applied sciences

Article

Uncertainty in Automated Ontology Matching: Lessons from an Empirical Evaluation

Inès Osman [1,*], Salvatore Flavio Pileggi [2] and Sadok Ben Yahia [3,4]

[1] LIPAH-LR11ES14, Faculty of Sciences of Tunis, University of Tunis El Manar, Tunis 1060, Tunisia
[2] Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW 2007, Australia; salvatoreflavio.pileggi@uts.edu.au
[3] Department of Software Science, Tallinn University of Technology, 12618 Tallinn, Estonia; sadok.ben@taltech.ee
[4] The Maersk Mc-Kinney Moller Institute, Center of Industrial Software, University of Southern Denmark, 6400 Sønderborg, Denmark
* Correspondence: ines.osman@fst.utm.tn

Abstract: Data integration is considered a classic research field and a pressing need within the information science community. Ontologies play a critical role in such processes by providing well-consolidated support to link and semantically integrate datasets via interoperability. This paper approaches data integration from an application perspective by looking at ontology matching techniques. As the manual matching of different sources of information becomes unrealistic once the system scales up, the automation of the matching process becomes a compelling need. Therefore, we have conducted experiments on actual non-semantically enriched relational data with the support of existing tools (pre-LLM technology) for automatic ontology matching from the scientific community. Even considering a relatively simple case study—i.e., the spatio–temporal alignment of macro indicators—outcomes clearly show significant uncertainty resulting from errors and inaccuracies along the automated matching process. More concretely, this paper aims to test on real-world data a bottom-up knowledge-building approach, discuss the lessons learned from the experimental results of the case study, and draw conclusions about uncertainty and uncertainty management in an automated ontology matching process. While the most common evaluation metrics clearly demonstrate the unreliability of fully automated matching solutions, properly designed semi-supervised approaches seem to be mature for more generalized application.

Keywords: data integration; ontology matching; uncertainty

Citation: Osman, I.; Pileggi, S.F.; Yahia, S.B. Uncertainty in Automated Ontology Matching: Lessons from an Empirical Evaluation. *Appl. Sci.* **2024**, *14*, 4679. https://doi.org/10.3390/app14114679

Academic Editors: Pierluca Sangiorgi and Antonio Pagliaro

Received: 9 May 2024
Revised: 28 May 2024
Accepted: 28 May 2024
Published: 29 May 2024

1. Introduction

Data integration, defined as "*the problem of combining data residing at different sources and providing the user with a unified view of these data*" [1], can be considered a well-covered research field, as the many contributions in the literature clearly confirms. Its relevance is determined by the practical implications in the different application domains, and it is well-recognized within the information science community [2]. We remark that many modern systems work at a semantic level [3], where data integration may be understood at different levels (e.g., concept [4] or multi-media [5]).

Semantic technology has been largely adopted in data integration [1], as it is definitely central to implement a more holistic approach, where data integration is considered a part of a more complex knowledge-building process. The proper adoption of semantic technology is extremely effective in supporting data integration and reuse via interoperability [6].

More in general, associating formal semantics with data is a key step in the fields of artificial intelligence and database management [2]. In addition, the analysis of semantic data can underpin sophisticated data mining techniques [7,8].

In the context of this work, knowledge building is seen as the process of combining relational data in order to create rich data spaces in which semantics are formally defined [1]. While data integration aims at establishing a common, unified view of data from different sources, the specification of formal semantics enables a further level of complexity, as the “meaning” of data is also represented in a fully machine-interpretable context.

Ontology is a classic philosophical concept that is related to the study of “the nature of being”. It has become an active part of the computer world [9], where it is a rich data model aimed at the formal specification of semantics. It supports the representation and processing of knowledge in a machine-readable format according to a model close to the human one. The adoption of ontologies allows effective solutions for knowledge building, since semantic data are enabled in the Semantic Web [10,11] and are modeled according to an advanced interoperability model, which is commonly referred to as semantic interoperability [12].

We fully rely on an ontology-based approach to support the data integration process. The benefits of ontology in different application domains are well-known and have been extensively discussed from different perspectives by several contributions [2].

An ad hoc approach to knowledge building is time-consuming, error-prone, and, in general, very expensive. Looking at the increasing complexity and scale of systems, automated and semi-automated data integration processes are becoming more and more relevant to assure effectiveness and performance on a large scale. The automation of the knowledge-building process becomes required once the target system scales up.

Ontology matching is a crucial step in the knowledge-building process. It reconciles the differences between ontologies and resolves their heterogeneity problem. A seamless and systematic knowledge-building process may only be considered adequate by assuming manual matching of concepts from the different sources of information. However, since manual matching is far from being scalable, automation becomes a compelling need. Automated ontology matching systems use a similarity computation algorithm to find similarities between ontologies to be integrated. However, as extensively discussed in the rest of this paper, such automation leads to a situation of uncertainty since correctness and accuracy cannot be guaranteed in general terms. In this paper, we focus on uncertainty and uncertainty management in the automated process of ontology matching.

We approach the knowledge-building process from a practical perspective. Indeed, we apply a real-world case study dealing with the spatio–temporal alignment of macro indicators. We propose several experiments on non-semantically enriched data by adopting existing tools for dataset conversion and ontology matching. We first adopt a previously developed conversion tool that enables the systematic translation from raw data (relational tables) to semantic data (ontologies) [2]. We note that such a conversion process does not include any semantic enrichment of the original datasets. Then, available ontology matching tools from the community are applied to find similarities among the resulting ontologies (i.e., converted datasets). We have measured the uncertainty resulting from the process according to common evaluation metrics. Experimental results clearly show that, regardless of the performance of the automated ontology matching process, a lack of semantics inevitably introduces errors and inaccuracies, resulting in significant uncertainty. On the other side, even when considering non-semantically enriched data, the experiment has demonstrated a certain degree of reliability. Such results, in general terms, point to properly designed semi-supervised integration approaches in a continuously evolving context in which technology is becoming higher and higher performing (e.g., by adopting large language models (LLMs)).

Structure of the Paper

Section 2 proposes an overview of a reference model for the knowledge-building process, while Section 3 addresses the intrinsically related uncertainty, and Section 4 deals with evaluation metrics. The conducted experiment and its discussion are reported in Sections 5 and 6, respectively.

2. Knowledge-Building Process

From a theoretical perspective, the knowledge-building process as understood in this paper is not limited to data integration, as it also includes semantic enrichment and annotations [2]. Different ontology-based solutions have been proposed to integrate data within a range of scientific and business contexts [13–15] as well as to support integration among systems [16].

However, the reference model for the empirical evaluation does not assume a semantically enriched context. Indeed, the process is not based on pre-developed ontologies but, rather, on a conversion of relational data into an ontological format. A process of semantic enrichment in this specific context would be domain/application specific and, therefore, not fully systematic.

Overall, the process can be centralized, meaning that a global schema can be adopted to provide integrated access to information [13]. Knowledge building through data integration is understood as the process of semantically integrating raw data [1]. Ontologies can be used to define rich data spaces (knowledge) in which semantics are formally specified. In this work, the knowledge-building process takes as input heterogeneous raw datasets (assumed to be a relational database) and returns as output an integrated semantic data space represented by a knowledge graph. As shown in Figure 1, the process is composed of the following three main steps: (*i*) conversion of datasets into ontologies, (*ii*) ontology matching, and (*iii*) ontology integration/merging.

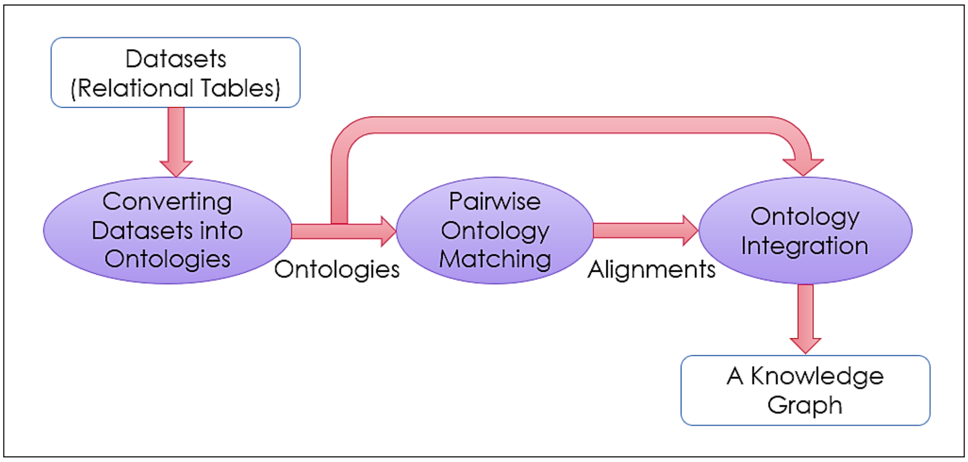


Figure 1. Overview of the reference model for knowledge building.

