Towards a Knowledge Graph for Science

Invited Article*

Sören Auer
TIB Leibniz Information Centre for Science and Technology and L3S Research Centre at University of Hannover
Hannover, Germany
auer@tib.eu

Viktor Kovtun
L3S Research Centre, Leibniz University of Hannover
Hannover, Germany
kovtun@l3s.de

Manuel Prinz
TIB Leibniz Information Centre for Science and Technology
Hannover, Germany
manuel.prinz@tib.eu

Anna Kasprzik
TIB Leibniz Information Centre for Science and Technology
Hannover, Germany
anna.kasprzik@tib.eu

Markus Stocker
TIB Leibniz Information Centre for Science and Technology
Hannover, Germany
markus.stocker@tib.eu

Maria Esther Vidal
TIB Leibniz Information Centre for Science and Technology and L3S Research Centre at University of Hannover
Hannover, Germany
maria.vidal@tib.eu

ABSTRACT

The document-centric workflows in science have reached (or already exceeded) the limits of adequacy. This is emphasized by recent discussions on the increasing proliferation of scientific literature and the reproducibility crisis. This presents an opportunity to rethink the dominant paradigm of document-centric scholarly information communication and transform it into knowledge-based information flows by representing and expressing information through semantically rich, interlinked knowledge graphs. At the core of knowledge-based information flows is the creation and evolution of information models that establish a common understanding of information communicated between stakeholders as well as the integration of these technologies into the infrastructure and processes of search and information exchange in the research library of the future. By integrating these models into existing and new research infrastructure services, the information structures that are currently still implicit and deeply hidden in documents can be made explicit and directly usable. This has the potential to revolutionize scientific work as information and research results can be seamlessly interlinked with each other and better matched to complex information needs. Furthermore, research results become directly comparable and easier to reuse. As our main contribution, we propose the vision of a knowledge graph for science, present a possible infrastructure for such a knowledge graph as well as our early attempts towards an implementation of the infrastructure.

*This invited article accompanies Sören Auer's WIMS2018 keynote.

WIMS '18, June 25–27, 2018, Novi Sad, Serbia

ACM ISBN 978-1-4503-5489-9/18/06…$15.00
https://doi.org/10.1145/3227609.3227689

CCS CONCEPTS

• **Information systems → Web searching and information discovery**;

KEYWORDS

Knowledge Graph, Science and Technology, Research Infrastructure, Libraries, Information Science

ACM Reference Format:
Sören Auer, Viktor Kovtun, Manuel Prinz, Anna Kasprzik, Markus Stocker, and Maria Esther Vidal. 2018. Towards a Knowledge Graph for Science: Invited Article. In *WIMS '18: 8th International Conference on Web Intelligence, Mining and Semantics, June 25–27, 2018, Novi Sad, Serbia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3227609.3227689

1 INTRODUCTION

The communication of scholarly information is document-centric. Researchers produce essays and articles that are made available in online and offline publication media as text documents. The entire library, technology, service and research landscape is geared towards this fundamental approach. This approach may indeed be justified if questions can be answered by individual articles. However, increasingly often answers do not just span multiple articles but also multiple scientific domains. In these cases researchers are not adequately supported by the existing infrastructure. Currently, in a best case scenario researchers obtain large, disordered amounts of more or less relevant documents, or more generally, digital objects.

With the current developments in areas such as knowledge representation, semantic search, human-machine interaction, natural language processing, and artificial intelligence it has become possible to completely rethink this dominant paradigm of document-centric information flows and transform scholarly communication into knowledge-based information flows by expressing and representing information as structured, interlinked and semantically rich knowledge graphs.

Ehrlinger and Wöß [10] have discussed the term "knowledge graph" and proposed a definition based on an analysis of current related research work. According to the authors a "knowledge graph acquires and integrates information in an ontology and applies a reasoner to derive new knowledge". The authors underline that "an ontology does not differ from a knowledge base", meaning that following their definition a knowledge graph acquires and integrates information in a kind of knowledge base. Since it is not within the scope of this article, we make no further attempt to refine this definition or provide an alternative. However, we suggest a specialization to science by proposing that a knowledge graph for science acquires and integrates scientific information in a knowledge base, and may apply a reasoner or other computational methods to derive new information.

