In *Proceedings of the 25th European Medical Informatics Conference*. Istanbul, Turkey, 2014.

[17] V. I. Levenshtein. "Binary codes capable of correcting deletions, insertions and reversals" In *Soviet physics doklady*, 1966, pp. 707.

[18] A. Maedche and S. Staab, "Measuring similarity between ontologies" In *Knowledge engineering and knowledge management: Ontologies and the semantic web*: Springer, 2002, pp. 251-263.

[19] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy". In *Proceedings of the International Conference on Research in Computational Linguistics,* 1997, pp. *19-33*.

[20] L. Stojanovic, *et al.*, "User-driven ontology evolution management" In *Knowledge engineering and knowledge management: ontologies and the semantic web*: Springer, 2002, pp. 285-300.

[21] N. F. Noy, *et al.*, "A framework for ontology evolution in collaborative environments," In *The Semantic Web-ISWC*, Springer, 2006, pp. 544-558.