

Introducing the HOBBIT platform into the Ontology Alignment Evaluation Campaign

Ernesto Jiménez-Ruiz^{1,2}, Tzanina Saveta³, Ondřej Zamazal⁴, Sven Hertling⁵, Michael Röder⁶, Irini Fundulaki³, Axel-Cyrille Ngonga Ngomo⁶, Mohamed Ahmed Sherif⁶, Amina Annane^{7,8}, Zohra Bellahsene⁸, Sadok Ben Yahia⁹, Gayo Diallo¹⁰, Daniel Faria¹¹, Marouen Kachroudi⁹, Abderrahmane Khiat¹², Patrick Lambrix¹³, Huanyu Li¹³, Maximilian Mackeprang¹², Majid Mohammadi¹⁴, Maciej Rybinski¹⁵, Booma Sowkarthiga Balasubramani¹⁶ and Cassia Trojahn¹⁷

¹ The Alan Turing Institute, London, United Kingdom

² Department of Informatics, University of Oslo, Norway

³ Institute of Computer Science - FORTH, Greece

⁴ University of Economics, Prague, Czech Republic

⁵ Data and Web Science Group, University of Mannheim, Germany

⁶ Paderborn University, Data Science Group, Pohlweg 51, D-33098 Paderborn, Germany

⁷ Ecole nationale Supérieure d'Informatique, Alger, Algeria

⁸ LIRMM, Université de Montpellier, CNRS, Montpellier, France

⁹ Université de Tunis El Manar, LIPAH-LR 1 ES14, 2092, Tunis, Tunisie

¹⁰ BPH Center - INSERM U1219, Team ERIAS & LaBRI UMR5800, University of Bordeaux

¹¹ Instituto Gulbenkian de Ciência, Portugal

¹² Human-Centered Computing, Freie Universität Berlin, 4195 Berlin, Germany

¹³ Linköping University & Swedish e-Science Research Centre, Linköping, Sweden

¹⁴ Delft University of Technology, The Netherlands

¹⁵ Dpto. Lenguajes y Ciencias de la Computación, University of Malaga, Spain

¹⁶ ADVIS Lab, Department of Computer Science, University of Illinois at Chicago, USA

¹⁷ Institut de Recherche en Informatique de Toulouse, France

Abstract. This paper describes the Ontology Alignment Evaluation Initiative 2017.5 pre-campaign. Like in 2012, when we transitioned the evaluation to the SEALS platform, we have also conducted a pre-campaign to assess the feasibility of moving to the HOBBIT platform. We report the experiences of this pre-campaign and discuss the future steps for the OAEI.

1 Introduction

The Ontology Alignment Evaluation Initiative¹ (OAEI) is a coordinated international initiative which organizes the evaluation of ontology matching systems [1,2]. The main goal of the OAEI is to compare systems and algorithms openly and on the same basis to allow anyone to draw conclusions about the best matching strategies. Furthermore, our ambition is to help tool developers to improve their systems through such evaluations.

The initiative started in 2004, and from 2006 until the present, the OAEI campaigns were held at the Ontology Matching workshop, collocated with the ISWC conference. Since 2011, we have been using an environment for automatically processing evaluations which was developed within the SEALS (Semantic Evaluation At Large Scale)

¹ <http://oaei.ontologymatching.org>

project². SEALS provided a software infrastructure for automatically executing evaluations and evaluation campaigns for typical semantic web tools, including ontology matching. In the OAEI 2017, a novel evaluation environment called HOBBIT was adopted for the novel HOBBIT Link Discovery track. In that OAEI campaign, all systems were executed under the SEALS client in all other tracks.

The good experience of the 2017 Link Discovery (*e.g.*, novel platform, online evaluation, automatic generation of result tables, attraction of link discovery developers, customization of the matching requirements of a benchmark task) track motivated the interest in assessing the possibility of transitioning the whole OAEI evaluation to HOBBIT. To that end, we decided to set-up an OAEI pre-campaign, as happened in the OAEI 2011.5 when the OAEI moved to SEALS,³ to evaluate potential risks and challenges. The nature of the link discovery tracks is different from the (traditional) OAEI tracks and we foresaw sources of uncertainty with respect to: (*i*) the use of a new evaluation environment, (*ii*) the adaptation of tracks with multiple tasks (like *multifarm*), (*iii*) the introduction of Docker to organisers and participants, (*iv*) the inclusion of interactivity capabilities, and (*v*) the storage of results. The objective of the Ontology Alignment Evaluation Initiative 2017.5 pre-campaign was, therefore, to evaluate the feasibility of moving some (traditional) OAEI tracks to the HOBBIT platform. In this paper, we report the experiences of this pre-campaign and future steps of the OAEI.

