ELSEVIER

Ontology mapping for semantically enabled applications

Ian Harrow[1], Rama Balakrishnan[2], Ernesto Jimenez-Ruiz[3,4], Simon Jupp[5], Jane Lomax[6], Jane Reed[7], Martin Romacker[8], Christian Senger[9], Andrea Splendiani[10], Jabe Wilson[11] and Peter Woollard[12]

[1] Ian Harrow Consulting Ltd, Whitstable, UK
[2] Genentech Inc., South San Francisco, CA, USA
[3] The Alan Turing Institute, London, UK
[4] Department of Informatics, University of Oslo, Oslo, Norway
[5] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK
[6] SciBite Ltd, Cambridge, UK
[7] Linguamatics Ltd, Cambridge, UK
[8] Roche Innovation Center, Basel, Switzerland
[9] OSTHUS GmbH, Aachen, Germany
[10] Novartis, Basel, Switzerland
[11] Elsevier RELX, London, UK
[12] GlaxoSmithKline, Stevenage, UK

In this review, we provide a summary of recent progress in ontology mapping (OM) at a crucial time when biomedical research is under a deluge of an increasing amount and variety of data. This is particularly important for realising the full potential of semantically enabled or enriched applications and for meaningful insights, such as drug discovery, using machine-learning technologies. We discuss challenges and solutions for better ontology mappings, as well as how to select ontologies before their application. In addition, we describe tools and algorithms for ontology mapping, including evaluation of tool capability and quality of mappings. Finally, we outline the requirements for an ontology mapping service (OMS) and the progress being made towards implementation of such sustainable services.

Introduction

Biomedical research is under a deluge of an increasing amount and variety of data. Diverse technologies enable more granular measurements from the laboratory bench to the clinical bedside for personalised treatments. To realise this promise, such data need to be brought together to build consistent biological knowledge bases [1]. As part of this process, different concepts, terminologies, and data models need to be reconciled. This reconciliation is supported by a variety of knowledge management resources, which cover a continuous spectrum of 'semantic expressivity' (Fig. 1).

At one extreme, we have simple lists, such as controlled vocabularies. Integration is significantly easier when different data sources use terms from a standardised list, instead of free text. Resources that have greater semantic expressivity enable more support for integration and interoperability, for instance leveraging synonyms or translating across languages. At the other extreme of the semantic spectrum, we have ontologies. These are a set of concepts in a subject area or domain that shows relations between concepts represented by properties.

Ontologies go beyond lists, thesauri, and taxonomies to provide a formal description of definitions of conceptual classes and their relations (one example being their hierarchical structure). 'Formal' means that definitions are based on a logical framework, such as the Web Ontology Language (OWL). This enables a representation of the meaning of concepts that is machine processable, ultimately allowing reasoning, generation of new knowledge, and automatic detection of

Corresponding author: Harrow, I. (ianharrowconsulting@gmail.com)