2.1. Dataset Conversion

Data spaces supported by Semantic Web technology assume data are available in a semantic format However, such a conversion of data into a semantic format is not always an obvious step, especially for non-technical users. In fact, the *Virtual Table model* [2] can be adopted in general terms to intuitively implement the conversion step. Along with RDF (Resource Description Framework) [12], OWL (Ontology Web Language) [17,18] is the most widely used language for representing and defining structured knowledge in the Semantic Web [10,11].

In this specific case, we believe that the conversion technique itself is not especially critical from a conceptual perspective, as it simply aims to preserve the original structure/logic as well as the original terminology. On the other side, as previously explained, the lack of semantic enrichment from relational data is a theoretical and practical limitation for the experiment. As rich semantics are not included in the input dataset, a process of enrichment based on contextual factors will be unreliable, even if manually performed. Indeed, associated metadata and related articles/documentation in the considered case study do not necessarily provide a comprehensive explanation to support the specification of formal semantics.

2.2. Ontology Matching

Also known as *ontology alignment*, ontology matching [19–27] is the process of finding semantic correspondences (mainly similarities) among entities from different ontologies. Each type of entity (*classes*, *object properties*, *datatype properties*, and *instances*) is normally matched in isolation, so *class-to-property*, *class-to-individual*, or *property-to-individual* correspondences are not addressed. Entity pairs that have the same name and meaning or have different names but the same meaning should be matched. Ontology matching is a critical step for ontology integration. Current ontology matching tools have become proficient in identifying equivalence correspondences between two ontologies. Ontology matching systems may use different matching algorithms.

2.3. Ontology Integration

Simple approaches for ontology integration include, among others, *Simple Union* [28]/ *Simple Merge* [29,30]. This consists of aggregating the input ontologies by adding bridging axioms that translate the alignments. The semantic correspondences of each alignment (resulting from the matching step) are transformed into bridging axioms in order to link the overlapping parts of the input ontologies [31,32].

3. Uncertainty Management

Automation naturally generates uncertainty, which is related to the accuracy of the mechanism. Such uncertainty becomes critical for the reliability of applications.

Focusing specifically on the process, uncertainty is mainly associated with the ontology matching phase and is commonly caused by two independent factors: the domain semantics and the strategy underlying the matcher.

Concerning the semantics, even manual approaches may lead to errors due to potential ambiguities. Indeed, people may not sufficiently understand the domain vocabulary and, therefore, may provide imprecise and incorrect correspondences. Additionally, certain domains can be intrinsically critical and naturally error-prone, as the intended correspondences maybe be unclear.

For example, concepts can have ambiguous semantics when they are closely related but not completely synonymous [33–35]. In this case, matching systems become uncertain [36]. Additional issues are related to the "Open World Assumption", where concepts are supposed to overlap and share a certain amount of common information (even if it is not yet specified) [36], as well as to different representations within the same domain or closely related domains [37].

As previously mentioned, significant uncertainty is introduced by matchers, regardless of their strategy (e.g., structural-, semantic-, or instance-based [38]).

Uncertainty management plays a crucial role in real-world applications, and it has been widely recognized as one of the key issues in data integration [39]. The actual challenge to manage uncertainty in such a context is to be able to maximize the reliability of the generated correspondences [40]. Techniques to reduce uncertainty (and, consequently, increase the quality of the output) may be manual or may assume a certain degree of automation.

One example from the latter class of solutions is alignment trimming [41], which aims to trim the alignment with a given contextual threshold in order to ensure that only the best correspondences are maintained in the alignment. Alternatively, different disambiguation techniques may be applied.

4. Evaluation Metrics

Several evaluation metrics have been defined and are commonly adopted within the research community to quantitatively assess the accuracy of alignments resulting from an automatic ontology matching system. Evaluation normally assumes a reference alignment, which is understood as the expected, intended, or "correct" alignment [42].

In practice, an alignment $\mathcal{A}$ returned by a given ontology matching tool can be compared to the reference alignment $\mathcal{R}$ by measuring the overlap between the two sets of correspondences [43]. The most common evaluation metrics for ontology alignment are adaptations of classical metrics from information retrieval.

In this section, we provide an overview of basic and advanced metrics as well as metrics to assess ambiguity.

4.1. Basic Evaluation Metrics

Basic metrics are characterized by their simplicity and are often combined to produce more sophisticated metrics. They are defined as follows [20]:

- **True positives** are correspondences that have been correctly found by an automatic ontology matching tool.

$$\boxed{\textbf{True Positives (TP)} = \mathcal{A} \cap \mathcal{R}}$$

- **False positives** are correspondences that have been falsely found by an automatic ontology matching tool.

$$\boxed{\textbf{False Positives (FP)} = \mathcal{A} - \mathcal{R}}$$

- **False negatives** are correct correspondences that have not been found by an automatic ontology matching tool.

$$\boxed{\textbf{False Negatives (FN)} = \mathcal{R} - \mathcal{A}}$$

- **True negatives** are false correspondences that have been correctly discarded by an automatic matching tool.

$$\boxed{\begin{array}{c}\textbf{True Negatives (TN)} = (E \times E') - (\mathcal{A} \cup \mathcal{R}) \\ E \text{ and } E' = \text{number of entities in the input ontologies } \mathcal{O}_1 \text{ and } \mathcal{O}_2. \\ (E \times E') = \text{set of all possible correspondences between } \mathcal{O}_1 \text{ and } \mathcal{O}_2.\end{array}}$$

Based on such definitions, the set of automatically identified correspondences comprises true positives and false positives ($\mathcal{A} = TP + FP$), and the set of expected correspondences is composed of true positives and true negatives ($\mathcal{R} = TP + TN$). Evidently, false positives and false negatives reduce the matching accuracy. Therefore, an efficient ontology matching system aims to minimize both of them.

4.2. Advanced Evaluation Metrics

Advanced evaluation metrics assume the following definitions [20,27,43,44]:

- **Precision** measures the "correctness" of an alignment. It reflects the share of correct correspondences among all the ones found. It is defined as the ratio of the number of correctly found correspondences (TP) over the total number of found correspondences (TP + FP). Given a *reference* alignment $\mathcal{R}$, the *Precision* of an alignment $\mathcal{A}$ is a function $P : \wedge \times \wedge \rightarrow [0,1]$ such that:

$$Precision = P(\mathcal{A}, \mathcal{R}) = \frac{|\mathcal{A} \cap \mathcal{R}|}{|\mathcal{A}|} = \frac{|TP|}{|TP| + |FP|}$$

- **Recall** measures the "completeness" of an alignment. It reflects the share of correct correspondences among all the expected correspondences. It is the ratio of the number of correctly found correspondences (TP) over the total number of expected correspondences to be found (TP + TN). Given a *reference* alignment $\mathcal{R}$, the *Recall* of an alignment $\mathcal{A}$ is a function $R : \wedge \times \wedge \rightarrow [0,1]$ such that:

$$Recall = R(\mathcal{A}, \mathcal{R}) = \frac{|\mathcal{A} \cap \mathcal{R}|}{|\mathcal{R}|} = \frac{|TP|}{|TP| + |TN|}$$

In the ideal case, *Precision* and *Recall* reach their highest value of 1.0.

- **Noise and Silence** are the complement measures of *Precision* and *Recall*. Given a *reference* alignment $\mathcal{R}$, the *Noise* and the *Silence* of an alignment $\mathcal{A}$ are, respectively, functions N and $S : \wedge \times \wedge \rightarrow [0,1]$ such that:

$$Noise = N(\mathcal{A}, \mathcal{R}) = 1 - Precision$$

$$Silence = S(\mathcal{A}, \mathcal{R}) = 1 - Recall$$

- **F-measure** combines *Precision* and *Recall*, as they cannot accurately assess the matching quality alone. Indeed, the ontology matching tool can have a high *Precision* and a low *Recall* and vice versa. *F-measure* is a combined metric that attaches different importance to *Precision* and *Recall*. Given a *reference* alignment $\mathcal{R}$ and a number α between 0 and 1 ($0 \leqslant \alpha \leqslant 1$), the *F-measure* of an alignment $\mathcal{A}$ is a function $F_\alpha : \wedge \times \wedge \rightarrow [0,1]$ such that:

$$F\text{-}measure(\alpha) = F_\alpha(\mathcal{A}, \mathcal{R}) = \frac{Precision \times Recall}{(1-\alpha) \times Precision + \alpha \times Recall}$$

The α parameter defines the relative balance between *Precision* and *Recall*, as higher α values give more importance to *Precision* with respect to *Recall*. When $\alpha = 1$, *F-measure* is equal to *Precision*, and when $\alpha = 0$, *F-measure* is equal to *Recall*. A value of $\alpha = 0.5$ defines the equal importance of both core metrics. Therefore, when $\alpha = 0.5$, *F-measure* becomes the harmonic mean of *Precision* and *Recall* as follows:

$$F1 = F\text{-}measure(0.5) = F_{0.5}(\mathcal{A}, \mathcal{R}) = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$F\text{-}measure(0.5)$, also called $F1$, is the most commonly used variant of $F\text{-}measure(\alpha)$ in IR since it balances the importance of *Precision* and *Recall* so that neither is compensated by the other.