The remainder of the paper is organised as follows. Section 2 introduces the HOBBIT platform. In Section 3, we present the overall evaluation methodology that has been used. Section 4 describes the evaluation data sets and Section 5 the participating systems. Section 6 overviews the lessons learned from the campaign; and finally, Section 7 summarizes the conclusions of this experience and discusses future plans for the OAEI.

2 HOBBIT platform

The HOBBIT platform is a generic, modular and distributed platform for Big Linked Data systems. It was designed to enable Big Data practitioners and Linked Data users to benchmark all steps of the data lifecycle at scale, *i.e.*, with all necessary contemplations of volume, velocity, value and veracity necessary to benchmark real applications. Some of its most important features within the context of link discovery include the support of (*i*) benchmarks that focus on the evaluation of the quality of a system using single consecutive requests as well as (*ii*) benchmarks aiming at evaluating the efficiency of Big Linked Data solutions, *e.g.*, by generating distributed parallel requests leading to a high workload. The HOBBIT project⁴ designed and develops the HOBBIT platform with the aim of providing an open-source, extensible, FAIR⁵ and scalable evaluation platform (in a fashion akin to GERBIL [3]) along with corresponding benchmarks and mimicking algorithms for real data sources of industrial scale. The platform being open-source means that it can be downloaded and installed locally for tests. The online instance of the platform allows (*i*) running public challenges and (*ii*) making sure that even people without the required infrastructure are able to run the benchmarks they are interested in. The platform, as well as the benchmarks that are designed and implemented in HOBBIT

² <http://www.seals-project.eu>

³ <http://oaei.ontologymatching.org/2011.5/>

⁴ <http://project-hobbit.eu>

⁵ Findable, Accessible, Interoperable and Reusable

are modelled as actors with which the platform interacts. The use cases relevant for end users supported by the platform include:

- *Benchmarking a System*: the user can select a benchmark to test a system. The platform loads appropriate configuration parameters for the benchmark, as well as the list of available systems for this benchmark. The user configures the benchmark and selects one of the available systems to benchmark.
- *Showing and Comparing Benchmark Results*: the user can view the results of a single benchmark run or select multiple, e.g., to compare several systems that have been evaluated with the same benchmark.
- *Adding a System*: the user adds the system that needs to be benchmarked in the platform by providing a docker image of the system and a system adapter which serves as a proxy between the benchmark and the system.

Figure 1 shows the layout of the HOBBIT platform components and how the different parts interact. The platform can be separated into two parts. The first part comprises platform components that are always running (right hand side of Figure 1). The second part contains all components that belong to a certain experiment (left hand side of Figure 1), i.e., the benchmark components as well as the benchmarked system.

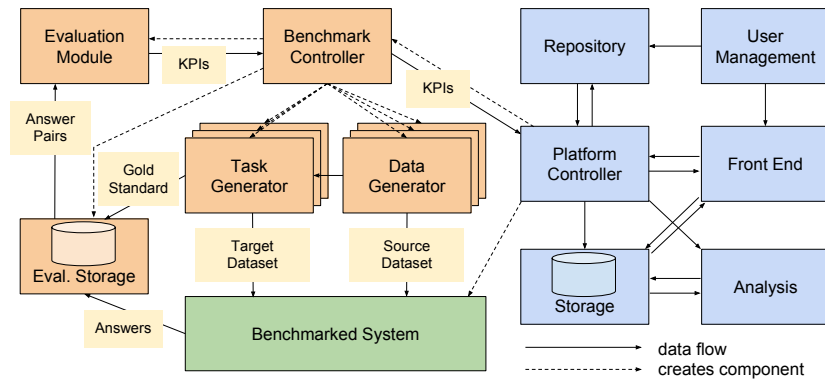


Fig. 1. Interaction of the components of HOBBIT Platform

The *Platform Controller* makes sure that the benchmark chosen by the user can be started and ensures that all nodes of the cluster are available. It communicates with the system to be benchmarked, ensures that it is working properly and generates the benchmark controller that is responsible for producing the data and task generators as well as the evaluation storage. The *Data Generator* produces the source dataset that is sent to the *Benchmarked System*, and the target dataset as well as the Gold Standard which are sent to the *Task Generator*. The *Task Generator* sends the target dataset to the *Benchmarked System* and forwards the Gold Standard to the *Evaluation Storage*. When the system finishes its task, it sends the answers to the *Evaluation Storage*. The *Evaluation Module* receives the system and the Gold Standard answers and returns the Key Performance Indicators (KPIs) for the experiment.