https://doi.org/10.1016/j.drudis.2019.05.020

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

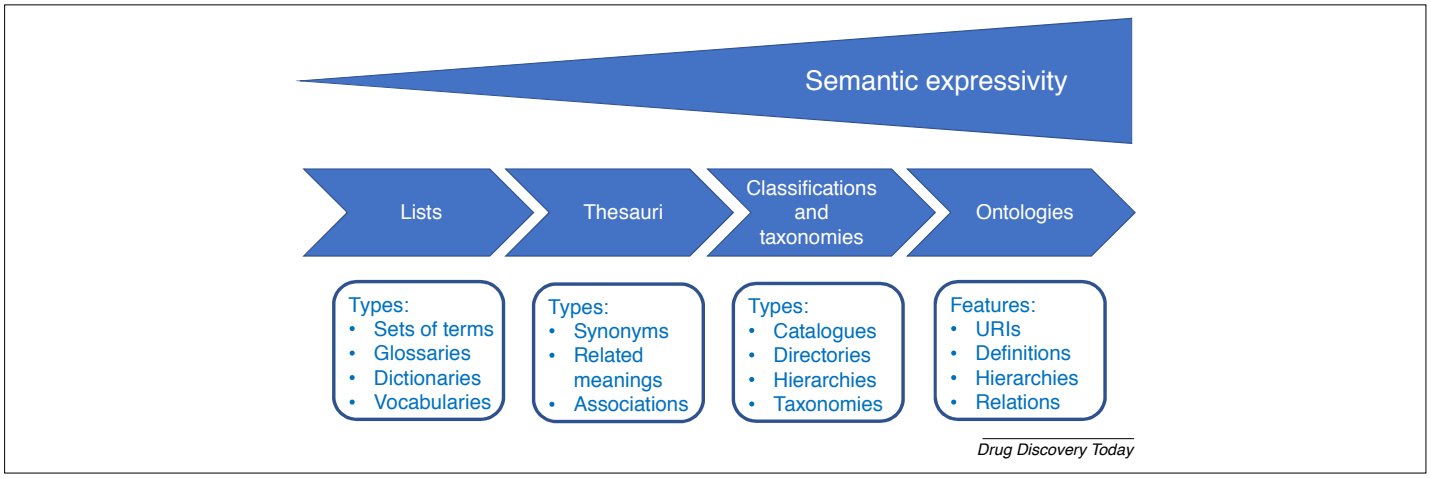

FIGURE 1

The spectrum of semantic expressivity for knowledge management resources. Abbreviation: URI, Uniform Resource Identifier.

inconsistencies in the semantic model [2]. In addition, Uniform Resource Identifiers (URI) uniquely reference each class to support machine processing and interoperability.

One of the strongest examples of a mature ontology in the biomedical sciences is Gene Ontology (GO) [3,4], which is used extensively by a multitude of applications and analytical tools [5]. Ideally, each domain in the biomedical sciences should be supported by a single reference ontology [6], an idea that was originally a strategic objective of the Open Biomedical Ontologies (OBO) consortium [6]. However, the real situation is different and finds numerous overlapping ontologies, each having their own contexts of application. This creates problems of reconciliation and even difficulties in selection of the most appropriate resource [7].

Overlap between ontologies happens for a variety of reasons. One of these is that a single reference ontology often provides insufficient coverage for a particular application, which gives rise to the development of application ontologies, such as the Experimental Factor Ontology (EFO). The EFO uses relevant parts of reference ontologies and cross references (or mappings) between them [8]. Mapping between ontologies expands the coverage across large domains, such as anatomy, disease, phenotype, and laboratory investigation. Mapping between ontologies in different domains requires the discovery of the evidence for a relationship through, for example, data or text mining [9,10].

Another reason is that many applications make use of classification systems, such as Medical Subject Headings (MeSH), Enzyme Commission (E.C.) nomenclature, Anatomical Therapeutic Chemical Classification of drugs (ATC), or Human Gene Nomenclature (HGNC), which, although powerful, were never designed as ontologies. However, it can be very useful to map between such classification systems and ontologies, which ontology-matching algorithms are able to do [11].

Although application ontologies and mapping to code lists can be built by manual curation, it is desirable to augment this process with ontology-matching algorithms [12]. These bring scalability while reducing the cost of maintenance.

Application of ontologies and their mappings

Ontology application

Controlled vocabularies have been used for many decades, especially by industry, to ensure the consistency of metadata while collecting data of an experiment or conducting analysis, often in laboratory information-management systems [13]. Seamlessly integrated into applications, controlled vocabularies and ontologies can speed up the entry of data sets and facilitate the subsequent retrieval of data through simple search interfaces. This is because experimental metadata for a biological assay can comprise many elements, including creation date, experimenter, batch/sample information (e.g., tissue or cell type, cell line, etc.), disease or normal status, and treatment (stimulant, compound, placebo, or time course) [14].

Ontologies already have an important role in annotating and organising the vast wealth of experimental, clinical, and real-world data and their day-to-day usage is well established in the scientific community. Therefore, it is not surprising that the important biomedical literature resource, PubMed, developed and applies the MeSH taxonomy for indexing and searching journal articles [15]. In pharmacovigilance, adverse events need to be reported to the US Food and Drug Administration (FDA) using the MedDRA ontology as a system to encode regulatory information [16]. Furthermore, the FDA has mandated that the Study Data Tabulation Model (SDTM), developed by the Clinical Data Interchange Standards Consortium (CDISC), must be used as the standard for the submission of study data. The controlled terminologies of SDTM are integrated with the NCI Thesaurus [17]. Real-world data provides a final example, where WHO classifications of disease, ICD-N, have been used for annotation [18]. Given that precision medicine, personalised healthcare, and translational medicine are increasingly driving modern research and development in the biopharmaceutical industry, it is vital to combine data from the vast number of public and private repositories using all these different classifications and ontologies. Therefore, ontology mapping is intrinsically tied to data integration, which is crucial for the successful discovery and development of innovative treatments of disease.