The variegated problems of document-centric information flows are fairly obvious. Among others, the expansion of scientific literature[1] makes it increasingly difficult to keep an overview of the current state of research. Furthermore, the creation, reading, and processing of scientific literature is tying up cognitive capacity. The ambiguity, intransparency and redundancy of publications also contributed to a lack of reproducability of research culminating in the reproducibility crisis [18]. A further problem rests in the fact that although the characters, words, and sentences can be indexed and searched, the structure and semantics of text, illustrations, references, symbols, etc. are currently hardly accessible to computers.

As our main contribution, we propose the vision of a knowledge graph for science, present a possible infrastructure for such a knowledge graph (Section 2) as well as our early attempts towards an implementation of this infrastructure (Section 3). We highlight some avenues for future work (Section 4) and provide a brief review of related work (Section 5).

2 SCIENCE GRAPH INFRASTRUCTURE

The science graph is a knowledge graph for scholarly communication. It is the core of a socio-technical infrastructure that develops and maintains the graph and operates services. The science graph represents scientific information. It does not merely link (metadata about) people, documents, datasets, institutions, grants, etc. but rather represents research contributions semantically, i.e., explicitly and formally. While an ontology that conceptualizes research contributions is yet to be developed, a semantic description for a research contribution should, at the very least, relate the problems) addressed by the contribution with the methods used and results obtained. Problems, methods, and results are semantic resources themselves. For instance, a result such as the statement "There is a significant difference in the mean duration of a phenomenon X (e.g., particle formation in the atmosphere) between winter and summer seasons" is not merely a natural language sentence but an identified semantic resource in the knowledge graph. This resource is furthermore linked to the research contribution (and thus the authors and affiliation) and the methods used to obtain the result.

The science graph is populated and curated by the infrastructure via four complementary sources. First, the infrastructure leverages existing metadata, data, taxonomies, ontologies, and information models. Second, it provides services that enable direct contributions from scientists who describe their research, supported by intelligent interfaces and automatically generated suggestions. Third, it implements automated methods for information extraction and linking. Fourth, it supports curation and quality assurance by domain experts, librarians and information scientists.

We argue that in order to succeed it is necessary to combine these different sources and curatorial methods. Automated procedures do not achieve the necessary coverage and accuracy while fully manual curation is too time-consuming. Moreover, librarians and information scientists lack the necessary domain expertise while scientists lack the necessary expertise in knowledge representation. By combining the four strategies we can bring their respective strengths to bear and compensate for relative weaknesses.

The science graph infrastructure provides services for interlinking, integration, visualization, exploration, and search. It enables scientists to gain a much faster overview of new developments in a specific field and identify relevant research problems. It represents the evolution of the scientific discourse in the individual disciplines and enables scientists to make their work more accessible to colleagues as well as partners in industry, policy, and society at large.

We suggest that a minimally viable infrastructure must comprise the following technical components. First, a data model for representing scholarly communication semantically. The data model can adopt RDF and Linked Data as a scaffold, but must add comprehensive provenance, evolution, and discourse information. Second, the infrastructure must include a scalable graph-storage backend to store information and expose a comprehensive API for interacting with the knowledge graph. Third, we require user interface widgets and components for collaborative authoring and curation of the graph and integration of these widgets into third-party services. Finally, the infrastructure must support semi-automated semantic integration, search, extraction, and recommendation services to support the curation of the knowledge graph.

3 INFRASTRUCTURE IMPLEMENTATION

At its core, the infrastructure consists of a scalable data management system with a flexible graph-based data model that can be accessed via lightweight APIs. To ensure maximum interoperability, it implements the long-established open standards RDF, RDF Schema, OWL, and Linked Data as well as W3C Data on the Web and the FAIR Data Principles. A central aspect is the preservation of provenance and evolution, so that changes can be tracked transparently at any time. The user interface supports flexible elements, which can be contributed by advanced users themselves to enable customized domain-specific interactions.

