

Research Contributions: Towards a Formal and Semantic Representation

Simone Angioni^{1*,†}, Allard Oelen^{2,†} and Angelo Salatino^{3,†}

¹CNR-ISTI, Pisa, Italy

²TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany

³Knowledge Media Institute, The Open University, UK

Abstract

Moving beyond the high-level metadata available in bibliographic databases is essential to achieve a richer, more accurate representation of research paper content. Current retrieval and evaluation frameworks rely heavily on basic structural attributes, which fail to capture the nature and intrinsic value of scholarly contributions. To this end, we propose a joint community effort with the workshop audience to collaboratively design and refine a dedicated Research Contribution Ontology. We believe establishing this domain-independent semantic model will provide new and better ways to retrieve content, and fundamentally transform scientific evaluation, supporting qualitative assessment frameworks that move beyond canonical, purely quantitative metrics. In this vision paper, we argue that the time is ripe for this evolution. With the rapid expansion of open-access literature, mature parsing software, and LLMs capable of structured extraction, we now possess both the data and technology. What is missing is the formal semantic representation to link these capabilities together. We present an initial core ontology outline as a concrete foundation for collective development and future integration into scholarly knowledge graphs.

Keywords

Research Contributions Ontology, Scientific Knowledge Graphs, Metascience, Ontology Engineering, Research Contributions

1. Introduction

Digital libraries such as the ACM Digital Library and IEEE Xplore, alongside bibliographic databases such as OpenAlex, Dimensions, and OpenAIRE, routinely index research articles using structural metadata. This metadata, which includes titles, authors, venues, and keywords, forms the backbone of current scholarly systems, crucially enabling both the retrieval of papers and the assessment of their impact through metrics like citation counts. However, relying solely on these high-level attributes limits the overall capacity of these databases. To extend their capabilities without discarding the current foundation, we posit that integrating deeper information regarding the actual content of a paper would lead to significantly improved retrieval accuracy and a much more refined approach to research evaluation. A vital and logical starting point for this content-driven enrichment is the explicit capture of a paper's contributions, clearly articulating the novel knowledge or advancements it has introduced to the wider scientific community. This leads to the following questions: what constitutes as a contribution in a research paper, and how do we formally characterise it?

Currently, various methods classify scholarly articles [1, 2, 3] according to the disciplines discussed in their titles, abstracts, and keywords. These classifiers infer, from the available text, the research concepts, or research areas that best describe the topic of the work, and use them to support indexing, search, and recommendation [4, 5, 6, 7, 8, 9]. However, such classifications primarily capture the paper's generical topics rather than its contributions, such as a new method, dataset, theoretical insight, empirical finding, benchmark, or piece of software.

6th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment, Oct 2026, Bari, Italy

*Corresponding author.

[†]These authors contributed equally.

✉ simone.angioni@isti.cnr.it (S. Angioni); allard.oelen@tib.eu (A. Oelen); angelo.salatino@open.ac.uk (A. Salatino)

id 0000-0002-6682-3419 (S. Angioni); 0000-0001-9924-9153 (A. Oelen); 0000-0002-4763-3943 (A. Salatino)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This distinction is crucial because two papers may belong to the same topic while making different kinds of contributions. Conversely, similar contributions may appear in completely distinct disciplines. Therefore, topic-based ontologies and multidisciplinary classification systems provide only a partial representation of scholarly work. Moreover, building such reliable topic ontologies is itself challenging, since new research areas emerge rapidly, terminology evolves, and novel topics are generally difficult to recognise before they become consolidated in the literature [10]. Capturing the relations between research areas is also non-trivial, especially when articles combine concepts, methods, and problems from multiple fields [11]. Current multidisciplinary classifications often indicate that multiple disciplines are involved, but they rarely specify how these disciplines are related within the work. For instance, whether one discipline provides the method, another the application domain, and another the evaluation setting. As a result, such classifications may assign equal importance to all involved areas while missing the underlying structure of their interaction.

To a certain extent, these classifications were largely sufficient when only abstracts were publicly accessible for most articles. However, with the rapid growth of the Open Access publishing model, up to 50% of articles produced in recent years have been published as Open Access, making it increasingly possible to access the full text of a large corpus of recent literature [12]. This creates an opportunity to gain a broader understanding of how research contributions are developed and structured in the literature. Thanks to open-access publishing and the availability of full-text publications, it is possible to extract contribution-level information and its publication as structured, interoperable metadata, thereby enriching scholarly communication with machine-readable descriptions of the knowledge produced by research and facilitating its integration into scholarly knowledge graphs and other Open Science infrastructures.