3 Methodology

The OAEI campaigns are typically divided into three phases: (i) preparation phase (datasets are prepared), (ii) execution phase (systems are tuned), and (iii) evaluation phase (systems are evaluated). In this OAEI pre-campaign, we focused on the preparation and execution phases given the time constraints and the challenges encountered during the migration to the HOBBIT platform.

3.1 Preparation phase

The preparation phase for the OAEI 2017.5 pre-campaign, unlike recent OAEI campaigns, was more demanding as the OAEI track organisers were required to migrate the SEALS datasets following the novel HOBBIT specifications. We provided the benchmark definitions for the (i) Largebio⁶ and (ii) Link discovery⁷ tracks to make the transition smoother. These (reference) datasets were made available by the end of January 2018. Next, we provide a brief summary of the main components of a HOBBIT benchmark.

HOBBIT benchmark definition. The HOBBIT workflow and format of benchmarks is generic as the platform was designed to accommodate benchmarks across the whole of the Linked Data lifecycle. This flexibility adds some complexity with respect to the SEALS benchmark generation. Note that, since in the OAEI multi-tasks benchmarks the source dataset may change, we have slightly modified the general HOBBIT workflow depicted in Figure 1. In the OAEI workflow, the TaskGenerator deals with both the source and target datasets to generate a Task. A benchmark is composed by the following classes:

BenchmarkController is the main class of the benchmark where the general benchmark execution workflow is specified.

DataGenerator generates the benchmark datasets (e.g., input ontologies and alignments) and prepares the datasets for the TaskGenerator. For multiple-task benchmarks it also deals with the preparation of queue names to be sent to the system.

Task includes the information of the source and target datasets and the expected results together with some other parameters like which type of entity should be matched (e.g., only classes).

TaskGenerator deals with the generation of the task(s) and sends the task(s) to the system and the *EvaluationModule*.

EvaluationModule compares the expected results (e.g., reference alignment) provided by the *TaskGenerator* and the computed results by a system, and generates the KPIs.

Each benchmark is also associated to a metadata file⁸ where the docker images of the benchmark are referenced, the KPIs defined, and the name of the benchmark's API specified (e.g., *bench:LargebioAPI*).

⁶ <https://gitlab.com/ernesto.jimenez.ruiz/largebio>

⁷ <https://github.com/hobbit-project/SpatialBenchmark/>

⁸ Metadata for largebio: <https://git.project-hobbit.eu/ernestoj/largebio>

Table 1. OAEI 2017.5 Benchmarks: HOBBIT APIs and KPIs. Each benchmark has its own API as they may define different input parameters. Only systems compliant with (*i.e.*, implementing) the API will be evaluated under a given benchmark.

Track	API	KPIs
Conference	bench:ConferenceAPI	Precision, Recall, F-measure
Anatomy	bench:AnatomyAPI	Precision, Recall, F-measure, Recall+
Largebio	bench:LargebioAPI	Precision, Recall, F-measure
Spimbench	bench:spimbenchAPI	Precision, Recall, F-measure
Link discovery	bench:LinkingAPI	Precision, Recall, F-measure

OAEI 2017.5 tracks. The preparation phase was complete in early March 2018 and led to four novel tracks running under the HOBBIT platform: *conference*, *anatomy*, *largebio*, and *instance matching - spimbench*. Note that the *link discovery* track was already running under HOBBIT in the OAEI 2017 campaign. The benchmarks are (briefly) described in Section 4.

3.2 Execution phase

The execution phase also brought the new challenge to developers of implementing a system compliant with the HOBBIT specifications. We provided the following sources of instruction to support system developers with the integration with HOBBIT: (*i*) General HOBBIT instructions,⁹ (*ii*) LogMap’s example implementing the interfaces for the *conference*, *anatomy*, *largebio* and *spimbench* tracks¹⁰, and (*iii*) the Maven framework to facilitate the wrapping of systems.¹¹

HOBBIT system definition. The interface of a system is defined via the *SystemAdapter* class (e.g., *LogMapSystemAdapter*). This class receives the dataset definition from the *DataGenerator* of a benchmark (e.g. set of tasks and matching requirements) and the individual tasks (source and target datasets) from the *TaskGenerator* of a benchmark. The results (e.g., a file containing the mappings in RDF Alignment format) are sent to the benchmark’s *EvaluationModule*. The system adapter class communicates to the benchmark classes in a special way since it is submitted to the HOBBIT platform as a docker image. Each system is also associated to a metadata file,¹² which explicitly mentions the APIs the system implements (e.g. `hobbit:implementsAPI bench:LargebioAPI`). This enables the automation of the evaluation of the OAEI benchmarks.