3.1 Ontology

As already stated by Ehrlinger and Wöß [10], ontologies are core elements of a knowledge graph insofar as that all information that is acquired as an input for the graph is integrated into a network of ontologies underlying the graph. In information science, the term "ontology" has many definitions – a majority of them build on the formulation originally proposed by Gruber [14] who defined an ontology as "an explicit specification of a conceptualization".

[1] National Science Foundation: Science and Engineering Publication Output Trends: https://www.nsf.gov/statistics/2018/nsf18300/nsf18300.pdf

Thus, since ontologies are the base of the conceptualization of scholarly communication they are a core element of the science graph infrastructure. Attempts to "conceptualize science" using methods from knowledge engineering are, predictably, not entirely novel, and there are already several existing suggestions for ontologies trying to cover the scientific research process.[2] However, fundamental questions such as: "What is research?", "What are the contents of scholarly communication?", "What are the relevant components of a research contribution?" are notoriously difficult to answer and answers are continuously hard to formalize.

We have decided to set out focusing on "research contribution" as an abstract central concept of a possible top-level ontology, postponing considerations of feasibility and usefulness until we have gathered more experience in order to tackle those questions. We suggest that a research contribution communicates one or more results in an attempt to address one or more problems using a set of methods. This nucleus is certainly subject to review and extension. The next necessary step will be to define specific and adequate knowledge engineering workflows for the development of a core ontology (or, more extensively, a network of top-level and domain ontologies) that can be used as the base of the science graph infrastructures in order to support the storing of information.

It is fairly obvious that any abstract concept and top-level ontology will need to be specialized and branched out for different fields of science. What we call "problem" may be more commonly known as "hypothesis" in the natural sciences and "research question" in engineering. Furthermore, these specialized concepts may be conceptualized differently, i.e., feature different attributes and accordingly entail different conclusions concerning their subconcepts. It is even less obvious how concepts are established and determined. A top-down approach whereby a small group of experts designs top-level and domain-specific ontologies as well as their alignment with existing ontologies seems to be a daunting task with uncertain outcome. A bottom-up approach whereby concepts, relations, and conceptualizations are crowd-sourced, thus emerging from the submission of semi-structured data by researchers, may be an interesting way to derive an ontology but comes with its own challenges, for instance the acquisition of the resources that are needed for a continuous curation and formalization of the submitted data.

The design of a practical ontology engineering workflow that finds the right balance between those options and incorporates as many of their positive aspects as possible will need careful attention and should occupy a large portion of the next phase in the construction of the science knowledge graph.

3.2 Backend

The backend features a layered architecture consisting of three layers: application layer, domain layer, and persistence layer. Figure 1 provides an overview of the layers and components.

Inspired by the Hexagonal Architecture [8], the application layer contains ports and adapters. These implement the interface to the outside world and contain the application logic needed so that clients can access the information contained in the knowledge graph. The domain layer contains the domain model from which the knowledge graph is built. It also contains the authentication and authorization component that allows us to store provenance information as part of the domain model. As the lowest layer in the architecture, the persistence layer is responsible for storing all data. Since we currently evaluate different storage options, the persistence layer includes abstractions that can be implemented differently for the respective storage solutions.

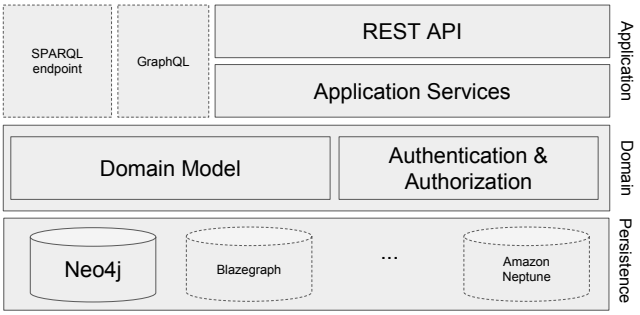

Figure 1: Architecture Diagram. The figure shows all layers and adapters. Possible adapters or storage solutions that are not currently available are displayed with dashed lines.