Driven by global initiatives such as DORA [13], the Leiden Manifesto [14], and, more recently, CoARA [15], which advocates a transition from purely quantitative, publication-centric assessment to a broader, more qualitative evaluation of researchers' contributions. Researchers should no longer be solely judged by problematic metrics and indicators, such as the h-index or the prestige of the journals and conferences in which they publish, but by the broader impact of their work on advancing all aspects of science. All research contributions should be assessed on the basis of their intrinsic quality and their potential to be reused, built upon, and integrated by others, fostering a more collaborative and cumulative mode of scientific progress. Specifically, CoARA's mission is to revolutionise assessment for both research and researchers. The version of the agreement, signed by 840 research organisations (e.g., universities and research centres) around the world, sets a shared direction for changes in assessment practices. Among the 10 commitments established by the agreement, three of them are focussed in reforming the research assessment from purely quantitative to qualitative assessments:

- Commitment 1: Recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research
- Commitment 2: Base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators
- Commitment 10: Evaluate practices, criteria and tools based on solid evidence and the state-of-the-art in research on research, and make data openly available for evidence gathering and research

These changes lead to the necessity of developing new research assessment frameworks that capture the quality, novelty, and reusability of the research.

In this context, we present the idea behind this vision paper: **Research Contributions Ontology**, a domain-independent model for representing research contributions, their types, relations, methods, evaluations, and evidence. Such an ontology could support scholarly knowledge graphs, research assessment, literature review, AI-powered research assistants, and a more transparent evaluation of scientific works.

The remainder of this paper is structured as follows. Section 2 describes the current state of the art regarding research contributions and prior works on ontologies. Section 3 outlines challenges and possible solutions for developing an ontology on research contributions. Lastly, Section 4 provides

a natural language summary of the article’s core contributions, with the corresponding structured representation detailed in Appendix B.

2. State of the Art

One of the earlier approaches for understanding scientific contributions is the nanopublication [16]. A nanopublication captures a single scientific assertion (according to their RDF schema) along with its provenance (how the assertion was produced) and publication information (who published it and when). This makes individual statements citable, attributable, and machine-readable independently of the paper they come from. However, different from our idea of a scholarly contribution, nanopublications treat single, isolated assertions and have no notion of “contribution”. Each assertion stands on its own, unrelated to other assertions describing the same piece of work, making it difficult to reconstruct which problem was addressed, which artifact was used (or produced), or which claim a given assertion actually supports. Micropublications [17] share the same idea, but with a richer scheme. A claim is supported by a larger list of arguments (i.e., evidence, interpretations, discussions, and challenges). Soldatova and King [18] proposed a general ontology of scientific experiments; the authors developed 218 concepts for experimental design, methodology, and results. The ontology served as a starting point for CoreSC [19], a sentence-level annotation scheme that marks up a scholarly article’s text with the components of a scientific investigation (Hypothesis, Motivation, Goal, Object, Method, Experiment, Result, Conclusion, etc.). Scoro [20], instead, models the roles that agents (authors, funders, reviewers, software) play in producing a specific contribution, rather than the content of the work itself. Dokieli [21] and RASH [22] focus on authoring scholarly articles as semantically annotated HTML: RASH defines a restricted markup language, a subset of HTML, for writing scientific articles, while dokieli is a browser-based tool for authoring and annotating articles as full HTML+RDFa documents. In recent years, various studies have developed domain-specific ontologies [23, 24, 25, 26, 27, 28]. Specifically, the Open Research Knowledge Graph [28] (ORKG) encodes research contributions in a machine-readable format, aiming to describe scholarly articles in a structured manner that extends beyond traditional metadata. Concretely, ORKG seeks to capture the knowledge presented within papers through the lens of their contributions. While their objective appears to align with our proposed Research Contribution Ontology on the surface, ORKG adopts a much broader definition of research contributions. Virtually all knowledge within a paper can be categorised into ORKG contributions, which encompass a wide variety of entities such as materials, problems, methods, and results. Consequently, ORKG contributions form complex structures without a strictly defined schema. In contrast, our more narrowly focussed concept allows us to establish a targeted ontology with a well-defined schema.

3. Research Contributions Ontology

In this section, we examine the utility of an ontology of research contributions. Specifically, we explore how it can be formally described, how it integrates with broader scholarly knowledge infrastructures, and the opportunities it unlocks across a variety of research disciplines.

