Cross-domain Semantic Drift Measurement in Ontologies Using the SemaDrift Tool and Metrics

Thanos G. Stavropoulos[1], Efstratios Kontopoulos[1], Albert Meroño Peñuela[2], Stavros Tachos[1], Stelios Andreadis[1], Ioannis Kompatsiaris[1]

[1] Information Technologies Institute, Thessaloniki, Greece
{athstavr, skontopo, staxos, andreadisst, ikom}@iti.gr
[2] Vrije Universiteit Amsterdam, The Netherlands
albert.merono@vu.nl

Abstract. Detecting and measuring semantic drift in different versions of ontologies across time is a novel area of research rapidly gaining attention. Nevertheless, only a few practical methods and tools have addressed the task. while even fewer are flexible enough to be efficiently applied to multiple domains. As the often domain-specific nature of ontologies may render methods and tools for measuring semantic drift ineffective, this paper presents the application and findings of the SemaDrift suite of methods and tools, presenting novel insights in several domains for the first time. While developed in the context of the PERICLES FP7 project, aimed at Digital Preservation, domain-independent text and structural similarity measures, available both as a software library and as a Protégé plugin for end-users, are now applied in the Dutch Historical Census and the BBC Sports Ontology. The two different domains demonstrate its applicability and ability to pinpoint the location, nature, origins and destinations of concept drift.

Keywords: Semantic drift; concept drift; semantic change; ontologies; Protégé.

1 Introduction

The evolution of semantics is an active area of research, especially challenged by the lack of universal metrics to address specificities and peculiarities pertinent to each domain. *Evolving semantics*, also referred to as *semantic change*, observe and measure the phenomenon of change in the meaning of concepts within knowledge representation models, along with their potential replacement by other meanings over time. In the Semantic Web [1], the representation of the underlying knowledge is typically assumed by ontologies. Thus, it can be easily perceived that semantic change can have drastic consequences on the use of ontologies in Semantic Web and Linked Data applications. In this setting, semantic change, i.e. the structural difference of the same concept in two ontologies [2], relates to various lines of research. Such examples are *concept* and *topic shift* [3], *concept change* [4], *semantic decay* [5], *ontology versioning* [6] and *evolution* [7]. A brief disambiguation of these terms can be found in [8].

Semantic drift can be defined as the phenomenon of ontology concepts gradually changing as our knowledge of the world evolves, obtaining possibly different meanings, as interpreted by various user communities or in a different context, risking their rhetorical, descriptive and applicative power [9]. *Concept drift* can refer to this language-related phenomenon, but also in abrupt parameter value changes in data mining [10].

This paper presents findings in two vastly diverse domains through applying a novel set of universal, domain agnostic semantic drift metrics across various domains using the SemaDrift suite of tools and metrics. The metrics, initially presented in [8], are embedded in respective software tools, that offer the means for domain experts to assess drift without programming knowledge. Namely, the SemaDrift plugin for the Protégé platform[1] aims at assisting a wider audience to monitor and manage concept drift, and was developed in the context of the PERICLES FP7 project[2], integrating and extending existing studies [3] and previously developed open, reusable methods [8].

The domains studied in this paper are (a) the CEDAR dataset containing historical Dutch census data, and (b) the BBC Sport Ontology for representing competitive sports events. In the historical census domain, the metrics help pinpoint historical occupation qualities for the population between 1869 and 1930, almost on the fly. Most importantly, using the same tool we move on to the BBC Sport Ontology were the metrics pinpoint the location, the nature, the origins and destinations of concept drift, across six versions. The ontologies used in this work and the SemaDrift outputs are publicly available online[3].

The rest of the paper is structured as follows: Section 2 presents related work in metrics and tools for measuring drift. Section 3 presents the underlying metrics and the SemaDrift framework. Sections 4 and 5 present the two proof-of-concept scenarios and report on our findings, while conclusions and future work are listed in the final section.

2 Related Work

Measures of semantic richness of Linked Data concepts have been investigated in [5], proving that increasing reuse of concepts decreases its semantic richness. Other studies have examined change detection between two ontologies at a structural or content level [2]. Concept drift has been measured either by clustering while populating ontologies [11] or by applying linguistic techniques on textual concept descriptions [12]. A vector space model by random indexing has been utilized to track changes of an evolving text collection [9] and to visualize the drift of vocabularies in a diachronic sample of the Linked Open Data cloud [13]. A strategy to represent change has been based on ontology evolution [7]. However, most of these techniques are not directly applicable to Semantic Web constructs or present limited statistical data.

