Semantic Web: A Review Of The Field

Pascal Hitzler
hitzler@ksu.edu
Kansas State University
Manhattan, Kansas, USA

ABSTRACT

We review two decades of Semantic Web research and applications, discuss relationships to some other disciplines, and current challenges in the field.

CCS CONCEPTS

• **Information systems → Graph-based database models**; **Information integration**; **Semantic web description languages**; **Ontologies**; • **Computing methodologies → Description logics**; **Ontology engineering**.

KEYWORDS

Semantic Web, ontology, knowledge graph, linked data

ACM Reference Format:
Pascal Hitzler. 2020. Semantic Web: A Review Of The Field. In *Proceedings of* . ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Let us begin this review by defining the subject matter. The term *Semantic Web* as used in this article is a field of research, rather than a concrete artifact – in a similar way in which, say, *Artificial Intelligence* denotes a field of research, rather than a concrete artifact. A concrete artifact, which may deserve to be called *The Semantic Web* may or may not some day come into existence, and indeed some members of the research field may argue that part of it has already been built. Sometimes the term *Semantic Web Technologies* is used to describe the set of methods and tools arising out of the field, in an attempt to avoid terminological confusion. We will come back to all this in the article in some way; however our focus will be on reviewing the research field.

This review will necessarily be rather subjective, as the field is very diverse not only in methods and goals which are being researched and applied, but also because the field is home to a large number of different but interconnected subcommunities, each of which would probably produce a rather different narrative of the history and the current state of the art of the field. I therefore do not strive to achieve the impossible task of presenting something close to a consensus – such a thing seems still elusive. However I do point out here, and sometimes within the narrative, that there are a good number of alternative perspectives.

The review is also necessarily very selective, because Semantic Web is a rich field of diverse research and applications, borrowing from many disciplines within or adjacent to computer science, and a brief review like this one cannot possibly be exhaustive or give due credit to all important individual contributions. I do hope that I have captured what many would consider key areas of the Semantic Web field. For the reader interested in obtaining a more detailed overview, I recommend perusing the major publication outlets in the field: The Semantic Web journal,[1] the Journal of Web Semantics,[2] and the proceedings of the annual International Semantic Web Conference.[3] This is by no means an exhaustive list, but I believe it to be uncontroversial that these are the most central publication venues for the field.

Now that we understand that Semantic Web is a field of research, what is it about? Answers to this question are again necessarily subjective as there is no clear consensus on this in the field.[4]

One perspective is that the field is all about the long-term goal of creating The Semantic Web (as an artifact) together with all the necessary tools and methods required for creation, maintenance, and application. In this particular narrative, The Semantic Web is usually envisioned as an enhancement of the current World Wide Web with machine-understandable information (as opposed to most of the current Web, which is mostly targeted at human consumption), together with services – intelligent agents – utilizing this information; this perspective can be traced back to the 2001 Scientific American Article [1] which arguably marks the birth of the field, and about which we will talk more below. Provision of machine-understandable information in this case is done by endowing data with expressive *metadata* for the data. In the Semantic Web, this metadata is generally in the form of *ontologies*, or at least a formal language with a logic-based semantics that admits reasoning over the meaning of the data. (Formal metadata is discussed at length below.) This, together with the understanding that intelligent agents would utilize the information, perceives the Semantic Web field as having a significant overlap with the field of Artificial Intelligence, and indeed for most of the 2000s major Artificial Intelligence conferences ran explicit "Semantic Web" tracks.

An alternative and perhaps more recent perspective on the question what the field is about, however, rests on the observation that

, ,

ACM ISBN 0001-0782/2020/04 - ArticleNumber…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

[1] http://www.semantic-web-journal.net/
[2] https://www.journals.elsevier.com/journal-of-web-semantics
[3] http://swsa.semanticweb.org/content/international-semantic-web-conference-iswc
[4] I would like to emphasize that this lack of consensus is as much a boon for the field, giving it diversity, as it is sometimes a disadvantage.

the methods and tools developed by the field have applications not tied to the World Wide Web, and which also can provide added value even without going all the way to establishing intelligent agents utilizing machine-understandable data. Indeed early industry interest in the field, which was substantial from the very outset, was aimed at applying Semantic Web Technologies to information integration and management. From this perspective, one could argue that the field is about establishing efficient (i.e., low cost) methods and tools for data sharing, discovery, integration and reuse, and the World Wide Web may or may not be a data transmission vehicle in this context. This understanding of the field moves it closer to databases, or the data management part of data science.

A much more restrictive, but perhaps practically rather astute, delineation of the field may be made by characterizing it as investigating foundations and applications of ontologies, linked data, and knowledge graphs (all discussed later), with the W3C standards[5] RDF, OWL, and SPARQL at its core; and we will explain and return to each of these points in more detail below.