Matching tools may need parameter tuning. In this case, *F-measure* is adopted as a driving factor to perform parameter tuning because values that maximize *F-measure* are considered the optimal ones. Hence, this metric is not only helpful in evaluating the quality of alignments but also in selecting input parameters for matching systems, such as the confidence threshold that returns the best *F-measure* value.

- **Overall** (or *matching accuracy* [45]) is explicitly developed for schema matching purposes, unlike the previously mentioned metrics. It measures the manual error correction effort. That is, it reflects the post-matching effort needed to fix the alignment by adding missing correspondences (FN) and removing false correspondences (FP). It is the ratio of the number of errors (FP + FN) over the total number of expected correspondences (TP + TN). Given a *reference* alignment $\mathcal{R}$, the *Overall* of an alignment $\mathcal{A}$ is a function $O : \wedge \times \wedge \rightarrow [-\infty, 1]$ such that:

$$Overall = O(\mathcal{A}, \mathcal{R}) = Recall \times \left(2 - \frac{1}{Precision}\right)$$

$$= 1 - \frac{|\mathcal{A} - \mathcal{R}| + |\mathcal{R} - \mathcal{A}|}{|\mathcal{R}|} = 1 - \frac{|FP| + |FN|}{|TP| + |TN|}$$

An *Overall* value ranges between $[-\infty, 1]$, where negative values are associated with "bad" matching performances. If an alignment $\mathcal{A}$ has a number of false positives higher than the number of true positives ($Precision < 0.5$), its *Overall* will have a negative value, which means that the alignment $\mathcal{A}$ is not worth the repair effort. Indeed, if more than half of the correspondences of $\mathcal{A}$ are false, the user would make less effort to manually match the ontologies from scratch than to manually modify the alignment of $\mathcal{A}$.

In the ideal case, when $Precision = Recall = 1$, *F-measure* and *Overall* reach their highest values of 1.0. It should be noted that *Overall* values are always lower than *F-measure* values [20].

4.3. Ambiguity Evaluation Metric

Ambiguity Degree [20] measures the "ambiguity" of an alignment. It is the proportion of ambiguous correspondences—i.e., entities that are matched to several entities. More formally, it is the proportion of entities from $\mathcal{O}_1$ that are matched to at least two entities from $\mathcal{O}_2$ and vice versa. The number of ambiguous correspondences in an alignment ($\#Amb$) is considered an absolute metric that varies according to the size of the alignment. Therefore, it is preferred to use a relative metric (%) reflecting the percentage of ambiguous correspondences in an alignment (regardless of the size of the alignment). The *Ambiguity Degree* of an alignment $\mathcal{A}$ is a function $Ambiguity : \wedge \times \wedge \rightarrow [0\%, 100\%]$ such that:

$$Ambiguity\ degree = Ambiguity(\mathcal{A}) = \frac{\left(|\#Amb(\mathcal{A})| \times 100\right)}{|\mathcal{A}|} = \frac{\left(|\#Amb(\mathcal{A})| \times 100\right)}{|TP| + |FP|}$$

A null value of the ambiguity degree (0%) is associated with no ambiguity in the alignment.

5. Case Study: Spatio–Temporal Alignment of Indicators

The experiment conducted addresses a case study of practical relevance that aims to integrate several independent indicators into a unique semantic data space. Successful integration is expected to provide a consistent view of the different indicators along the time and the spatial dimensions. Therefore, all semantically equivalent domain concepts should be matched.

In this section, we first provide a concise description of the experiment, with a focus on data sources, then, we address the tools adopted, and finally, we overview the results.

5.1. Experiment Description

The case study proposed in this paper has been inspired by the famous portal *Our World in Data* [46], which publishes and discusses heterogeneous indicators. In the portal, different datasets are published as independent `csv` files. For the most part, the terminology in the different files is supposed to be uniform.

The experiment assumes as an input the raw data as originally provided by the respective sources. It aims to achieve automatic integration. Details are reported in Table 1. As shown, each experiment involves a pair of input datasets.

Table 1. Experiment overview.

Experiment ID	Input Dataset 1	Input Dataset 2
Exp. 1	Countries of the World [47]	Country Profile Variables [48]
Exp. 2	Food Security Indicators (V_2.6) [49]	Prevalence of Undernourishment [50]
Exp. 3	Food Security Indicators (I_2.3) [51]	Prevalence of Undernourishment [50]
Exp. 4	Food Security Indicators (I_2.4) [51]	Prevalence of Undernourishment [50]
Exp. 5	Macroeconomic Data (GDP) [52]	GDP (current USD) [53]
Exp. 6	WDI Country [54]	WUP2018 Largest Cities [55]
Exp. 7	Total Life Expectancy—Historical [56]	Life Expectancy at Birth [57]
Exp. 8	Historical Gas Emissions [58]	List of Electoral Democracies [59]
Exp. 9	List of Electoral Democracies [59]	Sexual Violence in Childhood [60]
Exp. 10	Historical Gas Emissions [58]	Sexual Violence in Childhood [60]

The considered pairs can involve datasets belonging to the same domain as well as datasets from different domains.

Regardless of the meaning of the data, the original tables report the values of given indicators for different countries in different years. The actual structure may vary from case to case, but in most cases, it proposes typical patterns used to organize spatio–temporal data. For tables describing a particular indicator, rows represent the names of countries and columns represent years (or intervals of years) or vice versa. In some other cases, tables report the values of more than one indicator for different countries in a single year (or in a single interval of years): rows represent the names of countries, and columns represent indicators. More rarely, tables report the values of multiple indicators for different countries in different years: rows represent ID numbers (enumerated numbers/indexes), and columns contain a year or interval column, a country column, and indicator columns.

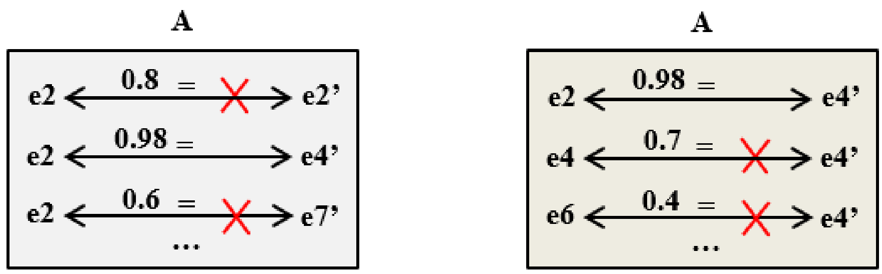
Overall, the test bed proposed cannot be considered critical, neither in terms of size nor complexity. It is, rather, characterized by a small scale and relatively low complexity. Input datasets are characterized by a number of columns that vary from 2 to 63 and rows in a range [46–540].

The experiment adopts available tools for alignment, as described later on in the paper.

After converting and matching the input dataset pairs, we will get the output alignments. First, we will evaluate the quality of the output alignments. Then, we will trim and/or disambiguate these alignments and evaluate them again. We aim to see how the alignment *trimming* and *disambiguation* processes can affect the quality and uncertainty of the output alignments.

For alignment *trimming*, we will use the Alignment API (https://moex.gitlabpages.inria.fr/alignapi/ (accessed on 20 October 2020)) [61,62]. That is a Java programming interface that facilitates the manipulation and evaluation of ontology alignments in RDF format. Given an alignment and a threshold value as an input, the Alignment API automatically trims the input alignment using the predefined method `cut()` and returns a new trimmed alignment. The trimmed output alignment will only contain correspondences that have a confidence measure greater than or equal to the chosen threshold value.

For alignment disambiguation, we apply an approach based on two major steps (Figure 2). First, we go through all the correspondences in the alignment, and we disambiguate each set of ambiguous correspondences having a source entity in common, as shown in Figure 2a; then, we go through all the correspondences again, and we disambiguate each set of ambiguous correspondences having a target entity in common (coming from $\mathcal{O}_2$), as shown in Figure 2b. In each step, disambiguation consists of keeping the strongest correspondence (i.e., the one with the highest similarity score) from the set of ambiguous correspondences and deleting the rest. If two ambiguous correspondences have exactly the same similarity score, we keep both of them.



(**a**) **Step 1**— left to right (**b**) **Step 2**— right to left

Figure 2. Alignment disambiguation—a simplified approach for the stable marriage model [63].

Finally, a reference alignment is manually provided by the authors for each target experiment for assessment purposes.