---

[1] The Protégé Ontology Editor: http://protege.stanford.edu

[2] PERICLES FP7 project: www.pericles-project.eu

[3] GitHub Repository with datasets and results: https://github.com/skontopo/MEPDaW2017

An appealing solution we have adopted transfers the notions of *label*, *extension* and *intension* from machine learning concept drift to semantic drift, further defining them in ontology terms [3]. Much philosophical debate examines how and by which properties a concept can be identified across time and appropriate formalization [14]. Some have utilized the notions of *perdurance* and *endurance* [15], so as to seek identity, by defining rigid properties that have to be persistent across instances and, thus, can identify entities [10]. In this work we adopt, implement and integrate the methods in [3] into a familiar application for knowledge engineers, targeting not only the lack of reproducible cross-domain metrics for semantic drift, but also the lack of similar graphical user interfaces.

3 Semantic Drift Metrics and the SemaDrift Platform

3.1 Semantic Drift Metrics

The drift metrics considered here implement and extend previous work in the field of concept drift [8], where highly applicable notions and metrics for measuring concept drift in the context of data mining have successfully been transferred to semantic drift. The method to measure concept drift in semantics considers two basic factors: (a) the different aspects of change, and (b) whether concept identity is known or not. The aspects of change can be:

- *Label*, which refers to the description of a concept, via its name or title;
- *Intension*, which refers to the characteristics implied by it, via its properties;
- *Extension*, which refers to the set of things it extends to, via its number of instances.

Meanwhile, the correspondence of a concept across versions can be either known or unknown, resulting in two different approaches for measuring change:

- *Identity-based approach* (i.e. known concept identity): Assessing the extent of shift or stability of a concept's meaning is performed under the assumption that its identity is known across ontologies. For instance, considering an ontology *A*, and its evolution, ontology *B*, each concept of *A* is known to correspond to a single, known concept of *B*.
- *Morphing-based approach* (i.e. unknown concept identity): Each concept is pertaining to just a single moment in time (ontology), while its identity is unknown across versions (ontologies), as it constantly evolves/morphs into new, even highly similar, concepts. Therefore, its change has to be measured in comparison to every concept of an evolved ontology.

The currently proposed method considers the more general morphing-based approach and considers drift as the dissimilarity of two maximally similar concepts in two versions [3]. Despite several methods have been proposed to seek identity correspondence across versions [10], they still can be domain or model dependent, mandating for ad-hoc expert knowledge in the form of annotations, user input or using explicit identities. In order to measure change, the meaning of each concept at a given point t (e.g. in time) is defined as a set of the three different aspects, as follows:

$$C^t = < label_t(C), int_t(C), ext_t(C) >$$

where C^t denotes the meaning of concept C at point t. Each of its aspects, $label_t(C)$, $int_t(C)$ for intensional and $ext_t(C)$ for extensional, is measured as follows:

$$label_t(C) = \{l, \mid \forall\langle C, rdfs{:}label, l\rangle \in T\}$$

$$int_t(C) = \{i \mid i = \langle C, p, x \rangle \vee i = \langle x, p, C \rangle, p = rdfs{:}domain \vee p = rdfs{:}range, \forall i \in T\}$$

$$ext_t(C) = \{x \mid \forall\langle x, rdf{:}type, C \rangle \in T\}$$

where T is the set of all triples in the ontology version t. Namely a) label is the $rdfs{:}label$ of a concept (a string), b) intension is a set of triples (i.e. the properties that involve the concept, calculated as the union of all RDF triples with C in the subject or object position of OWL Object Properties or OWL Datatype Properties) and c) extension is the set of strings (i.e. the names of instances with the concept as value of $rdf{:}type$). Due to the morphing based approach, each concept's drift is measured as the average drift to all concepts of the next ontology. Comparisons for strings are made using the Monge-Elkan algorithm [16], found to optimally suit strings in ontologies such as CamelCase or snake_case, and Jaccard similarity for sets.