OAEI 2017.5 participation. Ten systems were registered to participate in the OAEI 2017.5 campaign in March 2018. Only eight of them reported results or experiences during April and May: *OntoIdea*, *LogMap*, *SANOM*, *DisMatch*, *KEPLER*, *YAM-BIO*, *AML* and *RADON*. The participating system and proof-of-concept results are (briefly) presented in Section 5.

⁹ <https://project-hobbit.eu/challenges/oeai2017-5/oeai2017-5-tasks/>

¹⁰ LogMap [4]: <https://gitlab.com/ernesto.jimenez.ruiz/logmap-hobbit>

¹¹ Maven framework: <https://github.com/sven-h/ontMatchingHobbit>

¹² LogMap’s metadata: <https://git.project-hobbit.eu/ernestoj/logmapsystem>

4 Benchmarks

The OAEI 2017.5 pre-campaign included five tracks: *conference*, *anatomy*, *largebio*, *instance matching* - *spimbench*, and *link discovery*. Table 1 provides a summary of the benchmarks. This pre-campaign did not include the *multifarm* and the *interactive* tracks. In the case of the *multifarm* track the main restriction was to move thousands of matching tasks to a new environment. While for the *interactive* tracks the main limitation was technological as the inclusion of an “oracle” requires significant modifications on the HOBBIT pipeline. Next we briefly describe the datasets of the OAEI 2017.5 benchmarks.

Anatomy track. This track consists of finding an alignment between the Adult Mouse Anatomy ontology (AMA) and a part of the National Cancer Institute (NCI) Thesaurus (NCI-A). This data set has been used since 2007 with some improvements over the years [5]. The AMA ontology contains 2,744, while the NCI-A contains 3,304 concepts describing the human anatomy. Systems participating in the anatomy track are evaluated in terms of runtime, precision, recall and F-measure. In addition, the anatomy track measures the systems’ ability to find non-trivial correspondences (recall+) and checks whether the systems generate coherent alignments.

Conference track. This track consists of 21 test cases with ontologies from the domain of organising conferences. The conference track has been used since 2006 and it was gradually improved [6]. The advantage of the conference domain is the fact that it is generally understandable. The ontologies were developed independently and based on different resources, thus they capture the issues in organising conferences from different points of view and using different nomenclature. Finally, ontologies within this track are of small-medium size and relatively rich in OWL 2 axioms.

Largebio track. This track consists of finding alignments between the Foundational Model of Anatomy (FMA), SNOMED CT, and the National Cancer Institute Thesaurus (NCI) [7]. These ontologies are semantically rich and contain tens of thousands of classes. UMLS Metathesaurus has been selected as the basis for the track reference alignments (see [8] for details). UMLS is currently the most comprehensive effort for integrating independently-developed medical thesauri and ontologies, including FMA, SNOMED CT, and NCI. In this track we also put special attention to the number of unsatisfiabilities led by the mappings computed by a participating system.

SPIMBENCH track. The datasets in this track are produced using SPIMBENCH benchmark generator [9] with the aim to generate descriptions of the same entity where *value-based*, *structure-based* and *semantics-aware* transformations are employed on a source dataset in order to create the target dataset(s). The value-based transformations consider mainly typographical errors and different data formats, the structure-based transformations implement transformations applied on the structure of object and datatype properties and the semantics-aware transformations concern the instance level and take into account schema information. The latter are used to examine if the matching systems take into account RDFS and OWL constructs in order to discover correspondences between instances that can be found only by considering schema information.

Link discovery track. This track is composed of two tasks called: *linking* and *spatial*. The *linking* task measures how well systems can match traces that have been modified using string-based approaches along with addition and deletion of intermediate points.