Ontologies are one of the mechanisms to encode the semantics for an area of human knowledge in a machine-readable manner [19,20]. They are vital for capturing meaningful relationships to allow users to search or browse relationships and to identify patterns from analysis [21–24]. Consider modern search engines, such as Google and Bing, which use minimal context

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

and, in the case of Wikipedia, users are presented with a disambiguation page to select the relevant results. This contrasts with a search of scientific data and literature, which requires more consistent and reliable results by harnessing controlled vocabularies, classification systems, and ontologies (Fig. 1), especially when thousands, if not millions, of results need to be processed automatically. Consider the example of bone disease, as illustrated in Fig. 2, where we can see the positions of Legg–Calve–Perthes disease and Coxa Magna in the MeSH hierarchy, without any other prior knowledge. Such hierarchical structure of a taxonomy or ontology can also help with the visualisation of data, so that a user can start with a broad class, such as bone disease, and then move on to consider more specific, yet related, diseases.

Ontologies and their mappings have a central role in open semantically enabled applications, such as Open PHACTS [25] and Open Targets [26]. Commercial examples of similar applications are Elsevier's Pathway Studio [27] and Clarivate Analytics' MetaCore/MetaBase [28]. In the case of Open Targets, this public target validation application makes fundamental use of the EFO, which has been developed and optimised to support such applications [29]. Many of these powerful applications use automated

FIGURE 2

The relational position between two bone diseases, Legg–Calve–Perthes Disease and Coxa Magna, in the Medical Subject Heading (MeSH) hierarchy.

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

text-mining technology powered by ontologies to facilitate search for subject–verb–object triplets in scientific texts.

The evidence embedded in these applications is often integrated with graphical visualisation and statistical analysis, where mapped ontologies are the key components for being able to examine the underlying biology of a hypothesis or an experiment [26,30–32]. The mapped ontologies vary by application, but typically include GO, Disease Ontology (DO), Human Phenotype Ontology (HPO), EFO, MeSH, and NCBI taxonomy. Crucially, such applications provide links to the literature from which mappings were derived, which is important to assess the confidence in such information [30,33].

Mapping between ontologies

Ontology mapping (or matching) is central to providing semantic access across aggregated data used in knowledge-based products and services consumed by life science companies, academic institutions, and universities. When bringing together ontologies and related resources (Fig. 1), we are faced with different scenarios reflecting different use cases for mappings.

As mentioned earlier, often different ontologies are used to annotate the same or similar domains, for example, HPO and Mammalian Phenotype (MP) Ontology. These ontologies have been developed independently by different communities or might be customised to meet specific user needs. In this case, ontology mapping finds equivalence (exact or synonymous matches) or relationships in the hierarchy, which can be show narrow or broad semantic similarity. Another similar example is DO, which is used widely by the research community, whereas SNOMED CT is used mostly by healthcare workers and clinicians, for example in the National Health Service, UK (https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct). Translational applications require interoperability by mapping between these two important ontologies, which has been approached successfully through lexical mappings supplemented by Unified Medical Language System (UMLS) concepts [34,35].

Another application of ontology mapping within a domain is the predictive use of phenotype annotations in different model organisms. For example, rare human gene mutations can be annotated by relating homologous mutations to phenotypes in model organisms for diagnosis of rare inherited diseases [36].

Finally, matching can also relate ontology terms between closely related domains, such as disease and phenotypes [37]. In this case, we are looking at establishing more generic relations between concepts, effectively defining a knowledge network. This scenario is a frequent task in life sciences, where ontology matching can bridge different domains and support complex research questions.

Challenges and solutions for better mappings

Generating ontology mappings can provide several challenges. Words in language can have ambiguous meanings that depend on the context. For example, the English word ‘mole’, in anatomy it is a skin feature, in chemistry it is a unit of measure, for an animal there are numerous species of talpid ‘true’ mole or a distantly related, marsupial mole or golden mole. Beyond the scientific realm, mole can be a human surname, the name of various villages, rivers, and creeks, and is also an embedded spy in an organisation, and so on. This ambiguity means that it is insufficient to simply match class names, terms, or labels for successful ontology mapping. Therefore, it is important to make use of context to resolve ambiguity, which includes background knowledge and relations among concepts [38]. Another major challenge to mapping ontologies is managing the consequences of ontology dynamics, which reflect and represent how scientific understanding evolves [1,38]. This means that any derived mappings have to be maintained, while making sure that source identifiers and labels are retained. We expand on this challenge in the OMS section of this review.