Perhaps, indeed, each of the three perspectives just described has its merit, and the field exists in a confluence of these, with ontologies, linked data, knowledge graphs, being key concepts for the field, W3C standards around RDF, OWL and SPARQL constituting technical exchange formats which unify the field on a syntactic (and to a certain extent, semantic) level; the application purpose of the field is in establishing efficient methods for data sharing, discovery, integration, and reuse (whether for the Web or not); and a long-term vision that serves as a driver is the establishing of The Semantic Web as an artifact complete with intelligent agents applications, at some point in the (perhaps, distant) future.

In the rest of this article, we will lay out a timeline of the past of the field, and during this discussion we will cover a lot of ground regarding key concepts, standards, and prominent outcomes. After that, we will discuss some selected application areas, and some of the road and challenges that lie ahead.

2 A SUBJECTIVE TIMELINE

Declaring any specific point in time as the birth of a field of research is of course debateable at best. Nevertheless, the 2001 Scientific American Article [1], which was mentioned already above, is an early landmark and has provided significant visibility for the nascent field. And yes it was around the early 2000s when the field was in a very substantial initial upswing in terms of community size, academic productivity, and initial industry interest.

But of course there were earlier efforts. The DARPA Agent Markup Language (DAML) Program[6] ran from 2000 to 2006 with the declared goal to develop a Semantic Web language and corresponding tools. The European Union funded On-To-Knowledge project[7], running from 2000 to 2002, gave rise to the OIL language which was later merged with DAML, eventually giving rise to the Web Ontology Language (OWL) W3C standard which we will discuss in more detail below. And the more general idea of endowing data on the web with machine-readable or "-understandable" metadata can be traced back to the beginnings of the World Wide Web itself. For example, a first draft of the Resource Description Framework (RDF) was published as early as 1997.[8]

Our story of the field will commence from the early 2000s, and we will group the narrative into three overlapping phases, each driven by a key concept, i.e., under this reconstruction the field has shifted its main focus at least twice. From this perspective, the first phase was driven by *ontologies* and it spans the early to mid 2000s; the second phase was driven by *linked data* and stretches into the early 2010s. The third phase was and is still driven by *knowledge graphs*. We discuss each of these in the following.

2.1 Ontologies

For most of the 2000s, work in the field had the notion of *ontology* at its center, which of course has much older roots. According to a many-cited source from 1993 [5], an ontology is a formal, explicit specification of a shared conceptualization – though one may argue that this definition still needs interpretation, and is rather generic. In a more precise sense (and perhaps a bit post-hoc), an ontology is really a knowledge base (in the sense of symbolic artificial intelligence) of concepts (i.e., types or classes, such as "mammal" and "live birth") and their relationships (such as, "mammals give live birth"), specified in a knowledge representation language based on a formal logic. In a Semantic Web context, ontologies are a main vehicle for data integration, sharing, and discovery, and a driving idea is that ontologies themselves should be reuseable by others.

In 2004, the Web Ontology Language OWL became a W3C standard (the revision OWL 2 [11] was established in 2012), providing further fuel for the field. OWL in its core is based on a *description logic*, i.e., on a sublanguage of first-order predicate logic[9] using only unary and binary predicates and a restricted use of quantifiers, designed in such a way that logical deductive reasoning over the language is decideable [12]. Even after the standard was established, the community continued to have discussions whether description logics were the best paradigm choice, with rule-based languages being a major contender [28]. The discussion eventually settled, but the Rule Interchange Format RIF [25] which was later established as a rule-based W3C standard gained relatively little traction.[10]

Also in 2004, the Ressource Description Framework (RDF) became a W3C standard (the revision RDF 1.1 [32] was completed in 2014). In essence, RDF is a syntax for expressing directed, labelled and typed graphs.[11] RDF is more or less[12] compatible with OWL, by using OWL to specify an ontology of types and their relationships, and by then using these types as types in the RDF graph, and the

[5] The World Wide Web Consortium (W3C) calls its standards "Recommendations."
[6] http://www.daml.org/
[7] https://cordis.europa.eu/project/id/IST-1999-10132
[8] https://www.w3.org/TR/WD-rdf-syntax-971002/
[9] With some mild extensions not found in standard first-order predicate logic, such as counting quantifiers.
[10] Evidence for this is, e.g., given by comparing Google Scholar citation counts for the standards documents, which are two orders of magnitude lower for RIF.
[11] The full standard is more complicated, e.g., it allows things like using edge labels, or node types, also as nodes from which other edges originate, which would be in violation of what is usually considered a graph. Excessive use of such departures from standard graph structures are usually used sparingly, as the results are often hard to interpret.
[12] Syntactically, they are fully compatible, as RDF is a syntactic serialization format for OWL. However, RDF and OWL each carry a (more precisely, several) formal semantics that are not fully compatible between the languages. To the best of my knowledge, there is no single reference which discusses the exact relationship in detail, but [12] gives some indications.

relationships as edges. From this perspective, an OWL ontology can serve as a *schema* (or a logic of types) for the RDF (typed) graph.[13]

A W3C standard for an RDF query language, called SPARQL, followed in 2008 (with an update in 2013 [36] which then also became more fully compatible with OWL). Additional standards in the vicinity of RDF, OWL and SPARQL have been, or are being, developed, some of which have gained significant traction, e.g. ontologies such as the Semantic Sensor Networks ontology [7] or the Provenance ontology [20], or the SKOS Simple Knowledge Organization System [24].