In detail, if n_2 is the total number of concepts in t_2, we define label, intensional and extensional drifts of C between versions $t_1$and t_2 as follows:

$$label_{t_1 \to t_2}(C) = \frac{\sum_{i=1}^{n_2} MongeElkan\left(label_{t_1}(C), label_{t_2}(C_i)\right)}{n_2}$$

$$int_{t_1 \to t_2}(C) = \frac{\sum_{i=1}^{n_2} Jaccard\left(int_{t_1}(C), int_{t_2}(C_i)\right)}{n_2}$$

$$ext_{t_1 \to t_2}(C) = \frac{\sum_{i=1}^{n_2} Jaccard\left(ext_{t_1}(C), ext_{t_2}(C_i)\right)}{n_2}$$

A holistic aspect, $whole$ is defined as their average:

$$whole_{t_1 \to t_2}(C) = \frac{label_{t_1 \to t_2}(C) + int_{t_1 \to t_2}(C) + ext_{t_1 \to t_2}(C)}{3}$$

3.2 The SemaDrift Tool to Measure Semantic Drift

While the SemaDrift metrics discussed above can be used directly in third-party scripts and software via the SemaDrift API Library, a domain expert or researcher may not possess the ability to do so. Especially in the case that several domain ontologies need to be explored, as in this study, researchers need to use a common tool to work fast, reliably and without further adaptations as a common point of reference. The SemaDrift Protégé plugin serves this purpose[4].

The plugin's main panel is shown in Fig. 1. The tool provides a subset of the basic functions of the underlying SemaDrift API in a graphical manner. For that purpose, it exposes some of its functions and accommodates the outcomes in suitable user controls using the Java Swing library. This edition of the plugin focuses on ontology pairs, i.e. two versions of the same ontology, in order to provide more insight in them and their differences, fitting also into the Protégé workspace philosophy. Usually, the users work on a single ontology at a time, which is always displayed as a tree hierarchy of classes at the left pane. Then, plugins occupy the right pane, which is free to accommodate their functions (Fig. 1).

As a first step the user has to select the pair of ontologies for which to measure drift. To take advantage of the environment, the plugin assumes that the first selected ontology is the one currently loaded in Protégé, allowing also its in-depth visualization, reasoning and query execution. The second ontology can be selected from the SemaDrift pane using the "Browse" button to look through local or remote storage.

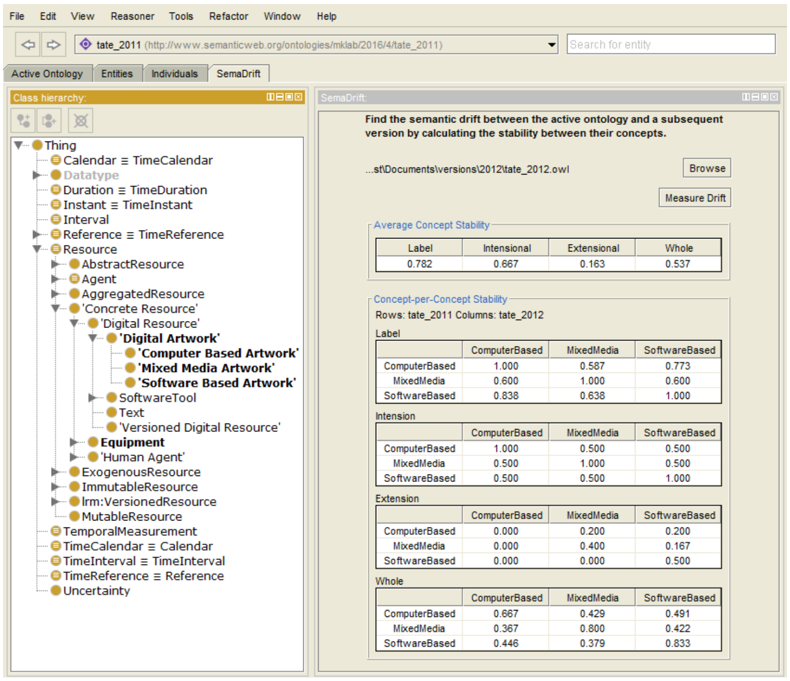

Fig. 1. SemaDrift Protégé plugin: The native tree hierarchy of the open ontology is shown on the left, while the plugin-provided content resides on the right, showing a second ontology to compare to, accompanied by the respective measurements.