Given an alignment to be evaluated and a reference alignment as input, the Alignment API [62] automatically returns the scores of the basic evaluation metrics (TP, TN, FP, and

FN) as well as the scores of the advanced evaluation metrics (*Precision*, *Recall*, *F-measure*, *Overall*, *Noise*, and *Silence*), which reflect the quality of the input alignment.

As for the ambiguity evaluation metric, we create a simple algorithm in Java that takes as input an alignment and returns the score of the *ambiguity degree* metric, reflecting the ambiguity degree of that input alignment.

Many applications that require the combination of multiple indicators are unlikely to be error-tolerant, meaning that the resulting alignment is expected to be entirely correct. Despite its relative simplicity, the case study proposed is relevant in different practical contexts and application domains. More concretely, the target case study is characterized by the need to compose indicators dynamically, and the resulting integration is expected to be precise and accurate. Practical examples may be identified, among others, in the areas of urban planning (e.g., [64]) and global sustainable development (e.g., [65]). Further requirements may characterize specific systems, such as real-time environments for disaster management [66].

5.2. Tools

The experiment relied on different research tools that allowed the implementation of the intended process. The focus is on data conversion and ontology matching.

5.2.1. Dataset Conversion

The data conversion tool [2] supports the conversion of a given relational table into an OWL ontology. It applies a virtual table approach to facilitate such a process, assuming a supervised environment. The user interface allows users to straightforwardly import a relational table through simple copy and paste functions. The tool requires a relatively simple user input, which is considered to be realistic for the scale and complexity characterizing this concrete case study.

5.2.2. Ontology Matching

Ontology matching tools take as an input a pair of ontologies in OWL and return alignments in RDF. The *Ontology Alignment Evaluation Initiative* (http://oaei.ontologymatching.org/ (accessed on 20 October 2020)) (OAEI) is probably the most known international initiative for the enhancement and evaluation of ontology matching tools. Outstanding results from the OAEI community are presented yearly in the *Ontology Matching Workshop* (OM), which is co-located with the prestigious *International Semantic Web Conference* (ISWC).

We adopted two of the most popular tools within the OAEI community:

- **LogMap** (https://github.com/ernestojimenezruiz/logmap-matcher (accessed on 20 October 2020)) [67] is a highly scalable ontology matching tool. It performs an iterative process based on structural matching. To achieve scalability, LogMap relies on lexical and structural indexes of the input ontologies.
- **AML** (https://github.com/AgreementMakerLight/AML-Project (accessed on 20 October 2020)) [41] also approaches ontology matching looking at scalability and, additionally, provides a comprehensive user interface. AML combines different individual matchers to optimize the process.

We adopted more than one tool in the experiment to highlight the possible impact of tools on the final outcome. More concretely, we adopted the online web interface versions of LogMap (http://krrwebtools.cs.ox.ac.uk/logmap/ (accessed on 20 October 2020)) and AML v3.1 (https://github.com/AgreementMakerLight/AML-Project/releases (accessed on 20 October 2020)), which also include graphical user interfaces.

It is critical to highlight the constant evolution of the underlying technology. Indeed, in line with more generic trends, recent advances are based on (or include) large language models (LLMs), such as in [68]. While expectations are significant, we consider that such a technological trend is still at an early stage. Therefore, we have focused on the more consolidated technological landscape (pre-LLM).

5.3. Experimental Results

Results are summarized in Table 2, which summarizes the quality of ontology alignments resulting from each experiment. Such results may be considered initial results, as the reported values refer to the process before dis-disambiguation. The table reports evaluation scores as previously described.

Table 2. Initial results overview (pre-disambiguation).

Exp.	Tool	$\mathcal{R}$ [†]	$\mathcal{A}$ [‡]	*Amb* [§]	*Precision*	*Recall*	*F-Measure*	*Overall*	*Ambiguity*
Exp. 1	LogMap AML	267	212 270	2 186	0.995 0.255	0.79 0.258	0.881 0.256	0.786 −0.49	0.94% 68.88%
Exp. 2	LogMap AML	446	200 227	0 15	1.0 0.929	0.448 0.473	0.619 0.627	0.448 0.437	0% 6.6%
Exp. 3	LogMap AML	434	194 227	0 25	0.984 0.903	0.44 0.472	0.6 0.62	0.433 0.421	0% 11%
Exp. 4	LogMap AML	55	61 331	0 317	0.426 0.033	0.472 0.2	0.448 0.056	−0.163 −5.61	0% 95.77%
Exp. 5	LogMap AML	137	3 3	0 0	1.0 1.0	0.021 0.021	0.042 0.042	0.021 0.021	0% 0%
Exp. 6	LogMap AML	102	31 158	0 129	1.0 0.177	0.3 0.274	0.466 0.215	0.3 −1.0	0% 81.64%
Exp. 7	LogMap AML	521	290 3680	0 3628	0.996 0.013	0.554 0.095	0.712 0.023	0.552 −6.871	0% 98.58%
Exp. 8	LogMap AML	195	190 190	0 2	1.0 0.989	0.974 0.964	0.987 0.976	0.974 0.953	0% 1%
Exp. 9	LogMap AML	196	177 192	0 2	1.0 0.973	0.9 0.954	0.949 0.963	0.9 0.928	0% 1.04%
Exp. 10	LogMap AML	197	179 187	0 0	1.0 0.994	0.91 0.944	0.952 0.968	0.91 0.939	0% 0%

[†] $\mathcal{R}$: number of correspondences in the reference alignment $\mathcal{R}$ = number of expected correspondences $= TP + TN$. [‡] $\mathcal{A}$: number of correspondences in the output alignment $\mathcal{A}$ = number of detected correspondences $= TP + FP$. [§] *Amb*: number of ambiguous correspondences in the output alignment $\mathcal{A}$.

In Table 2, we highlight the differences in terms of advanced metrics—i.e., *Precision*, *Recall*, *F-measure*, and *Overall*—for the two adopted tools.

More concretely, for the *Precision* metric, both tools provide a significantly different result in Experiments 1, 4, 6, and 7 and minor differences in the remaining experiments. The *Recall* metric presents remarkable differences in Experiments 1, 4, and 7. *F-measure* values are quite different in Experiments 1, 4, and 7, while a pointless difference also shows up in Experiment 6. Finally, *Overall* returns strongly different values in Experiments 1 and 7 and a more limited divergence in Experiments 4 and 6. In terms of performance, the ontology matching process, on average, results in high values for *Precision*, with some evident exceptions (Experiment 4, especially), while *Recall*, *F-measure*, and *overall* values associated with the different experiments present important differences.

It is also worth mentioning that the last three experiments present a clear cross-domain focus. They are associated with higher performance in terms of both *Precision* and *Recall*. Intuitively, integrating ontologies from different domains is relatively easier than integrating ontologies within the same domain (or contiguous domains). Indeed, the latter case is often characterized by fine-grained terminology and heterogeneity.

5.3.1. Disambiguation

Tables 3 and 4 present resulting disambiguated alignments and trimmed alignments, respectively. Additionally, in Table 5, we report the resulting trimmed and disambiguated alignments.

Table 3. Results after alignment *disambiguation*.

Exp.	Tool	$\mathcal{R}$ †	$\mathcal{A}$ ‡	*Amb* §	*Precision*	*Recall*	*F-Measure*	*Overall*	*Ambiguity*
Exp. 1	LogMap	267	211	0	1.0	0.79	0.882	0.79	0%
	AML		118	0	0.559	0.247	0.342	0.052	0%
Exp. 2	LogMap	446	200	0	1.0	0.448	0.619	0.448	0%
	AML		218	0	0.958	0.468	0.629	0.448	0%
Exp. 3	LogMap	434	194	0	0.984	0.44	0.6	0.433	0%
	AML		210	0	0.952	0.46	0.621	0.437	0%
Exp. 4	LogMap	55	61	0	0.426	0.472	0.448	−0.163	0%
	AML		26	0	0.346	0.163	0.222	−0.145	0%
Exp. 5	LogMap	137	3	0	1.0	0.021	0.042	0.021	0%
	AML		3	0	1.0	0.021	0.042	0.021	0%
Exp. 6	LogMap	102	31	0	1.0	0.3	0.466	0.3	0%
	AML		40	0	0.7	0.274	0.394	0.156	0%
Exp. 7	LogMap	521	290	0	0.996	0.554	0.712	0.552	0%
	AML		171	0	0.292	0.095	0.144	−0.136	0%
Exp. 8	LogMap	195	190	0	1.0	0.974	0.987	0.974	0%
	AML		189	0	0.994	0.964	0.979	0.958	0%
Exp. 9	LogMap	196	177	0	1.0	0.9	0.949	0.9	0%
	AML		191	0	0.979	0.954	0.966	0.933	0%
Exp. 10	LogMap	197	179	0	1.0	0.91	0.952	0.91	0%
	AML		187	0	0.994	0.944	0.968	0.939	0%

† $\mathcal{R}$: expected correspondences. ‡ $\mathcal{A}$: detected correspondences. § *Amb*: ambiguous correspondences in $\mathcal{A}$.