With all these key standards developed under the W3C, basic compatibility between them and other key W3C standards has been maintained. For example, XML serves as a syntactic serialization and interchange format for RDF and OWL. All W3C Semantic Web standards also use IRIs as identifiers for labels in an RDF graph, for OWL class names, for datatype identifiers, etc.

The DARPA DAML program ended in 2006, and subsequently there were few if any large-scale funding lines for fundamental Semantic Web research in the U.S. As a consequence, much of the corresponding research in the U.S. moved either to application areas such as data management in healthcare of defense, or into adjacent fields altogether. In contrast, the European Union Framework Programmes, in particular FP 6 (2002-2006) and FP 7 (2007-2013) provided significant funding for both foundational and application-oriented Semantic Web research. One of the results of this divergence in funding priorities is still mirrored in the composition of the Semantic Web research community, which is dominantly European. The size of the community is difficult to assess, but since the mid-2000s, the field's key conference, the International Semantic Web Conference, has drawn on average over 600 participants each year.[14] Given the interdisciplinary nature and diverse applications of the field, however, it is to be noted that much Semantic Web research or applications are published in venues for adjacent research or application fields.

Industry interest has been significant from the outset, but it is next to impossible to reconstruct reliable data on the precise level of related industry activity. University spin-offs applied state of the art research from the outset, and graduating PhD students – in particular the significant number produced in Europe – were finding corresponding industry jobs. Major and smaller companies have been involved in large-scale foundational or applied research projects, in particular under EU FP 6 and 7. Industry interest has changed focus with the research community, and we will come back to this throughout the narrative.

Some large-scale ontologies, often with roots pre-dating the Semantic Web community, matured during this time. For example, the Gene Ontology [35] had its beginnings in 1998 and is now a very prominent resource. Another example is SNOMED CT[15] which can be traced back to the 1960s but is now fully formalized in OWL and widely used for electronic health records [33].

[13]RDF Schema [32], which is part of the RDF standard, can serve this purpose as well but is much less expressive than OWL, and in terms of semantics not fully compatible with it – see the previous footnote.

[14]The much newer annual China Conference on Knowledge Graph and Semantic Computing, established in 2013, with primarily national focus, has by now grown to almost 1,500 participants.

[15]https://www.snomed.org/

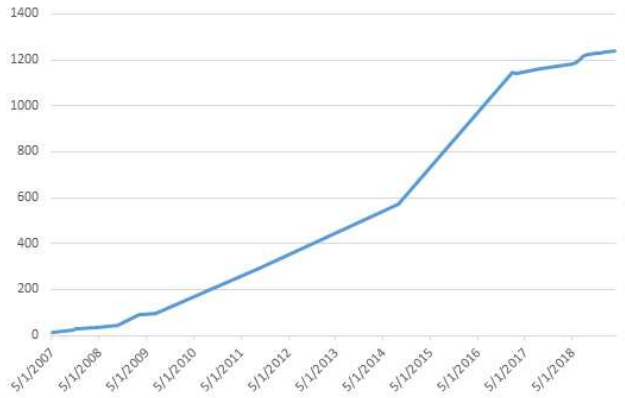

Figure 1: Number of RDF graphs in the Linked Open Data Cloud over time

As is so often the case in Computer Science research, though, initial over-hyped expectations on short-term massive breakthrough results gave way, around the mid-2000s, to a more sober perspective. Ontologies in the form in which they were mostly developed during this time – meaning often based on ad-hoc modeling as methodologies for their development were researched on but had not yet led to tangible results – turned out to be often hard to maintain and re-use. This, combined with the considerable up-front cost at that time to develop good ontologies[16] paved the way for a shift in attention by the research community, which can be understood as perhaps antithetical to the strongly ontology-based approach of the early 2000s.

2.2 Linked Data

The year 2006 saw the birth of "Linked Data" (or "Linked Open Data" if the emphasis is on open, public, availability under free licenses). Linked Data [3] would soon become a major driver for Semantic Web research and applications and persist as such until the early 2010s.

What is usually associated with the term "Linked Data" is that Linked Data consists of a (by now rather large) set of RDF graphs which are "linked" in the sense that many IRI identifiers in the graphs appear also in other, in fact sometimes in multiple, graphs. In a sense, the collection of all these linked RDF graphs can be understood as one very big RDF graph.

The number of publicly available linked RDF graphs has been showing significant growth in particular during the first decade as shown in Figure 1; the data is from the Linked Open Data Cloud website[17] which does not account for all RDF datasets on the Web. A 2015 paper [29] reports on "more than 37 billion triples[18] from over 650,000 data documents" which is also only a selection of all RDF graph triples that can be freely accessed on the World Wide Web. Large data providers, for example, often provide only a query interface based on SPARQL (a "SPARQL endpoint"), or use RDF for internal data organization but provide it to the outside only

[16]With it being rather unclear what "good" would mean.