[4] The SemaDrift suite is available at: http://mklab.iti.gr/project/semadrift-measure-semantic-drift-ontologies.

After both ontologies are available, pressing on the "*Measure Drift*" button will display the SemaDrift metric results. Stability, as a measure of drift, is shown in two sections: overall average stability per aspect and concept pair stability for all aspects. The first section constitutes the most generic, abstract measure of drift. It displays a table with the average drift of all concepts from the former ontology to the latter, per each of the four aspects: label, intension, extension and whole. Naturally, the measurements are derived using the metrics and algorithms for each aspect described in the previous section, yielding a value from zero (no similarity) to one (full similarity).

The second section of results is displayed in respective tables. Each table row corresponds to a concept of the former ontology and each column to a concept of the latter. Consequently, each cell holds the similarity metric (i.e. concept stability) between each pair of concepts. These similarity values between pairs can further be utilized by users for different purposes for example to generate similarity graphs or *morphing chains* such as those demonstrated in the rest of the paper.

4 Dutch Historical Censuses

Census data are essentially time series of systematic population records, and hence an important source for studying semantic drift of concepts involving culture and economics. In the Netherlands, the Dutch historical censuses are 17 country-wide population reports performed between 1795 and 1971, once every 10 years. In each of these reports, the government counted the population of the country and its demographic, occupational, and housing characteristics. In 1971, this detailed reporting stopped due to social concern on privacy. Nevertheless, the exhaustive, detailed, and aggregated characteristics[5] of these censuses have continued to attract the attention of historians and social scientists [17], who nowadays study them via a collection of 507 machine-readable spreadsheets, containing 2,288 census tables[6]. In order to improve more systematic and universal access to reproduce results of studies on this dataset, recent efforts have managed to publish it as Linked Data (LD) [18]. This LD set of the Dutch historical census will be further called the CEDAR dataset.

SemaDrift was used to analyze semantic drift in the CEDAR dataset and gain insights as to how *occupational concepts*, describing citizen jobs, changed in the period 1869-1930. For each of these two years, these occupational concepts are described with three attributes: (a) the *occupational concepts* themselves are SKOS concepts [19] using URIs of the Historical International Standard Classification of Occupations [20] (HISCO); (b) the *number of persons* having a specific job are associated to those SKOS concepts; and (c) a *number of labels* (in Dutch) are associated to these SKOS concepts. This implies that the *constraints* inherent to these data with respect to *intensions*, *extensions*, and *labels* are as follows: intensions do not exist for these occupational concepts, since no further formal descriptions are available; extensions are

5 Notably, the microdata registers (i.e. individual survey data), upon which these censuses are built, have been lost over time. Hence, the numbers are only aggregations, with no tracking information leading to the original individuals.

6 See http://volkstellingen.nl/

restricted by the cardinality of the concepts; and labels are assigned and abundant for all concepts in both years. An example is shown in Fig. 2.

A data transformation step is required before feeding the dataset into the tool to address not only the format but also to generate more meaningful properties. Namely, the data format, as originally shown in Fig. 2, is transformed to two OWL ontologies for the years 1869 and 1930. To do so, we convert every HISCO `skos:Concept` to an `owl:Class`, assigning them all `rdfs:label` in the original data. Furthermore, to obtain an accurate representation of extensions, we unroll the integer counts, as seen on Fig. 2, and generate as many anonymous instances as specified by these integers. This is done since, following a proper ontological representation, the extensional aspect actually refers to instances and not numerical properties. Finally, we assign these anonymous instances to their corresponding HISCO `owl:Class` using an `rdf:type` relation, thus each of them representing one person that carried the job indicated by the class.

```
cedar:hisco-13100 a skos:Concept ;
    skos:prefLabel "Hoogleeraar"@nl .

cedar:BRT-1930/37860 a qb:Observation ;
    cedar:occupation cedar:hisco-13100 ;
    qb:dataSet cedar:BRT-1930 ;
    cedar:populationSize "52"^^xsd:integer .
```

Fig. 2. Excerpt of the original CEDAR data. Census counts are modeled as RDF Data Cube observations, which carry information about the occupation class and the number of persons belonging to it.