A common approach to tackling the challenge of mapping between different ontologies is to map all the terms to a single ontology or knowledge resource. Many source ontologies contain embedded cross-references that can be used as curated matches to another ontology. An Ontology of Biomedical Associations (OBAN) is an example of such an approach, which was constructed as a large-scale, generic term-association model to support construction of a target validation knowledgebase [29]. PhenomeNET is a further example, where species-specific phenotype ontologies are mapped based on the overarching, anatomy ontology, UBERON, which identifies equivalent phenotype features through anatomical concepts across different species [39,40]. Similarly, the Monarch Initiative has built a platform for mapping between phenotypes and genotypes across species, and includes the Monarch Merged Disease Ontology, called MONDO [41].

Guidance, principles, and simple rules for the selection of ontologies

When several ontologies overlap to cover a scientific domain, we are faced with the problem of how to select which ontology to use. In clinical sciences, the best practice is mature enough to be governed by authorities to meet government regulations, as described earlier, whereas, in preclinical and translational research, best practices and data standards tend to be less mature and even absent. This situation promises to improve with the MIRO guidelines for Minimum Information for Reporting of an Ontology [42]. The Pistoia Guidelines were devised as a pragmatic step to support the selection of ontologies before the application and mapping of ontologies. These guidelines are available on a public wiki of Ontologies Mapping Resources, hosted by the Pistoia Alliance (https://pistoiaalliance.atlassian.net/wiki/spaces/PUB/pages/43089928/Ontologies+Mapping+Resources). They comprise of three types of guideline: general, technical, and content in Table 1. This table shows how the Pistoia guidelines align with the principles of the Open Biological and Biomedical Ontologies (OBO) Foundry (http://www.obofoundry.org), which are under constant development and review by the OBO community. In addition, Table 1 also shows alignment to the paper entitled ‘Ten Simple Rules for Selecting a Bio-ontology’ published by Malone *et al*. [7].

The suitability of ontologies for a particular application, such as gene expression analysis or mapping between ontologies, can be reviewed using the available rules and guidelines. The National Center for Biomedical Ontology has developed a tool for this purpose, called the Ontology Recommender 2.0 [43].

‘Sometimes an Ontology is Not Needed at All’ is the tenth simple rule of Malone *et al*. [7]. This is because more light-weight knowledge management systems might be sufficient (see Fig. 1 for

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

TABLE 1

Comparison of guidelines, principles, and simple rules for the selection of ontologies

Type	Pistoia Guidelines	OBO principles	Simple rules paper [7]
Generic	License	Open	Open
	Maintenance	Maintenance	Active development
	Versioning	Versioning	Previous versions available
	Users	Plurality of users documented; commitment to collaboration	Development with the community
	Locus of authority	Locus of authority	
Technical	Format	Format	
	URIs	URIs	Persistence of classes and relationships
	Relations	Relations	
	Textual definitions	Textual definitions	Textual definitions for domain experts
	Documentation	Documentation	
	Naming conventions	Naming conventions	Textual definitions for domain experts
	Conserved URIs		Persistence of classes and relationships
Content	Content delineation	Content delineation	Specific domain
	Content coverage		Current understanding reflected
	Content quality		Textual definitions for domain experts

examples). Therefore, selection of an ontology or related resource should be driven by understanding the needs of the users.

Ontologies mapping tool evaluation

Tool requirements and capabilities

A set of minimal requirements can be used to compare the numerous academic and commercial tools designed for mapping between ontologies. These functional requirements comprise of three aspects: (i) user Interface to include visualisation of source ontologies and mapping alignment editor; (ii) framework to include workflow and ontology matching (OM) algorithm; and (iii) import ontologies or mappings and export of mappings (Fig. 3). These requirements include elements of the ontology alignment life cycle that have been described by Euzenat and Shvaiko ([44] Chapter 3).

Such functional requirements can be used to compare and evaluate the capabilities of public and commercial ontology-mapping tools. This process was undertaken in 2016, and found that one academic tool (AML [45]) and two commercial tools [Infotech Soft (http://infotechsoft.com) and Mondeca (http://en.mondeca.com)] satisfied more than 80% of the functional requirements illustrated in Fig. 3.

Evaluation of ontology matching algorithms

OM algorithms are computational tools that map between two ontologies, and have wide application beyond life sciences [44]. The Ontology Alignment Evaluation Initiative (OAEI; http://oaei.ontologymatching.org) is a mature and open annual challenge that has operated since 2004. It provides a competitive platform to showcase and evaluate the performance of latest algorithms.

It is useful to consider the different features and techniques used by OM algorithms, which can be classified as summarised in Table 2 ([44] Chapter 3). They harness lexical features (e.g., different names, synonyms, and definitions of concepts), structural, logical, or hierarchical features (e.g., the relation one concept has with other concepts within an ontology), extended information about the source ontologies (e.g., usage in annotations), and exploit background information (e.g., UMLS) [45].