[17]https://lod-cloud.net/

[18]In RDF terminology, a *triple* consists of a node-edge-node piece of an RDF graph.

via human-readable web pages. Datasets in the Linked Open Data Cloud cover a wide variety of topics, including Geography, Government, Life Sciences, Linguistics, Media, Scientific Publications, Social Networking.

One of the most well-known and used linked datasets is DBpedia [22]. DBpedia is a linked dataset extracted from Wikipedia (and, more recently, also Wikidata which we discuss further below). The April 2016 release[19] covers about 6 millon entities and about 9.5 billion RDF triples. Due to its extensive topic coverage (essentially, everything in Wikipedia) and the fact that it was one of the very first linked datasets to be made available, DBpedia plays a central role in the Linked Open Data Cloud of interlinked datasets: Many other datasets link to it so that it has become a kind of hub for Linked Data.

There was quite some industry interest in Linked Data from the outset. For example, BBC[20] was one of the first significant industry contributors to the Linked Data Cloud and the New York Times Company [31] and Facebook [40] were early adopters. However industry interest seemed mostly be about utilizing Linked Data *technology*, e.g. for data integration and management, often without it being visible on the open World Wide Web.

During the Linked Data era, ontologies played a much less prominent role. They still often were used as schemas in that they informed the internal structure of RDF datasets, however compared to the overpromises and depth of research from the Ontologies era, the information in RDF graphs in the Linked Data Cloud was shallow and relatively simplistic. The credo sometimes voiced during this time was that ontologies cannot be reused, and that a much simpler approach based mainly on utilizing RDF and links between datasets held much more realistic promises for data integration, management, and applications on and off the Web. It was also during this time, that RDF-based data organization vocabularies with little relation to ontologies, such as SKOS [24], were developed.

It was also during this time, in 2011, when Schema.org appeared on the scene [6]. Initially driven by Bing, Google and Yahoo! – and slightly later joined by Yandex – Schema.org made public a relatively simple ontology[21] and suggested to website providers that they annotate (i.e. link) entities on their websites with the Schema.org vocabulary. In return, the Web search engine providers behind Schema.org promised to improve search results by utilizing the annotations as metadata. Schema.org saw considerable initial uptake: [6], from 2015, reports that over 30% of pages have Schema.org annotations.

Another prominent effort from this time period – launched in 2012 – is Wikidata [39], which started as a project at Wikimedia Deutschland funded among others by Google, Yandex and the Allen Institute for AI. Wikidata is based on a similar idea as Wikipedia, namely to crowdsource information; but while Wikipedia is doing this for encyclopedia-style texts (with human readers as the main consumers), Wikidata is about creating structured data that can be used by programs or in other projects. For example, many other Wikimedia efforts, including Wikipedia, use Wikidata to provide some of the information that they present to human readers. As of the time of this writing, Wikidata has over 66 million data items, has had over one billion edits since project launch, and has over 20,000 active users.[22] Database downloads are available in several W3C standards, including RDF.

During the early 2010s, the initial hype about Linked Data began to give way to a more sober perspective. While there were indeed some prominent uses and applications of Linked Data, it still turned out that integrating and utilizing Linked Data took more effort than some initially expected. Arguably [16], shallow non-expressive schemas as often used for Linked Data appeared to be a major obstacle to reuseability, and initial hopes that interlinks between datasets would somehow account for this lack did not really seem to materialize. This observation should not be understood as demeaning the significant advances Linked Data has brought to the field and its applications: Just having data available in some structured format which follows a prominent standard, means that it can be accessed, integrated, and curated with available tools, and then made use of – and this is much easier than if data is provided in syntactically and conceptually much more heterogeneous form. But the quest for more efficient approaches to data sharing, discovery, integration, and reuse was of course as important as ever, and commencing.

2.3 Knowledge Graphs

In 2012, a new term appeared on the scene when Google launched its *Knowledge Graph*. Pieces of the Google Knowledge Graph can be seen, e.g., by searching for prominent entities on google.com: next to the search results linking to Web pages a so-called *infobox* is displayed which shows information from the Google Knowledge Graph. An example for such an infobox is given in Figure 2 – this was retrieved by searching for the term *Kofi Annan*. One can navigate from this node to other nodes in the graph by following one of the active hyperlinks, e.g. to *Nane Maria Annan* who is listed with a spouse relationship to the *Kofi Annan* node. After following this link, a new infobox for *Nane Maria Annan* is displayed next to the usual search results for the same term.

While Google does not provide the Knowledge Graph for download, it does provide an API to access content[23] – the API uses standard schema.org types and is compliant with JSON-LD [34], which is essentially an alternative syntax for RDF standardized by the W3C.