Table 4. Results after alignment *trimming*.

Exp.	Tool	Threshold	$\mathcal{R}$ †	$\mathcal{A}$ ‡	*Amb* §	*Precision*	*Recall*	*F-Measure*	*Overall*	*Ambiguity*
Exp. 1	LogMap	–	267	212	2	0.995	0.79	0.881	0.786	0.94%
	AML	0.77		209	138	0.32	0.25	0.281	−0.28	66%
Exp. 2	LogMap	–	446	200	0	1.0	0.448	0.619	0.448	0%
	AML	–		227	15	0.929	0.473	0.627	0.437	6.6%
Exp. 3	LogMap	–	434	194	0	0.984	0.44	0.6	0.433	0%
	AML	–		227	25	0.903	0.472	0.62	0.421	11%
Exp. 4	LogMap	–	55	61	0	0.426	0.472	0.448	−0.163	0%
	AML	0.49		36	23	0.3	0.2	0.241	−0.254	63.88%
Exp. 5	LogMap	–	137	3	0	1.0	0.021	0.042	0.021	0%
	AML	–		3	0	1.0	0.021	0.042	0.021	0%
Exp. 6	LogMap	–	102	31	0	1.0	0.3	0.466	0.3	0%
	AML	0.46		36	8	0.777	0.274	0.4	0.196	22.22%
Exp. 7	LogMap	–	521	290	0	0.996	0.554	0.712	0.552	0%
	AML	0.51		50	0	0.98	0.094	0.171	0.092	0%
Exp. 8	LogMap	–	195	190	0	1.0	0.974	0.987	0.974	0%
	AML	–		190	2	0.989	0.964	0.976	0.953	1%
Exp. 9	LogMap	–	196	177	0	1.0	0.9	0.949	0.9	0%
	AML	–		192	2	0.973	0.954	0.963	0.928	1%
Exp. 10	LogMap	–	197	179	0	1.0	0.91	0.952	0.91	0%
	AML	–		187	0	0.994	0.944	0.968	0.939	0%

† $\mathcal{R}$: expected correspondences. ‡ $\mathcal{A}$: detected correspondences. § *Amb*: ambiguous correspondences in $\mathcal{A}$.

In Tables 3 and 5, the output no longer contains ambiguous correspondences. Therefore, the scores for the *ambiguity degree* are null. In Tables 4 and 5, we report the trimming *threshold* values considered in the different experiments. Thresholds were manually tuned to optimize the output.

Table 5. Results after alignment *trimming* and *disambiguation*.

Exp.	Tool	Threshold	$\mathcal{R}$ [†]	$\mathcal{A}$ [‡]	*Amb* [§]	*Precision*	*Recall*	*F-Measure*	*Overall*	*Ambiguity*
Exp. 1	LogMap	–	267	211	0	1.0	0.79	0.882	0.79	0%
	AML	0.77		90	0	0.722	0.243	0.364	0.149	0%
Exp. 2	LogMap	–	446	200	0	1.0	0.448	0.619	0.448	0%
	AML	–		218	0	0.958	0.468	0.629	0.448	0%
Exp. 3	LogMap	–	434	194	0	0.984	0.44	0.6	0.433	0%
	AML	–		210	0	0.952	0.46	0.621	0.437	0%
Exp. 4	LogMap	–	55	61	0	0.426	0.472	0.448	−0.163	0%
	AML	0.49		20	0	0.45	0.163	0.24	−0.03	0%
Exp. 5	LogMap	–	137	3	0	1.0	0.021	0.042	0.021	0%
	AML	–		3	0	1.0	0.021	0.042	0.021	0%
Exp. 6	LogMap	–	102	31	0	1.0	0.3	0.466	0.3	0%
	AML	0.46		29	0	0.965	0.274	0.427	0.264	0%
Exp. 7	LogMap	–	521	290	0	0.996	0.554	0.712	0.552	0%
	AML	0.51		50	0	0.98	0.094	0.171	0.092	0%
Exp. 8	LogMap	–	195	190	0	1.0	0.974	0.987	0.974	0%
	AML	–		189	0	0.994	0.964	0.979	0.958	0%
Exp. 9	LogMap	–	196	177	0	1.0	0.9	0.949	0.9	0%
	AML	–		191	0	0.979	0.954	0.966	0.933	0%
Exp. 10	LogMap	–	197	179	0	1.0	0.91	0.952	0.91	0%
	AML	–		187	0	0.994	0.944	0.968	0.939	0%

[†] $\mathcal{R}$: expected correspondences. [‡] $\mathcal{A}$: detected correspondences. [§] *Amb*: ambiguous correspondences in $\mathcal{A}$.

5.3.2. Metric Analysis

Tables 6–10 provide a contextual analysis by metric looking at the mainstream process. In these tables, the highest score is highlighted in bold for each experiment/tool.

Table 6. *Precision* analysis.

Exp.	Tool	Initial Alignment [1]	Disambiguation [2]	Trimming [3]	Dis. and Trim. [4]
Exp. 1	LogMap	0.995	**1.0**	0.995	**1.0**
	AML	0.255	0.559	0.32	**0.722**
Exp. 2	LogMap	1.0	1.0	1.0	1.0
	AML	0.929	**0.958**	0.929	**0.958**
Exp. 3	LogMap	0.984	0.984	0.984	0.984
	AML	0.903	**0.952**	0.903	**0.952**
Exp. 4	LogMap	0.426	0.426	0.426	0.426
	AML	0.033	0.346	0.3	**0.45**
Exp. 5	LogMap	1.0	1.0	1.0	1.0
	AML	1.0	1.0	1.0	1.0
Exp. 6	LogMap	1.0	1.0	1.0	1.0
	AML	0.177	0.7	0.777	**0.965**
Exp. 7	LogMap	0.996	0.996	0.996	0.996
	AML	0.013	0.292	**0.98**	**0.98**
Exp. 8	LogMap	1.0	1.0	1.0	1.0
	AML	0.989	**0.994**	0.989	**0.994**
Exp. 9	LogMap	1.0	1.0	1.0	1.0
	AML	0.973	**0.979**	0.973	**0.979**
Exp. 10	LogMap	1.0	1.0	1.0	1.0
	AML	0.994	0.994	0.994	0.994

[1] Table 2. [2] Table 3. [3] Table 4. [4] Table 5.

Regardless of the more or less relevant numerical variations, the experiment points out the high and holistic value of trimming and disambiguating for performance. Indeed, by excluding untrustworthy correspondences and minimizing the number of false positives, the alignment quality significantly increases.

Table 7. *Recall* analysis.

Exp.	Tool	Initial Alignment [1]	Disambiguation [2]	Trimming [3]	Dis. and Trim. [4]
Exp. 1	LogMap AML	0.79 **0.258**	0.79 0.247	0.79 0.25	0.79 0.243
Exp. 2	LogMap AML	0.448 **0.473**	0.448 0.468	0.448 **0.473**	0.448 0.468
Exp. 3	LogMap AML	0.44 **0.472**	0.44 0.46	0.44 **0.472**	0.44 0.46
Exp. 4	LogMap AML	0.472 **0.2**	0.472 0.163	0.472 **0.2**	0.472 0.163
Exp. 5	LogMap AML	0.021 0.021	0.021 0.021	0.021 0.021	0.021 0.021
Exp. 6	LogMap AML	0.3 0.274	0.3 0.274	0.3 0.274	0.3 0.274
Exp. 7	LogMap AML	0.554 **0.095**	0.554 **0.095**	0.554 0.094	0.554 0.094
Exp. 8	LogMap AML	0.974 0.964	0.974 0.964	0.974 0.964	0.974 0.964
Exp. 9	LogMap AML	0.9 0.954	0.9 0.954	0.9 0.954	0.9 0.954
Exp. 10	LogMap AML	0.91 0.944	0.91 0.944	0.91 0.944	0.91 0.944

[1] Table 2. [2] Table 3. [3] Table 4. [4] Table 5.

Table 8. *F-measure* analysis.

Exp.	Tool	Initial Alignment [1]	Disambiguation [2]	Trimming [3]	Dis. and Trim. [4]
Exp. 1	LogMap AML	0.881 0.256	**0.882** 0.342	0.881 0.281	**0.882** **0.364**
Exp. 2	LogMap AML	0.619 0.627	0.619 **0.629**	0.619 0.627	0.619 **0.629**
Exp. 3	LogMap AML	0.6 0.62	0.6 **0.621**	0.6 0.62	0.6 **0.621**
Exp. 4	LogMap AML	0.448 0.056	0.448 0.222	0.448 **0.241**	0.448 0.24
Exp. 5	LogMap AML	0.042 0.042	0.042 0.042	0.042 0.042	0.042 0.042
Exp. 6	LogMap AML	0.466 0.215	0.466 0.394	0.466 0.4	0.466 **0.427**
Exp. 7	LogMap AML	0.712 0.023	0.712 0.144	0.712 **0.171**	0.712 **0.171**
Exp. 8	LogMap AML	0.987 0.976	0.987 **0.979**	0.987 0.976	0.987 **0.979**
Exp. 9	LogMap AML	0.949 0.963	0.949 **0.966**	0.949 0.963	0.949 **0.966**
Exp. 10	LogMap AML	0.952 0.968	0.952 0.968	0.952 0.968	0.952 0.968

[1] Table 2. [2] Table 3. [3] Table 4. [4] Table 5.