After using SemaDrift for the two ontologies, respective average per aspect stability and average concept-per-concept stability are generated. The latter is used to draw morphing chains for topics of interest. After observing the table, first, the most stable concept between both versions is the occupational class `hisco:-1` (stability of 0.917 – see Fig. 3), the class for occupations that cannot be classified elsewhere in HISCO. This is due a great label (0.750) and extensional (1.000) stability, which suggests that both the coherency of data coders w.r.t. unclassifiable jobs, and the population carrying those, remained stable in this period.

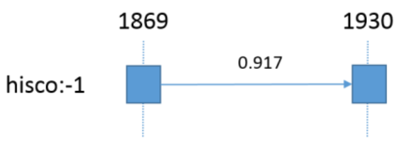

Fig. 3. Morphing chain of the whole aspect of the `hisco:-1` concept.

According to previous studies in extensional semantic drift in this dataset [21], other interesting classes from 1869 with expected extensional drift are:

- hisco:97125, loaders of ships, trucks, wagons or airplanes. These workers do not appear again in 1930, and the stability w.r.t. similar classes, like hisco:97145 (storehouse workers), is significantly lower (0.479). Their closest matches in terms of stability are varnishers and stone polishers (0.717);
- hisco:21110, general managers. This group does appear in 1930, but the similarity of their classes has greatly drifted (0.511). Many other occupational jobs, with loose semantic similarity, display more stability w.r.t. the original class;
- hisco:41025, working proprietors. Similarly, this group of workers shows a great deal of drift, to the extent of not having an equivalent class in 1930. This might be due to historical reasons, i.e. the late industrialization in the Netherlands and its effects on evolving old small business owners into upper-class company investors. Noticeably, the related class hisco:43200, commercial agents, displays certain stability (0.405).

It is important to underline some intrinsic limitations in the study of semantic drift within the CEDAR dataset. Besides the lack of concept intensions, many of these limitations are related to the problem of *identity*, as also reported by [3]. First, identity between *classes* cannot be assumed even between those of identical HISCO codes, since these are convoluted and culturally changing time periods. Secondly, identity between *instances* of these classes is even more volatile. Human annotated identity information such as the existence of owl:sameAs links between instances of different time periods would greatly improve the outcomes of the extensional drift analysis, but require manual labor.

The initial data transformation effort required in this scenario is justified, as the tool assumes proper ontology format (OWL) and design (instances instead of numeric properties according to their meaning). However, the tool itself with the existing morphing-based metrics is apparently very useful to very quickly gain access to insights regarding the evolution of semantic concepts that would otherwise require serious labor.

5 The BBC Sport Ontology

BBC is one of the pioneers in the field of ontology-based technologies, using them at an industrial level since 2010[7]. In the past, they found that conventional content management systems impose serious limitations on the flexibility of the ways that content is served, limiting the richness of the experience they offer to their visitors. To overcome these limitations and enhance the experience for website users, they turned to ontologies and Linked Data. An additional key benefit is that this approach also significantly reduces the time it takes for editors to create content that is easily discoverable across the website.

One of the first ontologies developed by BBC was the *Sport Ontology*[8], which initially started as an effort to represent information about the competitions, teams, play-

[7] BBC ontologies homepage: http://www.bbc.co.uk/ontologies

[8] BBC Sport Ontology homepage: http://www.bbc.co.uk/ontologies/sport

ers and matches of the 2010 World Cup. However, although it originated as a specific use case, the Sport Ontology has been extended and is now applicable to representing a wide range of competitive sporting events. The BBC now use this ontology to support their sports coverage, including coverage of both the 2012 London Olympics and the 2014 Brazil World Cup.

The ontology's significance for BBC, along with its potential applicability in various sports-related deployments, has led to our inclusion of the Sport Ontology in this study. However, compared to the case study presented in the previous subsection, the scope of this analysis is different in the sense that we are investigating design decisions from version to version, possibly influenced by the company's intended end-user applications and the public's demonstrated preference to certain pertinent aspects. Table 1 contains information regarding the versions of the Sport Ontology studied in this paper[9].

Table 1. Information regarding the BBC Sport Ontology versions studied in the paper.