The screenshot shows the HOBBIT web interface. At the top, there is a navigation bar with the HOBBIT logo, 'Upload', 'Benchmarks', 'Challenges', and 'Experiments' menus, and the user name 'Ernesto Jimenez Ruiz'. Below this is a 'Start Benchmarking' section. It contains three dropdown menus: 'Benchmark' (selected: OAEI Largebio benchmark), 'System' (selected: LogMap-HOBBIT), and 'Configuration Parameters' (selected: FMA-NCI-Small). Below the dropdowns is a blue 'Submit' button.

Fig. 2. Benchmarking LogMap with the FMA-NCI-Small *largebio* task.

The original datasets only contain coordinates, thus, we have replaced a number of those points with labels retrieved from Linked Data spatial datasets using the Google Maps¹³, Foursquare¹⁴ and Nominatim Openstreetmap¹⁵ APIs to be able apply string-based modifications implemented in LANCE [10]. This task also contains modifications of date and coordinate formats.

The *spatial* task measures how well systems can identify the DE-9IM (Dimensionally Extended nine-Intersection Model) topological relations between *LineStrings* and *Polygons* in two-dimensional spaces. The supported spatial relations are the following: *Equals*, *Disjoint*, *Touches*, *Contains/Within*, *Covers/CoveredBy*, *Intersects*, *Crosses*, *Overlaps*. The instances are represented in the Well-Known Text (WKT) format. For each relation, a different pair of source and target datasets is given to the participants.

5 Participation and proof-of-concept results

In this section we introduce the systems contributing to the OAEI 2017.5 campaign and provide an overview of how experiments are executed from the HOBBIT public instance.

5.1 HOBBIT experiments

Experiments can be executed via the HOBBIT public instance¹⁶ by following the *Benchmarks* menu. Note that, currently, only registered developers who are the owners of a system which conforms the specification (*i.e.*, API) of one or more benchmarks can run experiments. Figure 2 shows the interface to select a benchmark and evaluate a system implementing its API within the HOBBIT platform.

Every experiment is assigned a unique ID and, once they are finalized, registered users can access its results (see *Experiments* menu). In addition, one can also select several experiments for comparison purposes. For example, Figure 3 shows the results of LogMap for all six tasks of the *largebio* track.

¹³ <https://developers.google.com/maps/>

¹⁴ <https://developer.foursquare.com/>

¹⁵ <http://nominatim.openstreetmap.org/>

¹⁶ <https://master.project-hobbit.eu/>

Experiment ID	1516362047635	1516297290659	1516361094535	1516362166720	1516361695997	1516362213767
Experiment						
Benchmark	OAEI Largebio benchmark	OAEI Largebio benchmark	OAEI Largebio benchmark	OAEI Largebio benchmark	OAEI Largebio benchmark	OAEI Largebio benchmark
Challenge Task						
Error						
System	LogMap-HOBBIT	LogMap-HOBBIT	LogMap-HOBBIT	LogMap-HOBBIT	LogMap-HOBBIT	LogMap-HOBBIT
KPIs						
Fmeasure	0.91978609625668	0.83087531124305	0.79823297800825	0.72935435435435	0.78470769758271	0.70707766488720
Precision	0.94431372549019	0.85562130177514	0.94734442671529	0.83930885529157	0.95717273107486	0.86807610993657
Recall	0.89650037230081	0.80752047654504	0.68967806173249	0.64487222037836	0.66490412550842	0.59645554909936
timePerformance	15241	79131	55472	450092	204527	635372
Logs						
Benchmark Log	Download	Download	Download	Download	Download	Download
System Log	Download	Download	Download	Download	Download	Download
Parameter						
The name of the task within the Largebio track.	http://w3id.org/bench#FMA-NCI-Small	http://w3id.org/bench#FMA-NCI-Large	http://w3id.org/bench#FMA-SNOMED-Small	http://w3id.org/bench#FMA-SNOMED-Large	http://w3id.org/bench#SNOMED-NCI-Small	http://w3id.org/bench#SNOMED-NCI-Large

Fig. 3. Results of LogMap for all *largebio* tasks.

5.2 System overview

Table 2 lists the participating systems and links to available proof-of-concept results. The developers made a great effort adapting their systems to a new platform providing very useful insights about the feasibility of moving to the HOBBIT platform (see Section 6 for more details). Next, we provide a brief summary of the OAEI 2017.5 systems.

AgreementMakerLight (AML) [11, 12] is an all-purpose ontology alignment system inspired on AgreementMaker [13] and sharing its focus on flexibility and extensibility as main design paradigms. While initially primarily focused on the biomedical domain and on the use of background knowledge, its tool suite and capabilities were gradually extended to cover the full range of ontology matching tasks evaluated under the OAEI.