3.1. Why it matters?

A research contribution ontology is important because the current scholarly ecosystem relies almost entirely on unstructured narrative prose, rendering the rapidly growing global body of knowledge increasingly difficult to navigate, verify, and build upon. As scholarly publishing accelerates, neither traditional keyword indexing nor vector-based semantic retrieval is sufficient to guarantee the precise verification, logical synthesis, or systematic aggregation of complex research contributions across the literature. An ontology provides the explicit semantic framework needed to translate passive literature into an interconnected scientific knowledge graph. This provides the fundamental infrastructure

required for next-generation science: it enables machine-actionable reasoning over claims, simplifies cross-disciplinary discovery, and transforms how we track methodological rigour and scientific progress. Crucially, it anchors modern scientific communication in a structured format that both human researchers and computational agents can reliably process and build upon.

3.2. What is a Research Contribution?

Defining a research contribution is challenging due to the absence of a well-established and widely adopted definition within the literature. Existing proposals across the scholarly domain are often either too granular, conceptually ambiguous, or lack a well-defined schema, making it difficult to capture research contributions as structured, cohesive entities. Nevertheless, establishing clarity is essential to begin formalising an ontological schema that encapsulates research contributions, their distinct types, and their interrelated dependencies. This section provides a preliminary definition intended for refinement through collective feedback at the Sci-K workshop.

For the purposes of this work, a research contribution is defined as *a discrete, evidence-based unit of scholarly knowledge, methodology, artifact, or empirical finding generated by an investigation, which advances the state of the art.*

Identifying research contributions can be a cumbersome task. A common practice for authors is to explicitly list their contributions, often in the introduction. This does not mean that the paper's contributions are limited to the ones explicitly mentioned. Furthermore, more often than not, contributions are not explicitly mentioned but implicitly listed by means of discussing the research questions, challenges, objectives, hypothesis, findings, etc.

3.3. From Research Contribution to an Ontology

Ontology engineering conventionally proceeds from a specification of requirements and scope to a set of competency questions (CQs), being informal queries the ontology is expected to answer. Then drive the choice of classes, relations, and axioms, and are later reused to check whether the resulting model actually answers them [29, 30]. This cycle works well when a domain is well-delimited, and a community largely agrees on what a valid instance of it looks like. However, as previously stated, there is no widely accepted definition of “research contribution”. Moreover, because the ontology is meant to be domain-independent, a CQ that is answerable in one discipline is not guaranteed to be meaningful in another: “which benchmark was used” presupposes a notion of standardised evaluation that fits Computer Science far more naturally than it fits Historical scholarship. A CQ written for a cross-disciplinary model is therefore always a compromise between being specific enough to be testable and general enough to still hold outside the discipline from which it was written.

Research contributions across disciplines share a common skeleton even when their content differs completely. Something is proposed or used, something is done with it, something is found, and something is claimed and defended. We encode this shared skeleton in a small set of deliberately abstract classes and relations, and leave the discipline-specific vocabulary to subclasses and external vocabularies rather than to the core model. *Research Activity* (anything a contribution does to produce a result) and *Result* (anything a Research Activity proposes, finds, or a Contribution uses, or extends) are both abstract and never instantiated directly. As already said, they can be specialized according to domain-specific subclasses (e.g., a computer-science domain could specialise them into *Software*, *Algorithm*, *Benchmark*, *Evaluation*, ...). The relations that connect them to a contribution (*uses*, *producesResult*, ...) are defined once, at the level of the abstract class, and therefore hold regardless of which discipline-specific subclass is actually instantiated. The same logic could be applied to subject classification: rather than modeling a fixed disciplinary taxonomy, *Domain* deliberately points outward to whichever controlled vocabulary is appropriate for the field at hand (ACM CCS, MeSH, PhySH, CSO, ...), so the core ontology stays stable while the domain vocabulary underneath it can vary.

Putting these pieces together gives the model centred on *Contribution*: the specific advancement that a *Research Work* adds to existing knowledge (see Figure 1 in the appendix). A contribution addresses

one or more *Research Problems*, is typed by a *ContributionType* (methodological, empirical, technical, ...), may acknowledge a *Limitation*, uses/extends *Results*, and is supported by *Research Activities* that produce *Results*. It is characterised by *Claims*, each backed by *Evidence* and may *contradicts* previous Contributions. A single abstract *Research Element* superclass unifies every one of these classes for the purpose of *hasDomain* relation, so that a domain tag can be attached to a work, a problem, an artifact, an activity, or a result, and not only to the contribution itself.

Beyond this contribution-centric core, we would like the designed model to be: 1) **interoperable**: rather than modeling subject classification or citation semantics from scratch, the ontology delegates them to established external vocabularies/knowledge graphs; and 2) **extensible**: in the shown example “Result” and “Research Activity” are abstract precisely so that domain communities can plug in their own subclasses, or entire domain ontologies.