The backend is implemented around a data model that builds on RDF. Hence, triples consisting of a subject, a predicate and an object are the elementary entities. They also contain provenance information, such as the time of creation and the author. Resources are entities identified by an ID and carry a label for display purposes. Subjects and predicates are resources and are always referenced by their ID. Objects are either resources or (typed) literal values.

All data inserted into the knowledge graph will be made persistent via a layer that is agnostic of any specific storage technology. The compatibility of the data model with RDF means that data can be translated from and to RDF so that as storage technology one could use an off-the-shelf triple store. However, since we also want to make statements *about* triples which are not well supported in RDF we decided to use a linked property graph (LPG) instead, and accordingly our current implementation uses Neo4j.

Data modifications are preserved and can be queried. Currently, we only allow additions and deletions. Data can be modified and queried via a REST API implemented as an adapter of the application layer. We adopt JSON as the serialization format. Other possible adapters include a SPARQL endpoint or a GraphQL interface.

The knowledge graph can be queried openly and without registration. However, users are required to register in order to contribute data to the knowledge graph. Possible queries include the search for resources by ID or label, or the retrieval of lists of statements filtered by resource identifiers. The REST API is currently used by the frontend to power a user interface to the knowledge graph, as well as visualizations.

3.3 Frontend

The user interface provides access to the knowledge graph, specifically research contribution descriptions and resources they related to, currently in two primary forms: hierarchical and graphical. The main page of the user interface includes a search form to allow a

[2] https://derivadow.com/2011/04/19/science-ontology-take-three/

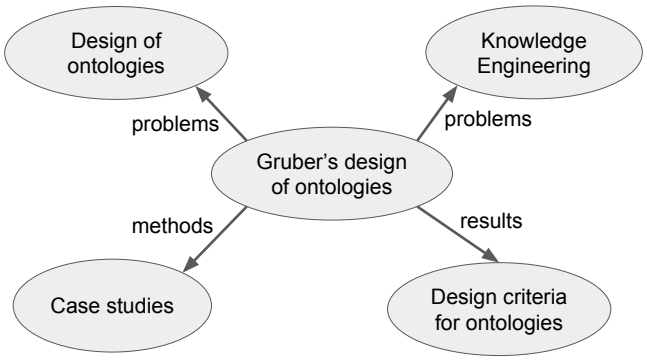

Figure 2: Graphical representation of a research contribution as a knowledge graph

search for resources by their labels. The resulting resources are then displayed in both graphical and hierarchical form.

At the top, the user interface visualizes the graphical form (Figure 2). It displays the found or selected resources as well as their relations to other resources. The graph visualized in this form can be navigated.. Nodes can be selected to display related information. The interface supports navigating from node to node while information about the currently selected node is automatically updated. While navigating, the displayed part of the graph is updated as well so that the user is presented with the nodes that have direct connections to the currently selected one, while other nodes disappear.

In the hierarchical view, the information is displayed in the form of nested lists (Figure 3). Here, the found or selected resources are elements of the outer list while the inner list contains the related resources and literals. Hence, the links in the hierarchical view present an additional way to navigate information. Each link triggers the selection of the corresponding resource. The information about the selected resource is reflected both in the hierarchical and graph views.

The hierarchical view enables users to add new information to the knowledge graph. Submitting information should be straightforward in order to reduce the burden on researchers. Currently, and subject to improvements, each top level resource in the hierarchy displays a button that provides the functionality to add related resources (e.g., a relationship to another resource or literal). When adding a new resource, the user is asked to provide the literal or the title for the resource as well as the relation between them.

4 FUTURE WORK

The work presented here delineates our initial steps towards a knowledge graph for science. By testing existing and developing new components, we have so far focused on some core technical aspects of the infrastructure. Naturally, there are a number of research problems and implementation issues as well as a range of socio-technical aspects that need to be addressed in order to realize the vision. Dimensions of open challenges are, among others:

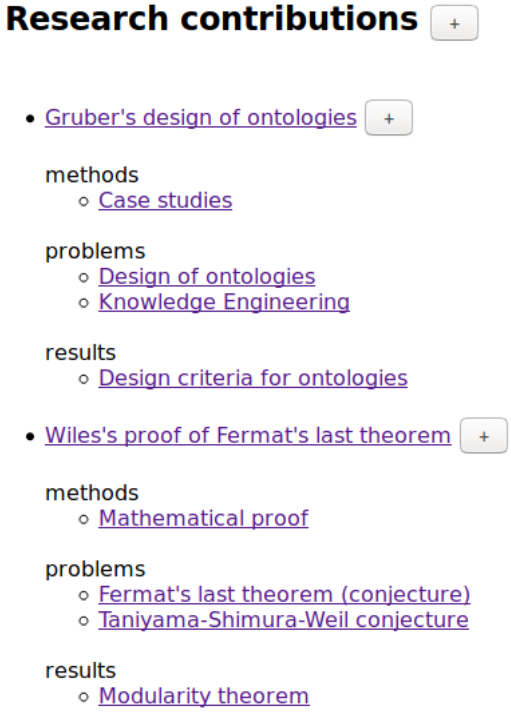

Figure 3: Hierarchical representation of two research contributions as a knowledge graph

- the low-threshold integration of researchers through methods of crowd-sourcing, human-machine interaction, and social networks;
- automated analysis, quality assessment, and completion of the knowledge graph as well as interlinking with external sources;
- support for representing fuzzy information, scientific discourse and the evolution of knowledge;
- development of new methods of exploration, retrieval, and visualization of knowledge graph information.

Several projects have demonstrated how to represent general encyclopedic and factual information in knowledge graphs (see Section 5). An open challenge is how to represent scholarly communication in specialized fields of science. Since precise conceptual structures emerge and evolve over time, the representation of discourse, opinion-forming, and evolution is of particular interest. A knowledge graph for science needs to accommodate fuzzy definitions, diverging opinions, and competing conceptualizations.

The integration of information from documents is critical and relies on natural language processing and information mining methods from text, image, and other media. The maturity of current methods is arguably insufficient to construct a rich knowledge graph from legacy documents in an automated manner. The science graph infrastructure thus relies on numerous complementary approaches to acquire information. Researchers are an important source but must be supported with automated suggestions and recommendations for populating the graph in order to reduce the

manual effort. Another open challenge is the question of how to organize collaboration and interaction among researchers, librarians, information scientists, and knowledge engineers.

An integration of particular interest is the one between the science graph infrastructure and publishers. Addressing the issue of when to best capture contributions to the knowledge graph by researchers, a collaboration with publishers could enable the acquisition of contributions at the time of article submission through the respective submission systems.

Existing scholarly communication incentive measures (e.g., citations, h/i-10 index, impact factor) are document-centric and thus rather coarse-granular. Clearly, we need incentive models for contributions to the knowledge graph for science. We argue that graph-centric measures are an opportunity for a more accurate assessment of scholarly contributions.

5 RELATED WORK

Knowledge graphs such as DBpedia [2], Yago [16] and WikiData [25] as well as similar industrial initiatives by Google, Bing, IBM, BBC, or Thomson Reuters have demonstrated that representing encyclopedic and factual knowledge using RDF and Linked Data is feasible.

However, while there has been a vast amount of work related to representing and managing bibliographic metadata, relatively little work focuses on representing the information contained inside scientific publications semantically. The Semantic Publishing and Referencing (SPAR) Ontologies [21] focus primarily on metadata but also on document structure to some extent.

There has been some work on enriching various document formats with semantic annotations. Examples include Dokie.li [6], RASH [22] or MicroPublications [7] for HTML and SALT [13] for LaTeX. We started representing key findings of survey articles focusing on semantically describing research problems, approaches, implementations and evaluations in [11] and integrating bibliographic information in a knowledge graph [23].

Other work focused on developing ontologies for representing scholarly knowledge in specific domains, for example mathematics [19], the RXNO ontology in chemistry or the OBO Foundry ontologies [24] in the life sciences. A knowledge graph for science must go beyond such efforts, by enabling the parallel and synchronized creation, curation and augmentation of both terminological/ontological as well as assertional and discourse knowledge. For representing provenance and discourse we can build on the PROV ontology [20] and Document Components Ontology [9].