DisMatch [14] is an experimental ontology matching system built around the idea of leveraging the recent advancements in semantic representations of texts within the context of the ontology alignment problem. The lexical matcher uses semantic similarity calculated from distributional representations of domain-specific words. In the experiments several relatedness measures were tested, based on different text representation methods, including DomESA [15] and Word2Vec’s Skip-Gram model [16].

Kepler [17] is an ontology alignment system able to deal with normal and large scale ontologies. Kepler is also able to cope with multilingual ontologies thanks to its translator module. Kepler exploits the expressiveness of the OWL language to detect and

Table 2. Systems participating in the OAEI 2017.5. Link to results requires *guest* log in. DisMatch and KEPLER tested the platform but they did not manage to produce results.

System	New in OAEI/HOBBIT?	Implemented APIs	Link to results
AML	No/No	bench:ConferenceAPI, bench:AnatomyAPI, bench:LargebioAPI, bench:spimbenchAPI, bench:LinkingAPI	https://goo.gl/ACG3kP
DisMatch	No/Yes	bench:AnatomyAPI, bench:LargebioAPI,	-
KEPLER	No/Yes	bench:ConferenceAPI, bench:AnatomyAPI, bench:LargebioAPI,	-
LogMap	No/Yes	bench:ConferenceAPI, bench:AnatomyAPI, bench:LargebioAPI, bench:spimbenchAPI	https://goo.gl/tFDJKB
OntoIdea	No/No	bench:LinkingAPI	https://goo.gl/mUjBPK
RADON	No/No	bench:LinkingAPI	https://goo.gl/G1nUDY
SANOM	No/Yes	bench:ConferenceAPI, bench:AnatomyAPI, bench:LargebioAPI,	https://goo.gl/D8nrJk
YAM-BIO	No/Yes	bench:AnatomyAPI, bench:LargebioAPI	https://goo.gl/A496ug

compute the similarity between ontology entities through six modules: preprocessing, partitioning, translation, indexation, candidate selection and final alignment generation.

LogMap [4] relies on lexical and structural indexes to enhance scalability. It also incorporates approximate reasoning and repair techniques to minimise the number of logical errors in the aligned ontology. *LogMap* comes with two variants: *LogMap-Bio* [18], which uses BioPortal [19] as a (dynamic) provider of mediating ontologies; and *LogMapLt*, a “lightweight” variant of *LogMap* that only applies (efficient) string matching techniques.

OntoIdea [20] is an instance matching tool implementing an enhanced version of the STRIM algorithm proposed in previous work [21]. The new version of the *OntoIdea* system identifies not only the “sameAs” relationships between instances, but also the “topological” relationships (*e.g.*, contains, equals, overlaps, covers, etc.) on geo-spatial datasets. The type of relationship is driven by the information associated to the entities (*i.e.*, text or geometry).

RADON [22] is one of the systems of the LIMES framework. It addresses the efficient computation of topological relations on geo-spatial datasets, which belong to the largest sources of Linked Data. The main innovation of the approach is a novel sparse index for geo-spatial resources based on *minimum bounding boxes* (MBB). Based on this index, it is able to discard unnecessary computations for DE-9IM relations. Extensive experiments show that *RADON* scales well and outperforms the state of the art by up to 3 orders of magnitude w.r.t. to its runtime.

SANOM [23] is an ontology alignment system that uses simulated annealing as the principal technique to find correspondences between two given ontologies. The system translates the alignment problem into a state optimization and then applies the simulated annealing to find the optimal alignment of two given ontologies. The optimality of a state is obtained by a complex fitness function which utilizes different lexical and structural similarity metrics.

YAM-BIO is an instance of a generic background knowledge based ontology-matching framework [24] which is publicly available on GitHub¹⁷. *YAM-BIO* instance uses *YAM++* [25] as matcher and the two biomedical ontologies *UBERON* and *DOID* as background knowledge. In the OAEI 2017.5, *YAM-BIO* adopted a derivation with a specific algorithm that reduces the path number by avoiding to reuse the same background knowledge concept more than once, and the rule-based mapping selection strategy. *YAM-BIO* relies on the *LogMap-Repair* [26] module to eliminate the inconsistent mappings in the generated alignments.