Knowledge graph technology has in the meantime found a prominent place in industry, including leading information technology companies other than Google, such as Microsoft, IBM, Facebook, eBay [27]. However, given the history of Semantic Web technologies, and in particular of linked data and ontologies as discussed above, it seems to be apparent that *knowledge graph* is mostly a new framing of ideas coming rather directly out of the Semantic Web field.[24] With, of course, some notable shifts in emphasis.

One of the differences is about *openness*: As the term Linked *Open* Data has suggested from the very beginning, the Linked Data efforts by the Semantic Web community mostly had open sharing

[19] https://blog.dbpedia.org/2016/10/19/yeah-we-did-it-again-new-2016-04-dbpedia-release/

[20] https://www.bbc.co.uk/academy/en/articles/art20130724121658626

[21] As of the writing of this manuscript, it has 614 classes and 902 relations, and consists primarily of a type hierarchy.

[22] https://www.wikidata.org/wiki/Wikidata:Statistics

[23] https://developers.google.com/knowledge-graph

[24] The term *knowledge graph* is of course also not new as such, it was alrady used, e.g., in the 1980s with a similar general meaning.

Figure 2: Google Knowledge Graph node – as shown after searching on google.com for the term *Kofi Annan*.

of data for reuse as one its goals, which means that linked data is mostly made freely available for download or by SPARQL endpoint, and the use of non-restricting licenses is considered of importance in the community. Wikidata as a knowledge graph is also unowned, and open. In contrast, the more recent activities around knowledge graphs are often industry-led, and the prime showcases are not really open in this sense [27].

Another difference is one of central control versus bottom-up community contributions: The Linked Data Cloud is in a sense the currently largest existing knowledge graph known, but it is hardly a concise entity. Rather, it consists of loosely inter-linked individual subgraphs, each of which is governed by its very own structure, representation schema, etc. Knowledge graphs, in contrast, are usually understood to be much more internally consistent, and more tightly controlled, artifacts. As a consequence, the value of "external links" – i.e., to external graphs without tight quality control – is put into doubt,[25] while quality of content and/or the undertlying schema comes more into focus.

However the biggest difference is probably the transition from academic research (which mostly drove the Linked Data effort) to use in industry. As such, recent activites around knowledge graphs are fueled by the strong industrial use cases and their demonstrated or perceived added value, even though there is, to the best of my knowledge, no published formal evaluation of their benefits.

Yet, many of the challenges and issues concerning knowledge graphs remain the same as they were for Linked Data, e.g., all items on the list of current challenges listed in [27] are very well-known in the Semantic Web field, many with substantial bodies of research having been undertaken.

3 SELECTED RELATIONSHIPS TO OTHER FIELDS AND DISCIPLINES

As we have discussed in the introduction, the Semantic Web field is not primarily driven by certain methods inherent to the field, which distinguishes it from some other areas such as machine learning. Rather, it is driven by a shared vision,[26] and as such it borrows from other disciplines as needed.[27]

For example, the Semantic Web field has strong relations to Knowledge Representation and Reasoning as a sub-discipline of Artificial Intelligence, as knowledge graph and ontology representation languages can be understood – and are closely related to – knowledge representation languages, with Description Logics, as the logics underpinning the Web Ontology Language OWL, playing a central role. Semantic Web application needs have also driven or inspired Description Logic research, as well as investigations into bridging between different knowledge representation approaches such as rules and description logics [19].

The field of databases is obviously closely related, where topics such as (meta)data management and graph-structured data have a natural home but are also of importance for the Semantic Web field. However, in Semantic Web research emphasis is much more strongly on conceptual integration of heterogeneous sources, e.g., how to overcome different ways to organize data; in Big Data terminology, Semantic Web emphasis is primarily on the *variety* aspect of data [17].

Natural Language Processing as an application tool plays an important role, e.g., for knowledge graph and ontology integration, for natural language query answering, as well as for automated knowledge graph or ontology construction from texts.

Machine Learning, and in particular deep learning, are being investigated as to their capability to improve hard tasks arriving in a Semantic Web context, such as knowledge graph completion (in the sense of adding missing relations), dealing with noisy data, and so on [4, 10]. At the same time, Semantic Web technologies are being investigated as to their potential to advance explainable AI [10, 21].

Some aspects of Cyber-Physical Systems and the Internet of Things are being researched on using Semantic Web technologies, e.g., in the context of smart manufacturing (Industry 4.0), smart energy grids, and building management. [30]

Some areas in the life sciences have already a considerable history of benefiting from Semantic Web technologies. We already mentioned SNOMED-CT and the Gene Ontology. Generally speaking, biomedical fields were early adopters of Semantic Web concept. Another prominent example would be the development of the ICD-11, which was driven by Semantic Web technologies [38].

Other current or potential application areas for Semantic Web technologies can be found wherever there is a need for data sharing,

[25] Early indicators of this have e.g. shown that many of the same-as links contained in the Linked Data Cloud link entities which should not as such be considered exactly the same [8].