OM algorithms produce a set of matches between the classes in the two ontologies being mapped. Such matches might express equivalence, binary, or multiple relations with a score of similarity. The quality of the predicted matches of the mapping results will depend on optimising the algorithm parameters, which will be specific for the ontologies being mapped.

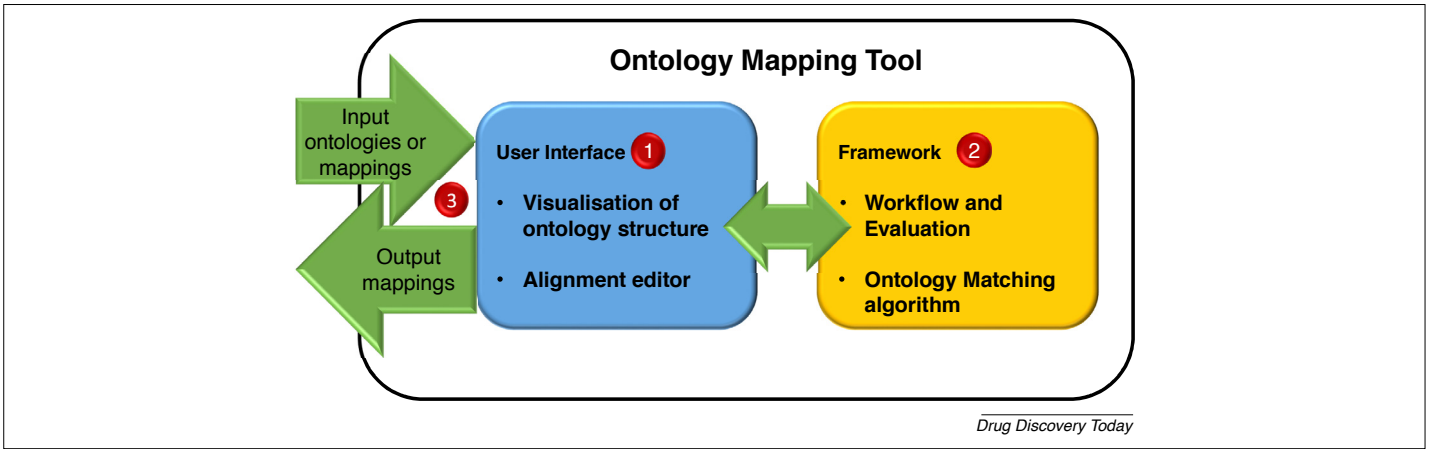

FIGURE 3

Functional requirements of an ontology mapping tool.

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

TABLE 2

Features and techniques used by OM algorithms

Level	Technical basis	Short description of technique
Element	String based	Often used to match names, identifiers, and name descriptions of ontological entities
	Language based	Considers names as words in some natural languages, such as English
	Constraint based	Deals with internal constraints applied to definitions of entities, such as types, multiplicity of attributes, and keys
	Informal resource based	Deduces relations between ontology entities based on how they relate to each other
	Formal resource based	Makes use of formal resources, such as domain-specific ontologies, upper ontologies. and linked data
Structure	Graph based	Compares source ontologies (including database schemas and taxonomies) as nodes on labelled graphs
	Taxonomy based	Hierarchical classifications consider only the specialisation relation
	Model based	Matches source ontologies based on semantic interpretation
	Instance based	Compares sets of instances of classes to decide whether they match
Knowledge	Fact or data based	Exploits facts or data stored in relevant knowledgebase or database

Numerous algorithms were tasked with matching pairs of disease and phenotype ontologies in the OAEI 2016 challenge (http://oaei.ontologymatching.org/2016). Predicted mappings were compared to a 'silver standard' from a consensus vote, given the absence of 'gold standard' mappings, in addition to limited manual evaluation. Four systems (AML [45], FCA-Map [46], LogMap(Bio) [47], and PhenoMF [40]) gave the highest performance for detection of equivalence matches, but all struggled to detect semantic similarity [37]. It is clear that a combination of automated and manual curation is required to generate high-quality mappings [48]. This is analogous to the workflow for protein annotation, where a combination of automated and manual curation is used to produce and maintain the protein knowledgebase, UniProt [49].

Toward services for ontology mappings

Service requirements

Ontologies are dynamic entities that evolve over time. Common changes include: class addition; class deprecation; combination of classes; and hierarchical relationships. Therefore, ontology mappings are not static resources and need to evolve in concert with their source ontologies; it follows that any ontology mapping needs to be provided not only as a one-off process, but also as an ongoing service [1].