6 Discussion and lessons learned

We collected feedback from eight platform developers pertaining to the transition from *SEALS* to *HOBBIT*. A common tenor found in most of the feedback from the systems pertained to the balance between complexity and guarantees. The *HOBBIT* platform requires (i) the systems to be implemented using the Docker stack of technologies, (ii) the implementation of a single interface to ensure a set of standardized communication processes and (iii) debugging by using log files collected by the platform across the distributed infrastructure it employs for benchmarking.

The use of Docker and associated technologies was largely regarded positively. Whereas the developers unfamiliar with Docker suggested that the supplementary effort necessary to create docker packages was considerable, most developers regarded the use of this technology as a step towards a better integration of tools and more controlled run of benchmarks. To ensure that the development with Docker can be carried efficiently, *HOBBIT* allows for single Docker files to be ran using different configurations such as to ease the deployment and use. This feature will be made more prominent to ensure that developers make more extensive use thereof.

Participating systems had to implement an API defined by the benchmark to receive the datasets which should be linked and to return the generated results. The participants found a template for this step very helpful and would like to have an even simpler template in the future to reduce the amount of adaptations. Especially when the benchmark API is adapted to support even more complex tasks like the *multifarm* task, a provided template eases the participation. The prepared template could take care of receiving the different ontologies and storing them in single files following the predefined structure of the *multifarm* task before starting the linking process.¹⁸ Such an extension would enable a backwards compatibility to older solutions which are based on the directory structure. Additionally, a clearer distinction of the necessary and optional steps when implementing the system adapter was rated as helpful.

The online instance of the *HOBBIT* platform is based on a cluster infrastructure and offers its services to a public community. Since the final evaluation is carried out

¹⁷ <https://github.com/AminaANNANE/GenericBKbasedMatcher>

¹⁸ <https://www.irit.fr/recherches/MELODI/multifarm/>

on this instance, the participants were asked to make sure that their systems can be deployed on the platform. The development process in itself can however be eased significantly by testing locally. The HOBBIT platform provides two means for doing so: a software development toolkit (SDK) and a recipe-based deployment infrastructure. The HOBBIT SDK¹⁹ allows to develop and debug a system adapter locally. To this end, the SDK simulates a platform running in a cluster and ensure rapid local development without the overheads (e.g., long waiting times) created by a shared cluster. HOBBIT also supports complete local deployments (instead of simulations) through the recipe-based framework Exoframe²⁰ for developers who would prefer not using a simulation. Therewith, it allows developers to follow a three-step process: (1) install the HOBBIT SDK or the HOBBIT platform via Exoframe, (2) develop and test your system locally, (3) upload the system image(s) and execute it (them) using the online instance.

7 Conclusions and next OAEI steps

The OAEI 2017.5 pre-campaign was instrumental to understand the strengths of HOBBIT but also the challenges into moving to a new platform. The feedback obtained from system developers has been very valuable for the next steps of the OAEI campaigns and the future development of the HOBBIT platform (e.g., support for more complex tasks, storage of computed alignment). The OAEI 2018 campaign²¹ will continue using the HOBBIT platform together with the SEALS infrastructure, with some tracks like *large-bio* providing a dual evaluation mode (i.e., both HOBBIT and SEALS). This way, system developers, organisers and HOBBIT developers will have additional time to guarantee a successful migration to the new evaluation platform. From the infrastructure point of view, the HOBBIT SDK will make the developing and debugging phase under the HOBBIT easier. In addition, we will continue offering the Maven-based framework to facilitate the submission to both HOBBIT and SEALS.

Acknowledgements. EJR was funded by the Centre for Scalable Data Access (SIR-IUS), the RCN project BigMed, and the Alan Turing project AIDA. OZ is supported by the CSF grant no. 18-23964S. DF was supported by the ELIXIR-EXCELERATE project (INFRADEV-3-2015). The work of MR is supported by grant TIN2017-86049-R (Ministerio de Economía, Industria y Competitividad, Spain). This work has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 688227 (HOBBIT).