[26] Another discipline not primarily driven by methods, but rather by shared vision or goals is, e.g., Cybersecurity.

[27] See e.g. the ISWC 2006 keynote by Rudi Studer on *Semantic Web: Customers and Suppliers*, see http://videolectures.net/iswc06_studer_sc/.

discovery, integration and reuse, e.g., in geosciences or in digital humanities [15].

4 SOME OF THE ROAD AHEAD

Undoudbtedly, the grand goal of the Semantic Web field – be it the creation of The Semantic Web as an artifact, or providing solutions for data sharing, discovery, integration and reuse which make it completely easy and painless – has not been achieved yet. This of course does not mean that intermediate results are not of practical use or even industrial value, as our discussions above around knowledge graphs, schema.org, and the life science ontologies demonstrate.

Yet, to advance towards the larger goals, further advances are required in virtually every subfield of the Semantic Web. For many of these, discussions of some of the most pressing challenges can be found, e.g., in [2], in the contributions to the January 2020 special issue of the Semantic Web journal,[28] or in the above referenced [27] for industrial knowledge graphs, in [37] for ontology alignment, in [23] for information extraction, in [13] for question answering, or in [9] for ontology design patterns, etc. Rather than to repeat or recompile these lists, let us focus on the challenge which I personally consider to be the current, short-term, most major roadblock for the field at large.

There is a wealth of knowledge – hard and soft – in the Semantic Web community and its application communities, about how to approach issues around efficient data management. Yet, new adopters often find themselves confronted with a cacophony of voices pitching different approaches, little guidance as to the pros and cons of these different approaches, and a bag of tools which range from crude unfit-for-practice research prototypes to well-designed software for particular subproblems, but again with little guidance which tools, and which approaches, will help them best in achieving their particular goals.

Thus, what I see that the Semantic Web field most needs, at this stage, is consolidation. And as an inherently application-driven field, this consolidation will have to happen across its subfields, resulting in application-oriented processes which are well-documented as to their goals and pros and cons, and which are accompanied by easy-to-use and well-integrated tools supporting the whole process. For example, some of the prominent and popular software available, such as the Protégé ontology editor [26], the OWL API [14], Wikibase which is the engine underlying Wikidata,[29] or the ELK reasoner [18], are powerful and extremely helpful, but fall far short from working easily with each other in some cases, even though they all use RDF and OWL for serializations.

Who could be the drivers of such consolidation? For academics, there is often limited incentive to develop and maintain stable, easy-to-use software, as academic credit – mostly measured in publications and in the sum of acquired external funding – does often not align well with these activities. Likewise, complex processes are inherently difficult to evaluate, which means that top-tier publication options for such kinds of work are limited. Writing high-quality introductory textbooks as a means to consolidate a field is very time-consuming and returns very little academic credit. Yet, the academic community does provide a basis for consolidation, by developing solutions that bridge between paradigms, and by partnering with application areas to develop and materialize use-cases.

Consolidation of sorts is also of course already happening in industry, as witnessed by the adoption of Semantic Web technologies in start-ups and multinationals. Technical details – not even to speak of in-house software – underlying this adoption, as e.g. in the case of the industrial knowledge graphs discussed in [27], are however usually not shared, presumably to protect the own competitive edge. If this is indeed the case, then it may only be a matter of time before corresponding software solutions become more widely available.

5 CONCLUSIONS

Within its first roughly 20 years of existence, the Semantic Web field has produced a wealth of knowledge regarding efficient data management for data sharing, discovery, integration and reuse. The contributions of the field are best understood by means of the applications they have given rise to, including Schema.org, industrial knowledge graphs, Wikidata, ontology modeling applications, etc., as discussed throughout this paper.

It is natural to also ask about the key scientific discoveries which have provided the foundations for these applications; however this question is much more difficult to answer. As I hope has become clear from the narrative, advances in the pursuit of the Semantic Web theme require contributions from many computer science subfields, and one of the key quests is about finding out how to piece together contributions, or modifications thereof, in order to provide applicable solutions. In this sense, the applications (including those mentioned herein) showcase the major scientific progress of the field as a whole.

Of course many of the contributing fields have also individually made major advances in the past 20 years, and sometimes central individual publications have decisively shaped the narrative of a subfield. Reporting in more detail on such advances would be a worth while endeavor but will constitute a separate piece in its own right. The interested reader is encouraged to follow up on the references given, which in turn will point to the key individual technological contributions which lead to the existing widely used standards, the landmark applications reported herein, and the current discussion on open technical issues in the field to which references have been included.

The field is seeing mainstream industrial adoption, as laid out in the narrative. However the quest for more efficient data management solutions is far from over and continues to be a driver for the field.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant 2033521 *A1: KnowWhereGraph: Enriching and Linking Cross-Domain Knowledge Graphs using Spatially-Explicit AI Technologies.* Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

[28] http://www.semantic-web-journal.net/issues
[29] https://wikiba.se/

REFERENCES

[1] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American* 284, 5 (May 2001), 34–43.