At the same time, the experiment highlights a significant number of missing correspondences as well as a more contained number of false positives. Hence, the resulting alignment still presents relevant uncertainty, which may be considered absolutely critical in many applications and is an object of discussion in the next section.

Table 9. *Overall* analysis.

Exp.	Tool	Initial Alignment [1]	Disambiguation [2]	Trimming [3]	Dis. and Trim. [4]
Exp. 1	LogMap	0.786	**0.79**	0.786	**0.79**
	AML	− 0.49	0.052	− 0.28	**0.149**
Exp. 2	LogMap	0.448	0.448	0.448	0.448
	AML	0.437	**0.448**	0.437	**0.448**
Exp. 3	LogMap	0.433	0.433	0.433	0.433
	AML	0.421	**0.437**	0.421	**0.437**
Exp. 4	LogMap	− 0.163	− 0.163	− 0.163	− 0.163
	AML	− 5.61	− 0.145	− 0.254	**− 0.03**
Exp. 5	LogMap	0.021	0.021	0.021	0.021
	AML	0.021	0.021	0.021	0.021
Exp. 6	LogMap	0.303	0.303	0.303	0.303
	AML	−1.0	0.156	0.196	**0.264**
Exp. 7	LogMap	0.552	0.552	0.552	0.552
	AML	− 6.871	− 0.136	**0.092**	**0.092**
Exp. 8	LogMap	0.974	0.974	0.974	0.974
	AML	0.953	**0.958**	0.953	**0.958**
Exp. 9	LogMap	0.9	0.9	0.9	0.9
	AML	0.928	**0.933**	0.928	**0.933**
Exp. 10	LogMap	0.91	0.91	0.91	0.91
	AML	0.939	0.939	0.939	0.939

[1] Table 2. [2] Table 3. [3] Table 4. [4] Table 5.

Table 10. *Ambiguity* analysis.

Exp.	Tool	Initial Alignment [1]	Disambiguation [2]	Trimming [3]	Dis. and Trim. [4]
Exp. 1	LogMap	0.94 %		0.94 %	
	AML	68.88 %	**0%**	66 %	**0%**
Exp. 2	LogMap	0 %		0 %	
	AML	6.6 %	**0%**	6.6 %	**0%**
Exp. 3	LogMap	0 %		0 %	
	AML	11 %	**0%**	11 %	**0%**
Exp. 4	LogMap	0 %		0 %	
	AML	95.77 %	**0%**	63.88 %	**0%**
Exp. 5	LogMap	0 %		0 %	
	AML	0 %	**0%**	0 %	**0%**
Exp. 6	LogMap	0 %		0 %	
	AML	81.64 %	**0%**	22.22 %	**0%**
Exp. 7	LogMap	0 %		0 %	
	AML	98.58 %	**0%**	0 %	**0%**
Exp. 8	LogMap	0 %		0 %	
	AML	1 %	**0%**	1 %	**0%**
Exp. 9	LogMap	0 %		0 %	
	AML	1 %	**0%**	1 %	**0%**
Exp. 10	LogMap	0 %		0 %	
	AML	0 %	**0%**	0 %	**0%**

[1] Table 2. [2] Table 3. [3] Table 4. [4] Table 5.

6. Discussion

In qualitative terms, the uncertainty characterizing the experiments may be related to different factors, which are briefly discussed in this section. In general terms, the critical analysis should be considered in context: namely, looking at the techniques adopted as well as at the reference process and the associated underlining assumptions.

We remark on role of the limited semantics characterizing relational data that, without proper semantic enrichments, are likely to introduce intrinsic ambiguities. Moreover, the uncertainty generated across the knowledge building process is going to be propagated at an application level and affects the whole system engineering process (e.g., ontological modeling [69] and ontology-driven analysis [70]).

6.1. Granularity

First of all, the different granularity characterizing the raw datasets definitely plays a role, as the performance of the current ontology matching tools generally decrease when dealing with ontologies addressing a different level of abstraction/detail.

An example is reported in Table 11, where country names (e.g., *Sudan*) and parts of countries (e.g., *South Sudan*) coexist. In the first example in the table, "`Sudan`" in $\mathcal{O}_1$ is considered equivalent to "`Sudan`" in $\mathcal{O}_2$, which is false in this case; indeed, `Sudan` in $\mathcal{O}_2$ is actually North Sudan, and `Sudan` in $\mathcal{O}_1$ is the former Sudan that is composed of both the northern and southern parts. The entities "`Sudan`" and "`South Sudan`" from $\mathcal{O}_2$ should rather be sub-entities of "`Sudan`" from $\mathcal{O}_1$ (using a *subsumption* relation, not an *equivalence* relation). Ontology matching tools cannot identify such *subsumption* correspondences between entities.

Table 11. Examples of uncertainty resulting from different granularity.

Ontology $\mathcal{O}_1$	Ontology $\mathcal{O}_2$
• `Sudan`	• `Sudan` • `South_Sudan`
• `Gaza_Strip` • `West_Bank`	• `State_of_Palestine`
• `Netherlands_Antilles`	• `Sint_Maarten_(Dutch_part)` • `Bonaire,_Sint Eustatius_and_Saba`
• `Jersey` • `Guernsey`	• `Channel_Islands`
• `Agriculture`	• `Economy_Agriculture` • `Employment_Agriculture`

Similar errors may be hard to predict, as $\mathcal{O}_1$ and $\mathcal{O}_2$ do not present different granularity in general terms. A more appropriate level of data semantics may contribute to mitigate uncertainty by enabling ontology matching tools to work on a more extensive and accurate knowledge-base.

6.2. Complexity

Similar considerations apply to the time dimension, despite its apparent simplicity. For instance, in Experiment 3, $\mathcal{O}_1$, the time format assumes "year-year" intervals (e.g., "`97-1999`"), while $\mathcal{O}_2$ adopts a separate year format (e.g., "`1997`", "`1998`", "`1999`").

The complexity of the notation can be significant in ontologies. Some examples from the experiment are reported in Table 12. For example, the entity "`Near_East`" from $\mathcal{O}_1$ should be matched to the entity "`WesternAsia`" from $\mathcal{O}_2$. While the entity "`Asia_(EX._Near_East)`" should be matched to the union of the entities "`CentralAsia`", "`EasternAsia`", "`SouthernAsia`", and "`South-easternAsia`". While certain entities, for instance, "`Northern_Africa`" and "`NorthernAfrica`", were correctly matched, in general, the underling complexity of the notation is not correctly addressed.

As per previous considerations, enhanced data semantics would be extremely valuable in this case also to address complexity by progressively generating a potential semantically inter-operable data space.

Table 12. Examples of complexity in the notation.

Ontology $\mathcal{O}_1$	Ontology $\mathcal{O}_2$
• `Near_East` • `Asia_(EX._Near_East)`*	• `WesternAsia` • `CentralAsia` • `EasternAsia` • `SouthernAsia` • `South-easternAsia`

Table 12. *Cont.*

Ontology $\mathcal{O}_1$	Ontology $\mathcal{O}_2$
• Northern_Africa • Sub−Saharan_Africa	• NorthernAfrica • MiddleAfrica • EasternAfrica • WesternAfrica • SouthernAfrica
• Northern_America • Latin_Amer._and_Carib	• NorthernAmerica • CentralAmerica • SouthAmerica • Caribbean

* EX. = Except.

6.3. Domain and Ontology Design

There is no prescriptive or universally accepted approach to develop an ontological representation of a given domain. Therefore, ontologies may be designed according to different strategies and may present different patterns that reflect design choices. However, the tools adopted in the experiment do not address heterogeneous matching among entities. For instance, this means that a class can be matched with another class but not with an instance, an object property, or a data property. This can be considered to be an intrinsic limitation looking at the technological framework of the experiment.

Another potential factor is the inherent complexity of domain modeling. Some concrete examples from the experiments are reported in Table 13, where the concept "small" is intrinsically ambiguous and leads to significant uncertainty.

Table 13. Domain complexity.

Ontology $\mathcal{O}_1$	Ontology $\mathcal{O}_2$
• Small_island_developing_states • Caribbean	• Small_states • Pacific_island_small_states • Caribbean_small_states • Other_small_states

7. Conclusions

This paper deals with the empirical evaluation of a knowledge-building process based on automatic ontology matching systems. The focus is on the resulting uncertainty and its practical implications. More concretely, the case study aims to provide an integrated semantic data space from heterogeneous non-semantically enriched relational datasets by leveraging automatic matching tools (pre-LLM technology).