While there has been work on argumentation and reasoning in AI (e.g. [3, 12]) and philosophy (often using specialized formalisms), more work needs to be done to represent argumentation, concept drift and scholarly knowledge evolution in knowledge graphs.

The RDF data model and respective ontologies arguably appear adequate as a scaffold for representing scholarly knowledge. However, aspects such as provenance, evolution and discourse are more difficult to represent in pure RDF (see the ongoing discussion about reification). While there are meanwhile relatively elegant solutions such as RDF singleton properties [26], which can be used for representing and exchanging semantic data, we need to investigate how graph data management techniques (e.g. using the Gremlin graph query algebra [17]) can be employed for storing and managing the extremely large amounts of interconnected scholarly communication data and metadata. Hence, we argue that a knowledge graph for science can build but must extend the triple (or quad) data model in RDF.

The scholarly communication community has initiated numerous related projects. The Research Graph [1] is a prominent example for an effort that aims to link research objects, in particular publications, dataset, researcher profiles. The Scholix project [5], driven by a corresponding Research Data Alliance working group and associated organizations, aims at standardizing the information about the links between scholarly literature and data exchanged among publishers, data repositories, and infrastructures such as DataCite, Crossref, and OpenAIRE.

Other related projects include Research Objects [4], which proposes a machine readable abstract structure that relates the products of a research investigation, including articles but also data and other research artefacts, as well as the RMap Project [15], which aims at preserving "the many-to-many complex relationships among scholarly publications and their underlying data."

6 CONCLUSIONS

The transition from purely document-centric to a more knowledge-based view on scholarly communication is in line with the current digital transformation of information flows in general and is thus inevitable. However, this also creates a need for the implementation of corresponding tools and workflows supporting the switch. As of now, there are still very few of those tools, and their design and concrete features remain a challenge that is yet to be tackled – collaboratively and in a coordinated manner.

ACKNOWLEDGMENTS

The authors would like to thank the participants of a related workshop held at TIB on March 20, 2018 for their contributions to current developments on the Open Research Knowledge Graph, a project recently initiated and coordinated by TIB.

REFERENCES

[1] Amir Aryani and Jingbo Wang. 2017. Research Graph: Building a Distributed Graph of Scholarly Works using Research Data Switchboard. In *Open Repositories CONFERENCE* (2017-06-01). https://doi.org/10.4225/03/58c696655af8a

[2] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *The Semantic Web*. 722–735. https://doi.org/10.1007/978-3-540-76298-0_52

[3] Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. 2015. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation* 6, 1 (2015), 24–49. https://doi.org/10.1080/19462166.2014.1001791

[4] Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble. 2013. Why linked data is not enough for scientists. *Future Generation Computer Systems* 29, 2 (2013), 599–611. https://doi.org/10.1016/j.future.2011.08.004 Special section: Recent advances in e-Science.

[5] Adrian Burton, Hylke Koers, Paolo Manghi, Markus Stocker, Martin Fenner, Amir Aryani, Sandro La Bruzzo, Michael Diepenbroek, and Uwe Schindler. 2017. The Scholix Framework for Interoperability in Data-Literature Information Exchange. *D-Lib Magazine* Volume 23, 1/2 (2017). https://doi.org/10.1045/january2017-burton

[6] Sarven Capadisli, Amy Guy, Ruben Verborgh, Christoph Lange, Sören Auer, and Tim Berners-Lee. 2017. Decentralised Authoring, Annotations and Notifications for a Read-Write Web with dokieli. In *International Conference on Web Engineering*. 469–481. https://doi.org/10.1007/978-3-319-60131-1_33

[7] Tim Clark, Paolo N Ciccarese, and Carole A Goble. 2014. Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. *Journal of Biomedical Semantics* 5, 1 (2014). https://doi.org/10.1186/2041-1480-5-28

[8] Alistair Cockburn. 2018. Hexagonal architecture. http://alistair.cockburn.us/Hexagonal+architecture

[9] Alexandru Constantin, Silvio Peroni, Steve Pettifer, David Shotton, and Fabio Vitali. 2016. The document components ontology (DoCO). *Semantic Web* 7, 2 (2016), 167–181. https://doi.org/10.3233/sw-150177