[2] Abraham Bernstein, James A. Hendler, and Natalya Fridman Noy. 2016. A new look at the semantic web. *Commun. ACM* 59, 9 (2016), 35–37.

[3] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2009. Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.* 5, 3 (2009), 1–22.

[4] Claudia d'Amato. 2020. Machine Learning for the Semantic Web: Lessons Learnt and Next Research Directions. *Semantic Web* 11, 1 (2020), 195–203.

[5] Thomas R Gruber. 1993. A translation approach to portable ontology specifications. *Knowledge acquisition* 5, 2 (1993), 199–220.

[6] Ramanathan V. Guha, Dan Brickley, and Steve Macbeth. 2016. Schema.org: evolution of structured data on the web. *Commun. ACM* 59, 2 (2016), 44–51. https://doi.org/10.1145/2844544

[7] Armin Haller, Krzysztof Janowicz, Simon Cox, Danh Le Phuoc, Kerry Taylor, and Maxime Lefrancois (Eds.). 2017. *Semantic Sensor Network Ontology.* W3C Recommendation 19 October 2017. Available from http://www.w3.org/TR/vocab-ssn/.

[8] Harry Halpin, Patrick J. Hayes, and Henry S. Thompson. 2011. When owl: sameAs isn't the Same Redux: A preliminary theory of identity and inference on the Semantic Web. In *Workshop on Discovering Meaning On the Go in Large Heterogeneous Data 2011 (LHD-11), Barcelona, Spain, July 16, 2011.* 25–30.

[9] Karl Hammar, Eva Blomqvist, David Carral, Marieke van Erp, Antske Fokkens, Aldo Gangemi, Willem Robert van Hage, Pascal Hitzler, Krzysztof Janowicz, Nazifa Karima, Adila Krisnadhi, Tom Narock, Roxane Segers, Monika Solanki, and Vojtech Svátek. 2016. Collected Research Questions Concerning Ontology Design Patterns. In *Ontology Engineering with Ontology Design Patterns – Foundations and Applications*, Pascal Hitzler, Aldo Gangemi, Krzysztof Janowicz, Adila Krisnadhi, and Valentina Presutti (Eds.). Studies on the Semantic Web, Vol. 25. IOS Press, 189–198.

[10] Pascal Hitzler, Federico Bianchi, Monireh Ebrahimi, and Md Kamruzzaman Sarker. 2020. Neural-Symbolic Integration and the Semantic Web. *Semantic Web* 11, 1 (2020), 3–11.

[11] Pascal Hitzler, Markus Krötzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph (Eds.). 2012. *OWL 2 Web Ontology Language: Primer (Second Edition).* W3C Recommendation 11 December 2012. Available from http://www.w3.org/TR/owl2-primer/.

[12] Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. 2010. *Foundations of Semantic Web Technologies.* Chapman & Hall/CRC.

[13] Konrad Höffner, Sebastian Walter, Edgard Marx, Ricardo Usbeck, Jens Lehmann, and Axel-Cyrille Ngonga Ngomo. 2017. Survey on challenges of Question Answering in the Semantic Web. *Semantic Web* 8, 6 (2017), 895–920.

[14] Matthew Horridge and Sean Bechhofer. 2011. The OWL API: A Java API for OWL ontologies. *Semantic Web* 2, 1 (2011), 11–21.

[15] Eero Hyvönen. 2020. Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipidous Knowledge Discovery. *Semantic Web* 11, 1 (2020), 187–193.

[16] Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, and Amit P. Sheth. 2010. Linked Data Is Merely More Data. In *Linked Data Meets Artificial Intelligence, Papers from the 2010 AAAI Spring Symposium, Technical Report SS-10-07, Stanford, California, USA, March 22-24, 2010.* AAAI.

[17] Krzysztof Janowicz, Frank van Harmelen, James A. Hendler, and Pascal Hitzler. 2015. Why the Data Train Needs Semantic Rails. *AI Magazine* 36, 1 (2015), 5–14.

[18] Yevgeny Kazakov, Markus Krötzsch, and Frantisek Simancik. 2014. The Incredible ELK – From Polynomial Procedures to Efficient Reasoning with EL Ontologies. *J. Autom. Reasoning* 53, 1 (2014), 1–61.

[19] Adila Krisnadhi, Frederick Maier, and Pascal Hitzler. 2011. OWL and Rules. In *Reasoning Web. Semantic Technologies for the Web of Data – 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures (Lecture Notes in Computer Science)*, Axel Polleres, Claudia d'Amato, Marcelo Arenas, Siegfried Handschuh, Paula Kroner, Sascha Ossowski, and Peter F. Patel-Schneider (Eds.), Vol. 6848. Springer, 382–415.

[20] Timothy Lebo, Satya Sahoo, and Deborah McGuinness (Eds.). 2013. *PROV-O: The PROV Ontology.* W3C Recommendation 30 April 2013. Available from http://www.w3.org/TR/prov-o/.