The results clearly show the significant uncertainty resulting from a lack of data semantics that reduces the accuracy of automatic ontology matching methods, even considering a relatively simple case study. Indeed, regardless of the adopted matching tool, experiments behold a significant number of detected false positives and false negatives, due mostly to the intrinsic ambiguity of non-semantically enriched data. This uncertainty is always propagated at an application level. However, such uncertainty needs to be considered in the context of a constant technological evolution and of applications, which may be characterized by a different degree of error tolerance.

Looking at the technology adopted in this concrete study and at the reference knowledge building process, realistically, we believe that a further consolidation of current knowledge engineering practices to enhance data semantics would be a determinant to significantly reduce uncertainty and to improve data integration system performance accordingly. Last but not least, more consistent data semantics would fully exploit the potentiality of AI-powered solutions.

Future work will aim to manage uncertainty in practice by inferring potential relationships between the characteristics of the raw data to be integrated and the resulting integrated data space. Such an extended process that assumes semantically enriched data can fully exploit the advanced capabilities provided by large language models. Further, we are exploring more holistic alignment-based approaches.

Author Contributions: Conceptualization, all authors; methodology, all authors; formal analysis, I.O.; investigation, all authors; data curation, I.O.; writing—original draft preparation, I.O.; writing—review and editing, I.O. and S.F.P.; supervision, S.F.P and S.B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The experiment was conducted on secondary data. Sources are reported in the paper.

Acknowledgments: We offer sincere acknowledgment to the anonymous reviewers, who provided constructive feedback and an opportunity for reflection. Addressing these comments was a determinant to improve the originally submitted version of the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lenzerini, M. Data integration: A theoretical perspective. In Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Madison, WI, USA, 3–5 June 2002; ACM: New York, NY, USA, 2002; pp. 233–246.
2. Pileggi, S.F.; Crain, H.; Yahia, S.B. An Ontological Approach to Knowledge Building by Data Integration. In Proceedings of the 20th International Conference on Computational Science (ICCS), Amsterdam, The Netherlands, 3–5 June 2020; Springer: Cham, Switzerland, 2020; Volume 12143, pp. 479–493. _35 [CrossRef]
3. García-Sánchez, F.; Colomo-Palacios, R.; Valencia-García, R. A social-semantic recommender system for advertisements. *Inf. Process. Manag.* **2020**, *57*, 102153. [CrossRef]
4. Zhang, X. Concept integration of document databases using different indexing languages. *Inf. Process. Manag.* **2006**, *42*, 121–135. [CrossRef]
5. Lin, W.C.; Chang, Y.C.; Chen, H.H. Integrating textual and visual information for cross-language image retrieval: A trans-media dictionary approach. *Inf. Process. Manag.* **2007**, *43*, 488–502. [CrossRef]
6. Noy, N.F. Semantic integration: A survey of ontology-based approaches. *ACM Sigmod Rec.* **2004**, *33*, 65–70. [CrossRef]
7. Delgado, M.; SáNchez, D.; MartıN-Bautista, M.J.; Vila, M.A. Mining association rules with improved semantics in medical databases. *Artif. Intell. Med.* **2001**, *21*, 241–245. [CrossRef]
8. Dou, D.; Wang, H.; Liu, H. Semantic data mining: A survey of ontology-based approaches. In Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), Anaheim, CA, USA, 7–9 February 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 244–251.
9. Guarino, N. Formal ontology, conceptual analysis and knowledge representation. *Intl. J. Hum.-Comput. Stud.* **1995**, *43*, 625–640. [CrossRef]
10. Berners-Lee, T.; Hendler, J.; Lassila, O. The semantic web. *Sci. Am.* **2001**, *284*, 34–43. [CrossRef]
11. Shadbolt, N.; Berners-Lee, T.; Hall, W. The semantic web revisited. *IEEE Intell. Syst.* **2006**, *21*, 96–101. [CrossRef]
12. Decker, S.; Melnik, S.; Van Harmelen, F.; Fensel, D.; Klein, M.; Broekstra, J.; Erdmann, M.; Horrocks, I. The semantic web: The roles of XML and RDF. *IEEE Internet Comput.* **2000**, *4*, 63–73. [CrossRef]
13. Gardner, S.P. Ontologies and semantic data integration. *Drug Discov. Today* **2005**, *10*, 1001–1007. [CrossRef]
14. Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L.J.; Eilbeck, K.; Ireland, A.; Mungall, C.J.; et al. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **2007**, *25*, 1251–1255. [CrossRef] [PubMed]
15. Zhang, H.; Guo, Y.; Li, Q.; George, T.J.; Shenkman, E.; Modave, F.; Bian, J. An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 41. [CrossRef] [PubMed]
16. Mate, S.; Köpcke, F.; Toddenroth, D.; Martin, M.; Prokosch, H.U.; Bürkle, T.; Ganslandt, T. Ontology-based data integration between clinical and research systems. *PLoS ONE* **2015**, *10*, e0122172.
17. McGuinness, D.L.; Van Harmelen, F. OWL Web ontology language overview. *W3C Recomm.* **2004**, *10* , 2004.
18. Motik, B.; Patel-Schneider, P.F.; Parsia, B.; Bock, C.; Fokoue, A.; Haase, P.; Hoekstra, R.; Horrocks, I.; Ruttenberg, A.; Sattler, U.; et al. OWL 2 web ontology language: Structural specification and functional-style syntax. *W3C Recomm.* **2012**, *27*, 159.
19. Choi, N.; Song, I.; Han, H. A survey on ontology mapping. *ACM SIGMOD Rec.* **2006**, *35*, 34–41. [CrossRef]