The proposed ontology is intended neither as a comprehensive ontology of scholarly communication nor as a replacement for ontologies describing publications, research organisations, datasets, software, or scientific experiments. Its scope is deliberately narrower: it models the contribution layer that connects a scholarly work to the advancement it claims to make, together with the research problem addressed, the activities and results involved, and the claims and evidence through which that advancement is supported. In this sense, a Contribution should not be equated with a publication, result, claim, or research activity. A publication may report several contributions, a contribution may involve several results and claims, and those elements may themselves be described using existing domain or infrastructure ontologies. The purpose of the core ontology is therefore to provide the semantic relations that bind these elements into a coherent unit of scholarly contribution.

3.4. Relation to Scholarly Knowledge Infrastructures

A research contribution ontology serves as a critical structural component for modern scholarly knowledge infrastructures. Major digital platforms and scientific knowledge graphs, such as OpenAlex, OpenAIRE, ORKG and related domain-specific knowledge bases, have successfully advanced the aggregation, linking, and open discovery of bibliographic metadata. However, these systems often operate at the macro-level of documents, authors, institutions, and high-level concepts, leaving the internal substance of a paper largely unstructured. The ORKG platform does focus on a more fine-grained representation of knowledge represented in scholarly articles. However, as outlined in Section 2, their broad description of contributions makes it more difficult to extract core contributions from articles and therefore does not directly support the intended uses cases of our ontology.

By offering a formalised, fine-grained schema for research contributions, this ontology provides the underlying semantic framework that these infrastructures require to represent the actual content of scientific work. Integrating this contribution model enables platforms like OpenAIRE to expand beyond bibliographic and citation networks toward deep semantic mapping of claims, methodologies, and datasets. Similarly, for crowd-sourced and automated curation engines such as ORKG, the ontology offers a standardised conceptual blueprint to harmonise user-defined comparison matrices and automated extractions. ORKG by design supports external ontology integration; therefore, adopting the proposed contribution ontology within their knowledge graph can provide direct benefits in the form of novel use cases. Ultimately, this will pave the way towards transforming scientific knowledge graphs from passive document indexes into active, machine-actionable reasoning environments, fostering seamless interoperability across global scholarly infrastructures.

3.5. Opportunities for Metascience, AI, and Research Assistance

The creation of a formalised machine-readable ontology for research contributions unlocks significant opportunities across metascience, artificial intelligence, and scholarly workflows. By moving beyond unstructured textual representations toward explicit semantic structures, the academic ecosystem can the improve research execution, peer review, automated domain discovery, and meta-research.

In the domain of literature review and automated research assistance, structured ontological representations provide a foundational framework for fine-tuning LLMs [31]. Current NLP models often struggle with domain-specific jargon, shifting terminology, and conceptual ambiguity across subfields [32, 33]. A shared ontology enables models to learn fine-grained relations between topics, methodologies, and findings, allowing AI assistants to map research landscapes accurately even when authors use disparate terms for identical underlying problems. This capability significantly mitigates the cognitive burden and stress on researchers, particularly those exploring unfamiliar disciplines, by automatically surfacing relevant prior work, established benchmarks, utilised code bases, and documented limitations without requiring the exhaustive manual analysis of hundreds of papers.

From a metascience perspective, a formal ontology of research contributions transforms how we study, evaluate, and monitor the scientific ecosystem itself. By mapping contributions into explicit semantic entities, meta-researchers, reviewers, and funding bodies can move beyond coarse bibliographic metrics, such as citation counts, to perform fine-grained analyses of scientific evolution, methodological rigour, and research assessment. This structured framework provides the tools to systematically audit scholarly claims, verify whether an investigation adheres to established domain gold standards, and assess consistency with prior literature. Furthermore, it establishes the structural groundwork for novel scientometric indicators that can track the emergence of new paradigms, quantify the reuse of specific methodologies or code bases, evaluate research culture, and enhance the integrity of peer review with higher precision.

3.6. Why the Time is Ripe?

In recent years, significant technological advancements have transformed the automated extraction and structuring of research contributions from a distant ideal into a highly viable opportunity. First, the development of robust PDF parsing tools, such as GROBID [34], MinerU [35], and Nougat [36], has made it possible to reliably convert static research articles into machine-actionable formats, typically XML or JSON. This progress addresses a historical bottleneck, as older generations of these parsing tools lacked the accuracy required for high-fidelity semantic tasks.

Furthermore, the emergence of LLMs represents a major pivotal point. Through prompt engineering, we can now easily design new solutions to extract complex contributions and automatically organise them according to a predefined ontological schema. Due to the increased size and capabilities of LLMs, they can produce highly structured outputs in predefined schemas via targeted extractions. Beyond basic prompting, we can also employ more complex, programmatic orchestration frameworks. An example is the Prompt Declaration Language [37], a declarative programming framework that allows developers to define structured prompt templates, control LLM call sequences with programmatic loops and conditionals, and enforce strict output schemas, thereby ensuring highly reliable, deterministic, and reproducible extraction pipelines. Combined with the technological advancements, the more widespread availability of open-access articles creates new opportunities for performing contribution extraction at scale.