[21] Freddy Lecue. 2020. On the Role of Knowlege Graphs in Explainable AI. *Semantic Web* 11, 1 (2020), 41–51.

[22] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.

[23] Jose L. Martinez-Rodgriguez, Aidan Hogan, and Ivan Lopez-Arevalo. 2020. Information Extraction meets the Semantic Web: A Survey. *Semantic Web* 11, 2 (2020), 255–335.

[24] Alistair Miles and Sean Bechhofer (Eds.). 2009. *SKOS Simple Knowledge Organization System.* W3C Recommendation 18 August 2009. Available from http://www.w3.org/TR/skos-reference/.

[25] Leora Morgenstern, Chris Welty, Harold Boley, and Gary Hallmark (Eds.). 2013. *RIF Primer (Second Edition).* W3C Working Group Note 5 February 2013. Available from http://www.w3.org/TR/rif-primer/.

[26] Mark A. Musen. 2015. The Protégé project: a look back and a look forward. *AI Matters* 1, 4 (2015), 4–12.

[27] Natalya Fridman Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. 2019. Industry-scale knowledge graphs: lessons and challenges. *Commun. ACM* 62, 8 (2019), 36–43.

[28] Peter F. Patel-Schneider and Ian Horrocks. 2006. Position paper: a comparison of two modelling paradigms in the Semantic Web. In *Proceedings of the 15th international conference on World Wide Web, WWW 2006, Edinburgh, Scotland, UK, May 23-26, 2006*, Les Carr, David De Roure, Arun Iyengar, Carole A. Goble, and Michael Dahlin (Eds.). ACM, 3–12.

[29] Laurens Rietveld, Wouter Beek, and Stefan Schlobach. 2015. LOD Lab: Experiments at LOD Scale. In *The Semantic Web – ISWC 2015 – 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II (Lecture Notes in Computer Science)*, Marcelo Arenas, Óscar Corcho, Elena Simperl, Markus Strohmaier, Mathieu d'Aquin, Kavitha Srinivas, Paul T. Groth, Michel Dumontier, Jeff Heflin, Krishnaprasad Thirunarayan, and Steffen Staab (Eds.), Vol. 9367. Springer, 339–355.

[30] Marta Sabou, Stefan Biffl, Alfred Einfalt, Lukas Krammer, Wolfgang Kastner, and Fajar J. Ekaputra. 2020. Semantics for Cyber-Physical Systems: A Cross-Domain Perspective. *Semantic Web* 11, 1 (2020), 115–124.

[31] Evan Sandhaus. 2010. Abstract: Semantic Technology at The New York Times: Lessons Learned and Future Directions. In *The Semantic Web - ISWC 2010 – 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010, Revised Selected Papers, Part II (Lecture Notes in Computer Science)*, Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, Jeff Z. Pan, Ian Horrocks, and Birte Glimm (Eds.), Vol. 6497. Springer, 355.

[32] Guus Schreiber and Yves Raimond (Eds.). 2014. *RDF 1.1 Primer.* W3C Working Group Note 24 June 2014. Available from http://www.w3.org/TR/rdf11-primer/.

[33] Stefan Schulz, Boontawee Suntisrivaraporn, Franz Baader, and Martin Boeker. 2009. SNOMED reaching its adolescence: Ontologists' and logicians' health check. *I. J. Medical Informatics* 78, Supplement-1 (2009), S86–S94.

[34] Manu Sporny, Dave Longley, Gregg Kellogg, Markus Lanthaler, and Niklas Lindström. 2014. *JSON-LD 1.0. A JSON-based Serialization for Linked Data.* W3C Recommendation 16 January 2014. Available from http://www.w3.org/TR/json-ld/.

[35] The Gene Ontology Consortium. 2008. The Gene Ontology Project in 2008. *Nucleic Acids Research* 36 (Database issue) (2008), D440–D444.

[36] The W3C SPARQL Working Group (Ed.). 2013. *SPARQL 1.1 Overview.* W3C Recommendation 21 March 2013. Available from http://www.w3.org/TR/sparql11-overview.

[37] Elodia Thieblin, Ollivier Haemmerle, Nathalie Hernandez, and Cassia Trojahn dos Santos. 2020. Survey on complex ontology matching. *Semantic Web* (2020), 689–727.

[38] Tania Tudorache, Csongor Nyulas, Natalya Fridman Noy, and Mark Musen. 2013. Using Semantic Web in ICD-11: Three Years Down the Road. In *The Semantic Web – ISWC 2013 – 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II (Lecture Notes in Computer Science)*, Harith Alani, Lalana Kagal, Achille Fokoue, Paul T. Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha F. Noy, Chris Welty, and Krzysztof Janowicz (Eds.), Vol. 8219. Springer, 195–211.

[39] Denny Vrandecic and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

[40] Jesse Weaver and Paul Tarjan. 2013. Facebook Linked Data via the Graph API. *Semantic Web* 4, 3 (2013), 245–250.