20. Euzenat, J.; Shvaiko, P. *Ontology Matching*, 2nd ed.; Springer: Heidelberg, Germany, 2013.
21. Granitzer, M.; Sabol, V.; Onn, K.W.; Lukose, D.; Tochtermann, K. Ontology Alignment—A Survey with Focus on Visually Supported Semi-Automatic Techniques. *Future Internet* **2010**, *2*, 238–258. [CrossRef]
22. Kalfoglou, Y.; Schorlemmer, W.M. Ontology Mapping: The state of the art. *Knowl. Eng. Rev.* **2003**, *18*, 1–31. [CrossRef]
23. Ochieng, P.; Kyanda, S. Large-Scale Ontology Matching: State-of-the-Art Analysis. *ACM Comput. Surv.* **2018**, *51*, 1–35. [CrossRef]
24. Otero-Cerdeira, L.; Rodríguez-Martínez, F.J.; Gómez-Rodríguez, A. Ontology Matching: A Literature Review. *Expert Syst. Appl.* **2015**, *42*, 949–971. [CrossRef]
25. Rahm, E. Towards Large-Scale Schema and Ontology Matching. In *Schema Matching and Mapping*; Data-Centric Systems and Applications; Springer: Berlin/Heidelberg, Germany, 2011; pp. 3–27. _1 [CrossRef]
26. Shvaiko, P.; Euzenat, J. Ontology Matching: State of the Art and Future Challenges. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 158–176. [CrossRef]
27. Mohammadi, M.; Rezaei, J. Evaluating and comparing ontology alignment systems: An MCDM approach. *J. Web Semant.* **2020**, *64*, 100592. [CrossRef]
28. Raunich, S.; Rahm, E. Towards a Benchmark for Ontology Merging. In Proceedings of the OTM Confederated International Workshops: On the Move to Meaningful Internet Systems, Rome, Italy, 10–14 September 2012; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7567; pp. 124–133. _20 [CrossRef]
29. Osman, I.; Ben Yahia, S.; Diallo, G. Ontology integration: Approaches and challenging issues. *Inf. Fusion* **2021**, *71*, 38–63. [CrossRef]
30. Osman, I.; Pileggi, S.F.; Yahia, S.B.; Diallo, G. An Alignment-Based Implementation of a Holistic Ontology Integration Method. *MethodsX* **2021**, *8*, 101460. [CrossRef] [PubMed]
31. Predoiu, L.; Feier, C.; Scharffe, F.; de Bruijn, J.; Martín-Recuerda, F.; Manov, D.; Ehrig, M. *State-of-the-Art Survey on Ontology Merging and Aligning V2*; EU-IST Integrated Project IST-2003-506826 SEKT; 2005 ; p. 79.
32. De Bruijn, J.; Ehrig, M.; Feier, C.; Martín-Recuerda, F.; Scharffe, F.; Weiten, M. Ontology mediation, merging and aligning. *Semant. Web Technol.* **2006**, 95–113. [CrossRef]
33. Po, L.; Sorrentino, S. Automatic generation of probabilistic relationships for improving schema matching. *Inf. Syst.* **2011**, *36*, 192–208. [CrossRef]
34. Jan, S.; Li, M.; Al-Sultany, G.; Al-Raweshidy, H. Ontology alignment using rough sets. In Proceedings of the 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shanghai, China, 26–28 July 2011; IEEE: Piscataway, NJ, USA, 2011; Volume 4; pp. 2683–2686.
35. Blasch, E.P.; Dorion, É.; Valin, P.; Bossé, E. Ontology alignment using relative entropy for semantic uncertainty analysis. In Proceedings of the IEEE 2010 National Aerospace & Electronics Conference, Dayton, OH, USA, 14–16 July 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 140–148.
36. Zhang, Y.; Panangadan, A.V.; Prasanna, V.K. UFOM: Unified fuzzy ontology matching. In Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), Redwood City, CA, USA, 13–15 August 2014; IEEE: Piscataway, NJ, USA, 2014, pp. 787–794.
37. Bharambe, U.; Durbha, S.S.; King, R.L. Geospatial ontologies matching: An information theoretic approach. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 2918–2921.
38. Marie, A.; Gal, A. Managing uncertainty in schema matcher ensembles. In Proceedings of the International Conference on Scalable Uncertainty Management (SUM), Washington, DC, USA, 10–12 October 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 60–73.
39. Gal, A. Managing uncertainty in schema matching with top-k schema mappings. *J. Data Semant. VI* **2006**, *6*, 90–114.
40. Dong, X.L.; Halevy, A.; Yu, C. Data Integration with Uncertainty. *VLDB J.* **2009**, *18*, 469–500. [CrossRef]
41. Faria, D.; Pesquita, C.; Santos, E.; Palmonari, M.; Cruz, I.F.; Couto, F.M. The AgreementMakerLight Ontology Matching System. In Proceedings of the OTM 2013 Conferences—Confederated International Conferences, Graz, Austria, 9–13 September 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 527–541. [CrossRef]
42. Do, H.H.; Melnik, S.; Rahm, E. Comparison of Schema Matching Evaluations. In Proceedings of the Web, Web-Services, and Database Systems, NODe 2002 Web and Database-Related Workshops, Erfurt, Germany, 7–10 October 2002; Springer: Berlin/Heidelberg, Germany, 2002; Volume 2593, pp. 221–237. [CrossRef]
43. Ehrig, M.; Euzenat, J. Relaxed Precision and Recall for Ontology Matching. In Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies, CEUR-WS.org, Banff, AB, Canada, 2 October 2005; Volume 156.
44. Euzenat, J. Semantic Precision and Recall for Ontology Alignment Evaluation. In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India, 6–12 January 2007; pp. 348–353.
45. Melnik, S.; Garcia-Molina, H.; Rahm, E. Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching. In Proceedings of the 18th International Conference on Data Engineering, San Jose, CA, USA, 26 February–March 2002; IEEE Computer Society: Piscataway, NJ, USA, 2002; pp. 117–128. [CrossRef]
46. Lab, G.C.D. Our World in Data. Available online: https://ourworldindata.org (accessed on 6 October 2020).
47. Lasso, F. Countries of the World: Country Names Linked to Region, Population, Area, Size, GDP, Mortality, and More. Kaggle. 2018. Available online: https://www.kaggle.com/datasets/fernandol/countries-of-the-world (accessed on 21 October 2020).

48. SRK. Country Statistics—UNData: Dataset of Economic, Social, Infra and Environmental Indicators. Kaggle. 2018. Available online: https://www.kaggle.com/datasets/sudalairajkumar/undata-country-profiles (accessed on 21 October 2020).
49. FAO. Statistics—Food Security Indicators. Food and Agriculture Organization of the United Nations. 2014. Available online: https://www.fao.org/fileadmin/user_upload/food-security-capacity-building/docs/Nutrition/NairobiWorkshop/5.WFP_IndicatorsFSandNutIntegration.pdf (accessed on 21 October 2020).
50. WorldBank. Prevalence of Undernourishment (% of Population). The World Bank Group. 2020. Available online: https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SN.ITK.DEFC.ZS (accessed on 21 October 2020).
51. FAO. Statistics—Food Security Indicators. Food and Agriculture Organization of the United Nations. 2020. Available online: https://openknowledge.fao.org/server/api/core/bitstreams/6ca1510c-9341-4d6a-b285-5f5e8743cc46/content/sofi-2022/food-security-nutrition-indicators.html (accessed on 21 October 2020).
52. Barro, R.J. Barro-Ursua Macroeconomic Data. 2010. Available online: https://scholar.harvard.edu/barro/publications (accessed on 21 October 2020).
53. WorldBank. GDP (Current US$). The World Bank Group. 2020. Available online: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD (accessed on 21 October 2020).
54. WorldBank. World Development Indicators (WDI). The World Bank Group. 2020. Available online: https://databank.worldbank.org/source/world-development-indicators (accessed on 21 October 2020).
55. WUP. World Urbanization Prospects: The 2018 Revision, Online Edition. United Nations, Department of Economic and Social Affairs, Population Division. 2018. Available online: https://population.un.org/wup/ (accessed on 21 October 2020).
56. Zijdeman, R.; Ribeira da Silva, F. Life Expectancy at Birth (Total). IISH Data Collection, V1. 2015. Available online: https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/LKYT53 (accessed on 21 October 2020).
57. WorldBank. Life Expectancy at Birth, Total (Years). The World Bank Group. 2020. Available online: https://data.worldbank.org/indicator/SP.DYN.LE00.IN (accessed on 21 October 2020).
58. ClimateWatch. Historical Emissions. Data Explorer. 2020. Available online: https://www.climatewatchdata.org/data-explorer/historical-emissions?historical-emissions-data-sources=climate-watch&historical-emissions-gases=all-ghg&historical-emissions-regions=All%20Selected&historical-emissions-sectors=total-including-lucf%2Ctotal-including-lucf&page=1&sort_col=2020&sort_dir=ASC (accessed on 21 October 2020).
59. FreedomHouse. 2020 List of Electoral Democracies. Freedom in the World, FIW. 2020. Available online: https://freedomhouse.org/report/freedom-world (accessed on 21 October 2020).
60. Unicef. Sexual Violence in Childhood. UNICEF DATA, 2020. Available online: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://data.unicef.org/wp-content/uploads/2020/06/Sexual-violence-in-childhood-database_June-2020.xls&ved=2ahUKEwi03tXc7rGGAxUhja8BHYp2HDcQFnoECC0QAQ&usg=AOvVaw2ceQzKhiWHX3Nle116RVFD (accessed on 21 October 2020).
61. Euzenat, J. An API for Ontology Alignment. In Proceedings of the Third International Semantic Web Conference, ISWC 2004, Hiroshima, Japan, 7–11 November 2004; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3298, pp. 698–712. _48 [CrossRef]
62. David, J.; Euzenat, J.; Scharffe, F.; dos Santos, C.T. The Alignment API 4.0. *Semant. Web* **2011**, *2*, 3–10. [CrossRef]
63. Gale, D.; Shapley, L.S. College admissions and the stability of marriage. *Am. Math. Mon.* **1962**, *69*, 9–15. [CrossRef]
64. Pileggi, S.F.; Hunter, J. An ontological approach to dynamic fine-grained Urban Indicators. *Procedia Comput. Sci.* **2017**, *108*, 2059–2068. [CrossRef]
65. Pileggi, S.F. Is the World Becoming a Better or a Worse Place? A Data-Driven Analysis. *Sustainability* **2020**, *12*, 88. [CrossRef]
66. Liu, X.; Zhang, J.; Ma, D.; Bao, Y.; Tong, Z.; Liu, X. Dynamic risk assessment of drought disaster for maize based on integrating multi-sources data in the region of the northwest of Liaoning Province, China. *Nat. Hazards* **2013**, *65*, 1393–1409. [CrossRef]
67. Jiménez-Ruiz, E.; Grau, B.C. Logmap: Logic-based and Scalable Ontology Matching. In Proceedings of the 10th International Semantic Web Conference (ISWC), Bonn, Germany, 23–27 October 2011; Volume 7031, pp. 273–288. [CrossRef]
68. Hertling, S.; Paulheim, H. OLaLa: Ontology matching with large language models. In Proceedings of the 12th Knowledge Capture Conference 2023, Pensacola, FL, USA, 5–7 December 2023; pp. 131–139.
69. Pileggi, S.F. Ontological Modelling and Social Networks: From Expert Validation to Consolidated Domains. In Proceedings of the International Conference on Computational Science, Prague, Czech Republic, 3–5 July 2023; Springer: Cham, Switzerland, 2023; pp. 672–687.
70. Sheth, A.P.; Ramakrishnan, C. Semantic (Web) technology in action: Ontology driven information systems for search, integration, and analysis. *IEEE Data Eng. Bull.* **2003**, *26*, 40–48.