Ontology version	v2.10	v2.11	v2.12	v2.13	v3.0	v3.2
Date created	20-Feb-14	27-Mar-14	16-Sep-14	9-Feb-15	8-Apr-15	14-May-15
Class count	35	37	37	37	39	38
Property count	50	49	49	49	49	49
Individual count	27	21	33	41	41	41

As derived by SemaDrift (but also implied in the table), the ontology is extremely stable with regards to its intensional aspect (i.e. classes and properties), with most classes demonstrating a perfect stability of 1. An exception was observed for classes `CompetitiveSportingGroup` and `CompetitiveSportingOrganisation`, whose stabilities were reduced to an average of 0.9 each, due to changes in domains and ranges of respective properties in versions 2.11 and 2.12.

On the other hand, the ontology is less stable extensionally, which is mostly due to instances being added to specific classes in versions 2.11, 2.12 and 2.13. More specifically, the initially empty (i.e. no instances) class `RoundType` was populated with 12 instances in version 2.12 (e.g. `final`, `quarter-final`, `semi-final` etc.) and with 4 additional instances in version 2.13. Additionally, class `CompetitionType`, which initially had 17 instances (e.g. `domestic-cup`, `european-cup`, `international` etc.), was populated with 4 additional instances in version 2.13. Overall, version 2.13 was the one where the extension of the ontology was finalized. The relevant morphing chains illustrating the drifts of the "whole" aspect for the three ontology versions involving the respective classes are illustrated in Fig. 4.

---

[9] Note that versions prior to v2.10 were not available online on the BBC website.

Finally, the changes in the two most recent versions of the ontology (3.0 and 3.2) were minimal, thus indicating that the ontology has stabilized.

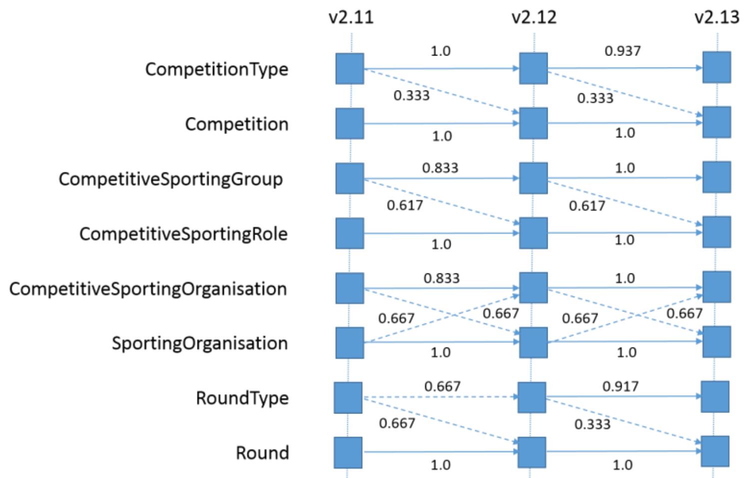

Fig. 4. Morphing chains illustrating the drifts of the "whole" aspect in versions 2.11, 2.12 and 2.13 of the BBC Sport Ontology.

6 Conclusions and Future Work

This paper employed novel ways to measure semantic drift in two different domains: historical censuses and competitive sports events. SemaDrift is a suite of tools, a software library and an application, that can measure drift aspects for ontologies on-the-fly. Linked data for the Dutch historical census from 1869 to 1930 were transformed to OWL and processed to show interesting insights for the semantic change in the population's occupation concepts. Moreover, semantic drift was studied for six versions of the BBC Sport Ontology. Using the same tool to gain insights for unrelated domains demonstrated its universal and cross-domain properties. Also, its usefulness is shown, as it gives access to insights otherwise hard to obtain, such as to assess the nature of the drift (extensional), locate it in time and track the migration of meaning from concept to concept through morphing chains.

Future work will be focused on expanding to more domains and extending the tools. As already apparent in this study, the tool may handle most ontologies out-of-the-box enabling researchers without programming knowledge to do more. However, the historical censuses have uncovered not only a minor change in format but also in ontology design. As both these matters were solved by writing a transformation script, such scripts may be incorporated into the tool for future use. Furthermore, the lack of matching identities, elaborated on in previous studies [8], may be handled by alternative metrics. Finally, additions to the tool's GUI include handling more ontologies, adding visual aids and drawing abilities for morphing chains evolving the tool into a one-stop-shop for semantic drift measurement.

Acknowledgments. This research received funding by the European Commission Seventh Framework Programme under Grant Agreement Number FP7-601138 PERICLES.