Whereas the most frequently used ontologies are openly accessible, many researchers and organisations build their own ontologies, either to expand on a particular branch of a public ontology or for areas that are not well served. Therefore, there are two key use cases for an OMS: (i) mapping among public ontologies; and (ii) mapping between public and internal ontologies. The former can be achieved with a repository of mappings among popular public ontologies, which has the benefit that it can manage updates, utilise existing mappings, and generate new ones. The second use-case can be approached by providing tooling such that the user can generate bespoke mappings from their internal ontologies to public ontologies as required.

For these and other use cases, an OMS should be able to be used at all levels of an ontology, from single terms to entire branches and ontologies. This give the flexibility that researchers need in daily search and integration tasks. It is useful to contrast an OMS with an identifier mapping service, such as the BridgeDb framework, which is focused on mapping between database identifiers [50].

An OMS for mapping among public ontologies should be able to incorporate existing mapping sets, such as by utilising the cross-references between ontologies that are commonly supplied as part of the source ontology. An OMS should also harness an OM algorithm, in addition to curation, to enable mapping at scale across whole ontologies. Ideally, it should also allow the addition of user-curated content and validation of predicted mappings, assisted by 'crowd-sourcing', which has been used for ontology validation [51].

Existing standards to represent alignments should be used by an OMS ([44] Chapter 10). In addition, it should provide metadata for mappings, which include: (i) dynamics: all ontologies and any mappings between them will change over time; thus, the service needs to reflect such dynamics using both through manual curation and automation by OM algorithms. A subset of metadata should record such dynamics for interoperability and reuse; (ii) provenance. Users should have clear information on the provenance of any mapping, including ontology sources, version number, download date, and so on. Specifically, for each mapping, the service should provide annotation with suitable metadata and documentation, to enable interoperability and reuse; (iii) quality. The service should provide the quality metrics for, and within, mappings. This should include similarity scores for each match (expected to range from exact and equivalent to close similarity to broadly similar), an indication of confidence (e.g., validated or not) and global metrics, such as precision (correctness from samples) and recall (missing matches compared to standard mappings); (iv) license limitations. Some ontologies, for example SNOMED CT, have license restrictions, which might also apply to derived products, such as mappings. These restrictions should be captured as part of the OM metadata.

Implementing a prototype service

A prototype ontology mapping service has been implemented as part of the Pistoia Alliance Ontologies Mapping project (https://www.pistoiaalliance.org/projects/ontologies-mapping). The primary objective of this service is to provide mappings between ontologies, building on existing EMBL-EBI services for the life sciences [52]. In particular, the OM repository, OxO (https://www.ebi.ac.uk/spot/oxo) is being developed to store mappings (or cross-references) between terms from ontologies, vocabularies, and coding standards. OxO stores cross-references, which are curated mappings, embedded in >200 public ontologies hosted

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

by the Ontology Lookup Service (OLS). OxO makes it easier to combine mapping data sets that are labelled in different ways. Even if different standards are used to annotate data sets, they can still be made interoperable through OM. Companies can use the public-domain OMs in OxO to bridge the gap between public and private research data.

The Pistoia Alliance prototype aimed to build on OxO through development of an OM algorithm to predict mappings between public ontologies hosted by OLS. The prototype service focussed on the phenotype and disease ontology domain for ten mappings between five public ontologies, namely: HPO, DO, Orphanet Rare Disease Ontology (ORDO), MP, and MeSH. The mappings predicted by the algorithm, developed for the OMS, were compared with silver-standard mappings from consensus voting between top-performing algorithms in OAEI 2017 [37,53]. The predicted mappings from this prototype service are stored in the OxO repository, along with the curated cross-references (mappings) embedded in all the public ontologies hosted by OLS.

The OM algorithm (technical details will be disclosed in a planned technical paper), powering the OMS stored in OxO, is able to detect matches with high similarity score, where labels and synonyms are equivalent or similar between ontologies. OxO also stores the manually curated cross-references, which can be missed by the silver standards. This powerful combination of predicted mappings from an algorithm and curated mappings is an example of a solution that can deliver a scalable and sustainable mapping service.

Concluding remarks

This review shows the impressive progress made over recent years with engineering ontologies and their mappings by utilising modern tools and services. It describes how this progress enables better support for semantically aware applications. We highlight crucial challenges that must be recognised and overcome by public and private enterprise working together in sustainable ways to deliver the necessary tools and services. The important process of providing quality mappings between ontologies as a sustainable service should be supported so that it can mature as a standardised and consolidated activity.