References

1. Euzenat, J., Meilicke, C., Shvaiko, P., Stuckenschmidt, H., Trojahn dos Santos, C.: Ontology alignment evaluation initiative: six years of experience. *J. Data Semantics* **XV** (2011)
2. Euzenat, J., Shvaiko, P.: *Ontology matching*. 2nd edn. Springer-Verlag (2013)
3. Röder, M., Usbeck, R., Ngonga Ngomo, A.C.: Gerbil–benchmarking named entity recognition and linking consistently. *Semantic Web (Preprint)* (2017) 1–21
4. Jiménez-Ruiz, E., Grau, B.C., Zhou, Y., Horrocks, I.: Large-scale interactive ontology matching: Algorithms and implementation. In: *Europ. Conf. Artif. Intell. (ECAI)*. (2012)

¹⁹ <https://github.com/hobbit-project/java-sdk>

²⁰ <https://github.com/hobbit-project/exoframe-recipe-hobbit>

²¹ oei.ontologymatching.org/2018/

5. Dragisic, Z., Ivanova, V., Li, H., Lambrix, P.: Experiences from the anatomy track in the ontology alignment evaluation initiative. *Journal of biomedical semantics* **8**(1) (2017) 56
6. Zamazal, O., Svátek, V.: The Ten-Year OntoFarm and its Fertilization within the OntoSphere. *J. Web Semantics* **43** (2017) 46–53
7. Jiménez-Ruiz, E., Grau, B.C., Horrocks, I.: Is my ontology matching system similar to yours? In: 8th International Workshop on Ontology Matching. (2013) 229–230
8. Jiménez-Ruiz, E., Grau, B.C., Horrocks, I., Llavori, R.B.: Logic-based assessment of the compatibility of UMLS ontology sources. *J. Biomedical Semantics* **2**(S-1) (2011) S2
9. Saveta, T., Daskalaki, E., Flouris, G., Fundulaki, I., Herschel, M., Ngomo, A.C.N.: Pushing the limits of instance matching systems: A semantics-aware benchmark for linked data. In: WWW, Companion Volume. (2015)
10. Saveta, T., Daskalaki, E., Flouris, G., Fundulaki, I., Herschel, M., Ngomo, A.C.N.: Lance: Piercing to the heart of instance matching tools. In: Int'l Semantic Web Conference. (2015)
11. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The Agreement-MakerLight Ontology Matching System. In: OTM Conferences - ODBASE. (2013) 527–541
12. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: Results of AML in OAEI 2017. In: ISWC International Workshop on Ontology Matching (OM). (2017)
13. Cruz, I.F., Palandri Antonelli, F., Stroe, C.: AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies. *PVLDB* **2**(2) (2009) 1586–1589
14. Rybinski, M., del Mar Roldán García, M., García-Nieto, J., Montes, J.F.A.: Dismatch results for OAEI 2016. In: 11th International Workshop on Ontology Matching (OM). (2016)
15. Rybiński, M., Montes, J.F.A.: DomESA: a novel approach for extending domain-oriented lexical relatedness calculations with domain-specific semantics. *Journal of Intelligent Information Systems* **49**(3) (2017) 315–331
16. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
17. Kachroudi, M., Diallo, G., Ben Yahia, S.: OAEI 2017 results of KEPLER. In: 12th International Workshop on Ontology Matching. Volume 2032 of CEUR-WS. (2017) 138–145
18. Chen, X., Xia, W., Jiménez-Ruiz, E., Cross, V.V.: Extending an ontology alignment system with bioportal: a preliminary analysis. In: ISWC Posters & Demonstrations Track. (2014)
19. Fridman Noy, N., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.A.D., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* **37**(Web-Server-Issue) (2009)
20. Khiat, A., Benaissa, M., Belfedhal, M.A.: STRIM results for OAEI 2015 instance matching evaluation. In: 10th International Workshop on Ontology Matching. (2015) 208–215
21. Khiat, A., Mackeprang, M.: I-Match and OntoIdea results for OAEI 2017. In: 12th International Workshop on Ontology Matching. (2017) 135–137
22. Sherif, M.A., Dreßler, K., Smeros, P., Ngomo, A.N.: RADON - Rapid Discovery of Topological Relations. In: AAAI Conference on Artificial Intelligence. (2017) 175–181
23. Mohammadi, M., Atashin, A.A., Hofman, W., Tan, Y.: SANOM results for OAEI 2017. In: 12th International Workshop on Ontology Matching. (2017) 185–189
24. Annane, A., Bellahsene, Z., Azouaou, F., Jonquet, C.: Building an effective and efficient background knowledge resource to enhance ontology matching. *J. Web Semantics* (2018)
25. Ngo, D., Bellahsene, Z.: Overview of YAM++ - (not) yet another matcher for ontology alignment task. *J. Web Semantics* **41** (2016) 30–49
26. Jiménez-Ruiz, E., Meilicke, C., Grau, B.C., Horrocks, I.: Evaluating mapping repair systems with large biomedical ontologies. In: 26th Workshop on Description Logics. (2013) 246–257