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Sci. Am. 284, 28–37 (2001).
2. Tury, M., Bieliková, M.: An approach to detection ontology changes. In: Workshop proceedings of the sixth international conference on Web engineering - ICWE '06. p. 14. ACM (2006).
3. Wang, S., Schlobach, S., Klein, M.: Concept drift and how to identify it. J. Web Semant. 9, 247–265 (2011).
4. Uschold, M.: Creating, integrating and maintaining local and global ontologies. In: Proceedings of the First Workshop on Ontology Learning (OL-2000) in conjunction with the 14th European Conference on Artificial Intelligence (ECAI 2000), Berling, Germany.
5. Pareti, P., Klein, E., Barker, A.: A Linked Data Scalability Challenge: Concept Reuse Leads to Semantic Decay. In: Proceedings of the ACM Web Science Conference. ACM Press-Association for Computing Machinery (2016).
6. Yildiz, B.: Ontology Evolution and Versioning. Tech. Report, TU Vienna. (2006).
7. Stojanovic, L., Maedche, A., Motik, B., Stojanovic, N.: User-driven ontology evolution management. Knowl. Eng. Knowl. Manag. Ontol. Semant. Web. 133–140 (2002).
8. Stavropoulos, T.G., Andreadis, S., Riga, M., Kontopoulos, E., Mitzias, P., Kompatsiaris, I.: A Framework for Measuring Semantic Drift in Ontologies. In: 1st Int. Workshop on Semantic Change & Evolving Semantics (SuCCESS'16). CEUR Workshop Proceedings, Leipzig, Germany (2016).
9. Wittek, P., Darányi, S., Kontopoulos, E., Moysiadis, T., Kompatsiaris, I.: Monitoring Term Drift Based on Semantic Consistency in an Evolving Vector Field. In: 15th International Joint Conference on Neural Networks (IJCNN 2015). pp. 1–8. IEEE, Killarney, Ireland (2015).
10. Meroño-Peñuela, A., Hoekstra, R.: What is Linked Historical Data? In: Proceedings of the 19th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2014). pp. 282–287. Springer International Publishing (2014).
11. Fanizzi, N., Amato, C., Esposito, F.: Conceptual Clustering: Concept Formation, Drift and Novelty Detection. In: The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain, June 1-5, 2008, Proceedings. pp. 318–332. Springer (2008).
12. Gulla, J., Solskinnsbakk, G., Myrseth, P.: Semantic Drift in Ontologies. In: Proceedings of 6th International Conference on Web Information Systems and Technologies (WEBIST), Valencia, Spain. pp. 13–20 (2010).
13. Meroño-Peñuela, A., Wittek, P., Darányi, S.: Visualizing the Drift of Linked Open Data Using Self-Organizing Maps. In: Drift-a-LOD Workshop at the 20th International Conference on Knowledge Engineering and Knowledge Management (2016).
14. Guarino, N., Welty, C.: A Formal Ontology of Properties. In: Knowledge Engineering and Knowledge Management Methods, Models, and Tools. pp. 97–112. Springer Berlin

Heidelberg (2000).
15. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with DOLCE. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 166–181. Springer Berlin Heidelberg (2002).
16. Monge, A.E., Elkan, C.: The Field Matching Problem: Algorithms and Applications. In: 2nd Intl. Conf. Knowledge Discovery and Data Mining (KDD). pp. 267–270 (1996).
17. Ashkpour, A., Meroño-Peñuela, A., Mandemakers, K.: The aggregate Dutch historical censuses: Harmonization and RDF. Hist. Methods A J. Quant. Interdiscip. Hist. 48, 230–245 (2015).
18. Meroño-Peñuela, A., Ashkpour, A., Guéret, C., Schlobach, S.: CEDAR: the Dutch historical censuses as linked open data. Semant. Web. 1–14 (2015).
19. Miles, A., Bechhofer, S.: SKOS simple knowledge organization system refer-ence. (2009).
20. Leeuwen, M.H.D. van, Maas, I., Miles, A.: HISCO: Historical international standard classification of occupations. Leuven: Leuven University Press (2002).
21. Meroño-Peñuela, A., Guéret, C., Hoekstra, R., Schlobach, S.: Detecting and reporting extensional concept drift in statistical linked data. Presented at the (2013).