The current flood of big data in the life sciences, especially from 'omics sources, brings massive challenges for data management. Semantic alignment and data standardisation are vital to solve if we are going to harness modern technologies, such as machine learning, for future drug discovery. These important challenges are being met by the biopharma industry through the ongoing implementation of the Findable, Accessible, Interoperable and Reusable (FAIR) guiding principles for scientific data management and governance, which make data 'findable, accessible, interoperable, and reusable' [54,55]. The interoperability principles of FAIR are supported by the effective application of ontologies and their mappings to underpin integration between many relevant sources of data [2,56].

Acknowledgements

We would like to express gratitude for funding of the Ontologies Mapping project by paying members of the Pistoia Alliance Inc. We wish to thank Thomas Liener and Helen Parkinson for helpful discussions, which have influenced this article. S.J. benefited from funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 654248 (CORBEL). E.J.R. was supported by the AIDA project, funded by the UK Government's Defence & Security Programme in support of the Alan Turing Institute, the BIGMED project (IKT 259055) and the SIRIUS Centre for Scalable Data Access (Research Council of Norway, project no.: 237889).

References

1 Rathore, A.S. *et al.* (2017) Role of knowledge management in development and lifecycle management of biopharmaceuticals. *Pharm. Res.* 34, 243–256
2 Grob, A. *et al.* (2016) Evolution of biomedical ontologies and mappings: overview of recent approaches. *Comput. Struct. Biotechnol. J.* 14, 333–340
3 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29
4 Blake, J.A. *et al.* (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056
5 The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resource. *Nucleic Acids Res.* 45, D331–D333
6 Smith, B. *et al.* (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255
7 Malone, J. *et al.* (2016) Ten simple rules for selecting a bio-ontology. *PLoS Comput. Biol.* 12, e1004743
8 Malone, J. *et al.* (2010) Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* 6, 1112–1118
9 Singhai, A. *et al.* (2016) Pressing needs of biomedical text mining in biocuration and beyond: opportunities and challenges. *Database* 2016 baw161
10 Przbyia, P. *et al.* (2016) Text mining resources for the life sciences. *Database* 2016 baw145
11 Winnenburg, R.I. and Bodenreider, O.A. (2014) A framework for assessing the consistency of drug classes across sources. *J. Biomed. Semant.* 5, 30
12 Dragisic, Z. *et al.* (2017) Experiences from the anatomy track in the ontology alignment evaluation initiative. *J. Biomed. Semant.* 8, 56
13 Harland, L. *et al.* (2011) Empowering industrial research with shared biomedical vocabularies. *Drug Discov. Today* 16, 940–947
14 Perez-Riverol, Y. *et al.* (2014) Open source libraries and frameworks for mass spectrometry-based proteomics: a developer's perspective. *Biochim. Biophys. Acta* 1844, 63–76
15 Kim, S. *et al.* (2016) Meshable: searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms. *Bioinformatics* 32, 3044–3046
16 Wang, L. *et al.* (2014) Standardizing adverse drug event reporting data. *J. Biomed. Semant.* 5, 36
17 Anzai, T. *et al.* (2015) Responses to the Standard for Exchange of Nonclinical Data (SEND) in non-US countries. *J. Toxicol. Pathol.* 28, 57–64
18 Mennini, F.S. *et al.* (2017) Economic burden of diverticular disease: an observational analysis based on real world data from an Italian region. *Dig. Liver Dis.* 49, 1003–1008
19 Whetzel, P.L. *et al.* (2013) NCBO Technology: powering semantically aware applications. *J. Biomed. Semant.* 15 (Suppl. 1), S8
20 Hoehndorf, R. *et al.* (2015) The role of ontologies in biological and biomedical research: a functional perspective. *Brief Bioinform.* 16, 1069–1080
21 Croset, S. *et al.* (2016) Flexible data integration and curation using a graph-based approach. *Bioinformatics* 32, 918–925
22 Gomez-Cabrero, D. *et al.* (2014) Data integration in the era of omics: current and future challenges. *BMC Syst. Biol.* 8 (Suppl. 2), I1
23 Lapatas, V. *et al.* (2015) Data integration in biological research: an overview. *J. Biol. Res.* 22, 9
24 Zhang, H. *et al.* (2018) An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC Med. Inform. Decis. Mak.* 18 (Suppl. 2), 4
25 Williams, A.J. *et al.* (2012) Open PHACTS: semantic interoperability for drug discovery. *Drug Discov. Today* 17, 1188–1198
26 Koscielny, G. *et al.* (2017) Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 45, D985–D994
27 Yuryev, A. *et al.* (2009) Ariadne's ChemEffect and Pathway Studio knowledge base. *Expert Opin. Drug Discov.* 4, 1307–1318