4. Conclusion

In this vision paper, we have argued that current technologies are mature enough to extract, harvest, organise, and build upon the diverse research contributions outlined within academic papers. However, we are currently missing a formal and semantic framework to represent them. We firmly believe that developing a dedicated ontology will serve as a crucial stepping stone for a new generation of intelligent agents. These agents could then move away from flawed traditional metrics to evaluate research quality accurately, assist researchers in identifying knowledge gaps and generating hypotheses, and ultimately empower AI companions to develop a far deeper comprehension of the existing literature. Moreover, this ontology could serve as a starting point for a broader class of ontologies that aim to capture specific aspects of scholarly articles, e.g., rigour, impact, reproducibility; in domain-agnostic manners. This leads to practical, less formal ontologies that are easier to create and particularly suitable for AI reuse.

Beyond this vision paper, we are keen to collaborate with the Sci-K 2026 audience to begin designing a comprehensive ontology of research contributions. This shared framework must be capable of capturing the unique nuances of diverse academic fields, providing the vital bedrock for the future curation and consumption of Scientific Knowledge Graphs.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to improve the language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- [1] A. A. Salatino, F. Osborne, T. Thanapalasingam, E. Motta, The cso classifier: Ontology-driven detection of research topics in scholarly articles, in: *International Conference on Theory and Practice of Digital Libraries*, Springer, 2019, pp. 296–311.
- [2] R. You, Y. Liu, H. Mamitsuka, S. Zhu, Bertmesh: deep contextual representation learning for large-scale high-performance mesh indexing with full text, *Bioinformatics* 37 (2021) 684–692. doi:10.1093/bioinformatics/btaa837.
- [3] N. Gialitsis, S. Kotitsas, H. Papageorgiou, Scinobo: A hierarchical multi-label classifier of scientific publications, in: *Companion Proceedings of the Web Conference 2022, WWW '22*, ACM, 2022, p. 800–809. doi:10.1145/3487553.3524677.
- [4] P. Manghi, C. Atzori, A. Bardi, M. Baglioni, H. Dimitropoulos, S. La Bruzzo, I. Foufoulas, A. Mannocci, M. Horst, K. Iatropoulou, A. Kokogiannaki, M. De Bonis, M. Artini, A. Lempesis, A. Ioannidis, N. Manola, P. Principe, T. Vergoulis, S. Chatzopoulos, OpenAIRE Graph Dataset, 2026. doi:10.5281/ZENODO.3516917.
- [5] M. Stocker, A. Oelen, M. Y. Jaradeh, M. Haris, O. A. Oghli, G. Heidari, H. Hussein, A.-L. Lorenz, S. Kabenamualu, K. E. Farfar, M. Prinz, O. Karras, J. D'Souza, L. Vogt, S. Auer, FAIR scientific information with the Open Research Knowledge Graph, *FAIR Connect* 1 (2023) 19–21. doi:10.3233/FC-221513.
- [6] P. Knoth, D. Herrmannova, M. Cancellieri, L. Anastasiou, N. Pontika, S. Pearce, B. Gyawali, D. Pride, CORE: A Global Aggregation Service for Open Access Papers, *Scientific Data* 10 (2023) 366. doi:10.1038/s41597-023-02208-w.
- [7] J. Priem, H. Piwowar, R. Orr, OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts, 2022. doi:10.48550/ARXIV.2205.01833.
- [8] S. Angioni, A. Salatino, F. Osborne, A. Birukou, D. R. Recupero, E. Motta, Leveraging Knowledge Graph Technologies to Assess Journals and Conferences at Springer Nature, in: U. Sattler, A. Hogan, M. Keet, V. Presutti, J. P. A. Almeida, H. Takeda, P. Monnin, G. Pirrò, C. d'Amato (Eds.), *The Semantic Web – ISWC 2022*, volume 13489, Springer International Publishing, Cham, 2022, pp. 735–752. doi:10.1007/978-3-031-19433-7_42.
- [9] S. Angioni, A. Salatino, F. Osborne, D. R. Recupero, E. Motta, AIDA: A knowledge graph about research dynamics in academia and industry, *Quantitative Science Studies* 2 (2021) 1356–1398. doi:10.1162/qss_a_00162.
- [10] A. Salatino, T. Aggarwal, A. Mannocci, F. Osborne, E. Motta, A survey of knowledge organization systems of research fields: Resources and challenges, *Quantitative Science Studies* 6 (2025) 567–610. doi:10.1162/qss_a_00363.
- [11] T. Aggarwal, A. Salatino, F. Osborne, E. Motta, Leveraging large language models for generating research topic ontologies: A multi-disciplinary study, *ACM Trans. Intell. Syst. Technol.* (2026). doi:10.1145/3828662.
- [12] European Commission. Directorate-General for Research and Innovation, Visionary Analytics, Technopolis Group, Opix, Timelex, Ernst & Young, J. Antanavičius, A. Sadauskaitė, M. Kalanta,