[10] Lisa Ehrlinger and Wolfram Wöß. 2016. Towards a Definition of Knowledge Graphs. In *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMANTiCS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16)*, Michael Martin, Martí Cuquet, and Erwin Folmer (Eds.), Vol. 1695. CEUR-WS, Leipzig, Germany. http://ceur-ws.org/Vol-1695/paper4.pdf

[11] Said Fathalla, Sahar Vahdati, Sören Auer, and Christoph Lange. 2017. Towards a Knowledge Graph Representing Research Findings by Semantifying Survey Articles. In *Research and Advanced Technology for Digital Libraries*. 315–327. https://doi.org/10.1007/978-3-319-67008-9_25

[12] Thomas F. Gordon and Nikos Karacapilidis. 1997. The Zeno argumentation framework. In *Proceedings of the sixth international conference on Artificial intelligence and law - ICAIL '97*. ACM, 10–18. https://doi.org/10.1145/261618.261622

[13] Tudor Groza, Siegfried Handschuh, Knud Möller, and Stefan Decker. 2007. SALT - Semantically Annotated LaTeX for Scientific Publications. In *Extended Semantic Web Conference*. 518–32. https://doi.org/10.1007/978-3-540-72667-8_37

[14] Thomas R. Gruber. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5, 2 (June 1993), 199–220. https://doi.org/10.1006/knac.1993.1008

[15] Karen L. Hanson, Tim DiLauro, and Mark Donoghue. 2015. The RMap Project: Capturing and Preserving Associations Amongst Multi-Part Distributed Publications. In *Proceedings of the 15th ACM/IEEE-CE on Joint Conference on Digital Libraries - JCDL '15*. ACM, 281–282. https://doi.org/10.1145/2756406.2756952

[16] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence* 194 (2013), 28–61. https://doi.org/10.1016/j.artint.2012.06.001

[17] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust Disambiguation of Named Entities in Text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 782–792. http://dl.acm.org/citation.cfm?id=2145432.2145521

[18] John P. A. Ioannidis. 2005. Why Most Published Research Findings Are False. *PLOS Medicine* 2, 8 (08 2005). https://doi.org/10.1371/journal.pmed.0020124

[19] Christoph Lange. 2013. Ontologies and languages for representing mathematical knowledge on the Semantic Web. *Semantic Web* 4, 2 (2013), 119–158. https://doi.org/10.3233/SW-2012-0059

[20] Timothy Lebo, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. 2013. *PROV-O: The PROV Ontology*. Recommendation. W3C.

[21] Silvio Peroni. 2014. The Semantic Publishing and Referencing Ontologies. In *Semantic Web Technologies and Legal Scholarly Publishing*. Law, Governance and Technology, Vol. 15. Springer, Cham, 121–193. https://doi.org/10.1007/978-3-319-04777-5_5

[22] Silvio Peroni, Francesco Osborne, Angelo Di Iorio, Andrea Giovanni Nuzzolese, Francesco Poggi, Fabio Vitali, and Enrico Motta. 2017. Research Articles in Simplified HTML: a Web-first format for HTML-based scholarly articles. *PeerJ Computer Science* 3 (2017), e132. https://doi.org/10.7717/peerj-cs.132

[23] Afshin Sadeghi, Christoph Lange, Maria-Esther Vidal, and SÃűren Auer. 2017. Integration of Scholarly Communication Metadata Using Knowledge Graphs. In *Research and Advanced Technology for Digital Libraries*. 328–341. https://doi.org/10.1007/978-3-319-67008-9_26

[24] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H Scheuermann, Nigam Shah, Patricia L Whetzel, and Suzanna Lewis. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 25, 11 (2007), 1251–1255. https://doi.org/10.1038/nbt1346

[25] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A Free Collaborative Knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85. https://doi.org/10.1145/2629489

[26] Daya C. Wimalasuriya and Dejing Dou. 2010. Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science* 36, 3 (2010), 306–323. https://doi.org/10.1177/0165551509360123