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020

28 Shmelkov, E. *et al.* (2011) Assessing quality and completeness of human transcriptional regulatory pathways on a genome-wide scale. *Biol. Direct* 6, 15
29 Sarntivijai, S. *et al.* (2016) Linking rare and common disease: mapping clinical disease–phenotypes to ontologies in therapeutic target validation. *J. Biomed. Semant.* 7, 8
30 Kafkas, Ş. *et al.* (2017) Literature evidence in open targets — a target validation platform. *J. Biomed. Semant.* 8, 20
31 Maldonado, R. *et al.* (2018) Deep learning meets biomedical ontologies: knowledge embeddings for epilepsy. *AMIA Annu. Symp. Proc.* 2017, 1233–1242
32 Arguello Casteleiro, M. *et al.* (2018) Deep learning meets ontologies: experiments to anchor the cardiovascular disease ontology in the biomedical literature. *J. Biomed. Semant.* 9, 13
33 García-Campos, M.A. *et al.* (2015) Pathway analysis: state of the art. *Front. Physiol.* 6, 383
34 Raje, S.I. and Bodenreider, O. (2017) Interoperability of disease concepts in clinical and research ontologies: contrasting coverage and structure in the disease ontology and SNOMED CT. *Stud. Health Technol. Inform.* 245, 925–992
35 Dhombres, F. and Bodenreider, O. (2016) Interoperability between phenotypes in research and healthcare terminologies — investigating partial mappings between HPO and SNOMED CT. *J. Biomed. Semant.* 7, 3
36 Petri, V. *et al.* (2014) Disease pathways at the Rat Genome Database Pathway Portal: genes in context — a network approach to understanding the molecular mechanisms of disease. *Hum. Genomics* 8, 17
37 Harrow, I. *et al.* (2017) Matching disease and phenotype ontologies in the ontology alignment evaluation initiative. *J. Biomed. Semant.* 8, 55
38 Kamder, M.R. *et al.* (2017) A systematic analysis of term reuse and term overlap across biomedical ontologies. *Semant. Web* 8, 853–871
39 Mungall, C.J. *et al.* (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* 13, R5
40 Rodríguez-García, M.Á *et al.* (2017) Integrating phenotype ontologies with PhenomeNET. *J. Biomed. Semant.* 8, 58
41 Mungall, C.J. *et al.* (2017) The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* 45, D712–D772
42 Matentzoglu, N. *et al.* (2017) MIRO: guidelines for minimum information for the reporting of an ontology. *J. Biomed. Semant.* 9, 6
43 Martínez-Romero, M. *et al.* (2017) NCBO Ontology Recommender 2.0: an enhanced approach for biomedical ontology recommendation. *J. Biomed. Semant.* 8, 21
44 Euzenat, J. and Shvaiko, P. (2013) *Ontology Matching* (2nd edn.), Springer
45 Faria, D. *et al.* (2018) Tackling the challenges of matching biomedical ontologies. *J. Biomed. Semant.* 9, 4
46 Zhao, M. *et al.* (2018) Matching biomedical ontologies based on formal concept analysis. *J. Biomed. Semant.* 9, 11
47 Jimenez-Ruiz, E. and Cuenca Grau, B. (2011) LogMap: logic-based and scalable ontology matching. *Int. Semant. Web Conf.* 2011, 273–288
48 Dragistic, Z. *et al.* (2016) User validation in ontology alignment. *Int. Semant. Web Conf.* 2016, 200–217
49 UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212
50 van Iersel, M.P. *et al.* (2010) The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinf.* 11, 5
51 Mortensen, J.M. *et al.* (2016) Is the crowd better as an assistant or a replacement in ontology engineering? An exploration through the lens of the Gene Ontology. *J. Biomed. Inf.* 60, 199–209
52 Perez-Riverol, Y. *et al.* (2017) OLS Client and OLS Dialog — Open Source Tools to Annotate Public Omics Datasets. *Proteomics* 17 (19),
53 Arichi, M. *et al.* (2017) Results of the Ontology Alignment Evaluation Initiative 2017. *CEUR Workshop Proc.* 2032, 61–113
54 Wilkinson, M.D. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018
55 Wise, J. *et al.* (2019) Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discov. Today* 24, 933–938
56 Dhombres, F. and Charlet, J. (2017) Knowledge representation and management, it's time to integrate! *Yearb. Med. Inform.* 26, 148–151

Please cite this article in press as: Harrow, I. et al. Ontology mapping for semantically enabled applications, Drug Discov Today (2019), https://doi.org/10.1016/j.drudis.2019.05.020