- G. Pozzo, I. Bernotas, A. Elenskè, M. Eggert, L. Stoy, F. Blandinieres, A.-S. Chekaf, E. Seminaroti, N. Gruber, Y. Aslan, N. Manola, I. Grypari, M. Katsamakis, P. Tsiavos, A. Lempesis, L. Boghaert, A. Sachdev, P. Freguglia, K. Kazmierska, Economic Analysis of Options for Improving EU Legislative and Regulatory Frameworks with Impact on Access and Reuse of Publicly Funded R&I Results and of Publications and Data for Scientific Purposes, Publications Office of the European Union, LU, 2026. doi:10.2777/8969719.
- [13] San Francisco Declaration on Research Assessment, The San Francisco Declaration on Research Assessment, 2013. URL: <https://sfdora.org/read/>, accessed: April 28, 2026.
 - [14] Diana Hicks, P. Wouters, The Leiden Manifesto for research metrics, *Nature* 520 (2015) 9–11. doi:10.1038/520429a.
 - [15] M. Arentoft, S. Berghmans, L. Borrell-Damian, S. Bottaro, J.-E. Faure, V. Gaillard, K. Glinos, J. L. Albacete, R. Morais, J. Morris, M. Schiltz, K. Stroobants, Agreement on Reforming Research Assessment (2022). doi:10.5281/ZENODO.13480728.
 - [16] P. Groth, A. Gibson, J. Velterop, The anatomy of a nanopublication, *Information Services and Use* 30 (2010) 51–56. doi:10.3233/ISU-2010-0613.
 - [17] T. Clark, P. N. Ciccarese, C. A. Goble, Micropublications: A semantic model for claims, evidence, arguments and annotations in biomedical communications, *Journal of Biomedical Semantics* 5 (2014) 28. doi:10.1186/2041-1480-5-28.
 - [18] L. N. Soldatova, R. D. King, An ontology of scientific experiments, *Journal of The Royal Society Interface* 3 (2006) 795–803. doi:10.1098/rsif.2006.0134.
 - [19] M. Liakata, S. Teufel, A. Siddharthan, C. Batchelor, Corpora for the conceptualisation and zoning of scientific papers, in: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (Eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, 2010. URL: <https://aclanthology.org/L10-1440/>.
 - [20] D. Shotton, S. Peroni, Scoro, the scholarly contributions and roles ontology, 2017.
 - [21] S. Capadisli, A. Guy, R. Verborgh, C. Lange, S. Auer, T. Berners-Lee, Decentralised authoring, annotations and notifications for a read-write web with dokieli, in: *International Conference on Web Engineering*, Springer, 2017, pp. 469–481.
 - [22] S. Peroni, F. Osborne, A. Di Iorio, A. G. Nuzzolese, F. Poggi, F. Vitali, E. Motta, Research articles in simplified html: a web-first format for html-based scholarly articles, *PeerJ Computer Science* 3 (2017) e132.
 - [23] D. Dessì, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, E. Motta, H. Sack, Ai-kg: an automatically generated knowledge graph of artificial intelligence, in: *International semantic web conference*, Springer, 2020, pp. 127–143.
 - [24] D. Dessì, F. Osborne, D. Buscaldi, D. Reforgiato Recupero, E. Motta, Cs-kg 2.0: A large-scale knowledge graph of computer science, *Scientific Data* 12 (2025) 964.
 - [25] Y. Luan, L. He, M. Ostendorf, H. Hajishirzi, Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction, in: *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, pp. 3219–3232.
 - [26] P. Ciccarese, E. Wu, G. Wong, M. Ocana, J. Kinoshita, A. Ruttenberg, T. Clark, The swan biomedical discourse ontology, *Journal of biomedical informatics* 41 (2008) 739–751.
 - [27] A. Bandrowski, R. Brinkman, M. Brochhausen, M. H. Brush, B. Bug, M. C. Chibucos, K. Clancy, M. Courtot, D. Derom, M. Dumontier, et al., The ontology for biomedical investigations, *PloS one* 11 (2016) e0154556.
 - [28] M. Y. Jaradeh, A. Oelen, K. E. Farfar, M. Prinz, J. D'Souza, G. Kismihók, M. Stocker, S. Auer, Open research knowledge graph: Next generation infrastructure for semantic scholarly knowledge, in: *Proceedings of the 10th international conference on knowledge capture*, 2019, pp. 243–246. doi:10.1145/3360901.3364435.
 - [29] M. Grüninger, M. S. Fox, The role of competency questions in enterprise engineering, in: *Benchmarking—Theory and practice*, Springer, 1995, pp. 22–31.
 - [30] M. Uschold, M. King, et al., Towards a methodology for building ontologies, *Artificial Intelligence*

Applications Institute, University of Edinburgh, 1995.

- [31] F. Bolanos-Burgos, A. A. Salatino, F. Osborne, E. Motta, Modelling and classifying the components of literature reviews: a novel annotation schema and evaluation of transformer models, *PeerJ Computer Science* 12 (2026) e3921. doi:10.7717/peerj-cs.3921.
- [32] D. Eitan, M. Pirchi, N. Glazer, S. Meital, G. Ayach, G. Krendel, A. Shamsian, A. Navon, G. Hetz, J. Keshet, Combining language models for specialized domains: A colorful approach, 2023. URL: <https://arxiv.org/abs/2310.19708>.
- [33] M. Amin-Naseri, H. Kim, E. Hruschka, A dynamic self-evolving extraction system, 2026. URL: <https://arxiv.org/abs/2603.06915>.
- [34] P. Lopez, Grobid: Combining automatic bibliographic data recognition and term extraction for scholarship publications., in: *ECDL*, volume 9, 2009, pp. 473–474.
- [35] B. Wang, C. Xu, X. Zhao, L. Ouyang, F. Wu, Z. Zhao, R. Xu, K. Liu, Y. Qu, F. Shang, B. Zhang, L. Wei, Z. Sui, W. Li, B. Shi, Y. Qiao, D. Lin, C. He, Mineru: An open-source solution for precise document content extraction, 2024. URL: <https://arxiv.org/abs/2409.18839>.
- [36] L. Blecher, G. Cucurull, T. Scialom, R. Stojnic, Nougat: Neural optical understanding for academic documents, 2023. *arXiv:2308.13418*.
- [37] M. Vaziri, L. Mandel, C. Spiess, M. Hirzel, Pdl: A declarative prompt programming language, 2024. URL: <https://arxiv.org/abs/2410.19135>.

A. Ontology Outline

In Figure 1, we outline the ontology that we aim to refine or completely redefine during the workshop.

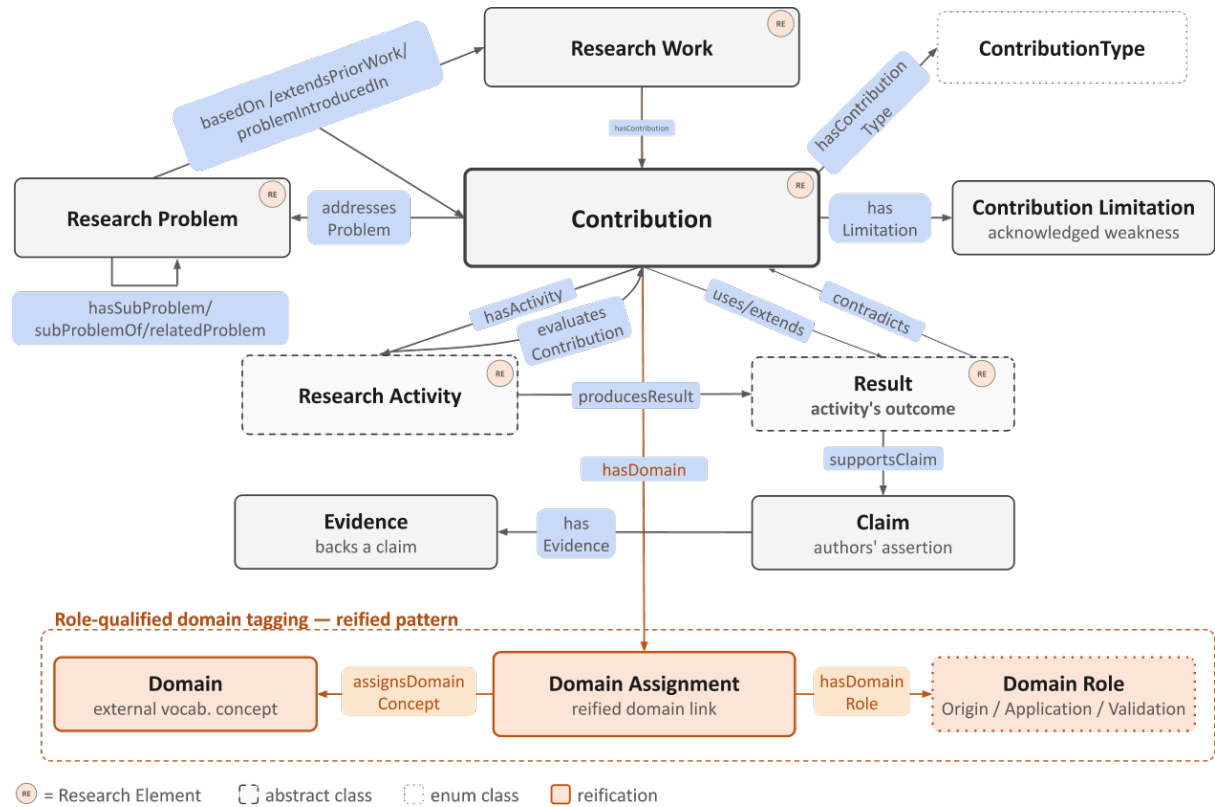


Figure 1: Ontology Diagram

B. Contribution KG of this paper

Table 1 and Table 2 outline a sample of Knowledge Graph extracted of this paper applying the ontology described in Section 3 and shown in Figure 1.

ID	Class	Label / Description
RW1	Research Work	Research Contributions: Towards a Formal and Semantic Representation
RW2	Research Work	ORKG — Open Research Knowledge Graph [28]
C1	Contribution	The proposed Research Contributions Ontology
RP1	Research Problem	What constitutes a research contribution, and how to formally characterise it
RP2	Research Problem	Multidisciplinary classifications don't specify how disciplines relate within one work (method vs. application domain vs. evaluation setting)
RP3	Research Problem	No formal semantic representation exists to link open-access text, mature parsers, and LLMs for structured contribution extraction
RAct1	Research Activity	Ontology specification / conceptual design process (Section 3.3) — generic instance, no matching subclass exists yet (see Notes sheet)
RES1	Result	The Research Contributions Ontology outline itself (Figure 1) — tangible result
CL1	Claim	"The time is ripe": mature parsing tools + LLMs + reproducible extraction frameworks make automated contribution extraction newly viable
CL2	Claim	A compact, discipline-neutral core of abstract classes offers a plausible general solution for modeling research contributions
CL3	Claim	ORKG's broad contribution definition yields complex, schema-less structures; this ontology's narrower focus enables a well-defined, targeted schema
CL4	Claim	Up to 50% of recently produced articles are published Open Access
CL5	Claim	Neither traditional keyword indexing nor vector-based semantic retrieval is sufficient to guarantee precise verification, logical synthesis, or systematic aggregation of complex research contributions
EV1	Evidence	GROBID [34], MinerU [35], Nougat [36] — mature PDF-to-structured-format parsing tools
EV2	Evidence	Prompt Declaration Language [37] — reliable, deterministic, reproducible LLM extraction pipelines
EV3	Evidence	EU Commission economic analysis report [12] — source of the ~50% Open Access figure
EV4	Evidence	Section 2's own account of ORKG's contribution schema (materials, problems, methods, results — no strictly defined schema)
LIM1	Limitation	Not intended to be complete or immediately operational — an initial outline offered for collaborative refinement, not a finished/working ontology
D1	Domain	Ontology Engineering / Semantic Web (Computer Science)
D2	Domain	Metascience / Research Assessment
DA1	Domain Assignment	Links C1 to D1
DA2	Domain Assignment	Links C1 to D2

Table 1

Entities and their class and labels extracted for this article

Subject	Relation	Object
RW1	hasContribution	C1
C1	hasContributionType	"Conceptual"
C1	addressesProblem	RP1
C1	addressesProblem	RP2
C1	addressesProblem	RP3
RP1	relatedProblem	RP3
C1	hasActivity	RAct1
RAct1	producesResult	RES1
C1	hasClaim	CL1
C1	hasClaim	CL2
C1	hasClaim	CL3
C1	hasClaim	CL4
C1	hasClaim	CL5
CL1	hasEvidence	EV1
CL1	hasEvidence	EV2
CL4	hasEvidence	EV3
CL3	hasEvidence	EV4
CL3	contradicts	RW2
C1	hasLimitation	LIM1
C1	hasDomain	DA1
DA1	assignsDomainConcept	D1
DA1	hasDomainRole	"Origin"
C1	hasDomain	DA2
DA2	assignsDomainConcept	D2
DA2	hasDomainRole	"Application"

Table 2
Entities and their relation