

The 6th International Workshop on Scientific Knowledge (Sci-K): Representation, Discovery, and Assessment

Anna M. Jacyszyn¹, Andrea Mannocci², Allard Oelen³, Francesco Osborne⁴, Georg Rehm⁵, Angelo Salatino⁴, Sonja Schimmler⁶ and Lise Stork⁷

¹FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen (DE)

²Institute of Information Science and Technologies, Italian Research Council (CNR-ISTI), Pisa (IT)

³TIB – Leibniz Information Centre for Science and Technology, Hannover (DE)

⁴The Open University, Milton Keynes (UK)

⁵Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Berlin (DE); Humboldt-Universität zu Berlin (DE)

⁶TU Berlin, Fraunhofer FOKUS, Berlin (DE)

⁷Informatics Institute, University of Amsterdam, Amsterdam (NL)

Abstract

Proceedings of the 6th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment; co-located with 25th International Semantic Web Conference (ISWC 2026), 26 October 2026, Bari, Italy. There were 29 papers submitted for peer-review to this workshop. Out of these, 14 papers were accepted for this volume, 8 as regular papers and 6 as short papers.

1. Preface

The International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment (Sci-K 2026) is an established workshop now at its sixth edition and co-located with the 25th International Semantic Web Conference, a premier conference in the *Semantic Web*. The Sci-K workshop brings together researchers and practitioners from diverse fields—including, but not limited to, Digital Libraries, Information Extraction, Machine Learning, Semantic Web, Knowledge Engineering, Natural Language Processing, Scholarly Communication, Science of Science, Scientometrics and Bibliometrics—as well as industry professionals, to explore innovative solutions and ideas for the production and consumption of Scientific Knowledge Graphs (SKGs) and assessing their impact on research. The workshop invited high-quality submissions centred on three main themes in the study of scientific knowledge: *representation*, *discovery*, and *assessment*. The topics of Sci-K are more relevant than ever, as the volume of scholarly articles keeps rising rapidly every year. Especially with the increase in AI usage for scholarly research, there is a growing need for structured, machine-readable knowledge representations.

The call for papers for Sci-K 2026 opened in April with a submission deadline on 24 July. The deadline was later extended until 5 August. In response to the call for papers, the workshop has received 29 submissions from researchers in 13 different countries: Canada, Estonia, France, Germany, India, Ireland, Italy, Spain, Sweden, Switzerland, Türkiye, United Kingdom, and USA (listed in alphabetic order). Each

Sci-K 2026: 6th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment

Authors contributed equally.

✉ anna.jacyszyn@fiz-karlsruhe.de (A. M. Jacyszyn); andrea.mannocci@isti.cnr.it (A. Mannocci); allard.oelen@tib.eu (A. Oelen); francesco.osborne@open.ac.uk (F. Osborne); georg.rehm@dfki.de (G. Rehm); angelo.salatino@open.ac.uk (A. Salatino); sonja.schimmler@tu-berlin.de (S. Schimmler); l.stork@uva.nl (L. Stork)

🌐 <https://www.fiz-karlsruhe.de/en/bereiche/lebenslauf-und-publicationen-dr-anna-jacyszyn> (A. M. Jacyszyn); <https://andremann.github.io> (A. Mannocci); <https://www.tib.eu/en/research-development/open-research-knowledge-graph/service-and-software-development/staff/allard-oelen> (A. Oelen); <https://people.kmi.open.ac.uk/francesco> (F. Osborne); <http://georg-rehm> (G. Rehm); <https://salatino.org> (A. Salatino); <https://sites.google.com/view/sonjaschimmler/> (S. Schimmler); <https://lisestork.github.io/> (L. Stork)

🆔 0000-0002-5649-536X (A. M. Jacyszyn); 0000-0002-5193-7851 (A. Mannocci); 0000-0001-9924-9153 (A. Oelen); 0000-0001-6557-3131 (F. Osborne); 0000-0002-7800-1893 (G. Rehm); 0000-0002-4763-3943 (A. Salatino); 0000-0002-8786-7250 (S. Schimmler); 0000-0002-2146-4803 (L. Stork)



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

paper was reviewed by at least three members of the program committee. Given the quality and the interesting topics covered by the submissions, on August 26 we announced the acceptance of a total of 14 papers: 8 long papers, and 6 short ones. The full program can be found on the Sci-K website: <https://sci-k.github.io/2026>. Accepted papers are listed with their abstracts in Section 3.

2. Sci-K Workshop Series

Sci-K 2026 builds on five previous successful editions (Sci-K 2021 to 2025) and keeps attracting a broad and diverse pool of attendees. The first edition (Sci-K 2021) was held on 13 April 2021 in conjunction with The Web Conference 2021. Its program consisted of two keynote talks and the presentation of 11 research papers. The second edition (Sci-K 2022) took place on 26 April 2022 at The Web Conference 2022. The program included the presentation of 5 long papers, 4 short papers, 2 vision papers, 2 keynote talks and a panel on “What’s next after Microsoft Academic Graph?” [1]. The third edition (Sci-K 2023) took place on 30 April 2023 at the 22nd International Web Conference. The program included the presentation of 4 long papers, 6 short papers, and a keynote speech from Matt Buys, Executive Director of DataCite, on the topic “Scaling the Global Data Citation Corpus: An International Collaboration” [2]. The fourth edition (Sci-K 2024) took place on 12 November 2024 at The 23rd International Semantic Web Conference, in Baltimore, MD, USA. The program included the presentation of 9 contributions: 7 long research papers and 2 short ones. This edition also featured a keynote speech given by Natasha Noy titled “Linking scientific knowledge: A dataset-centric view” [3].

The fifth edition (Sci-K 2025) took place on 2 November 2025, co-located with the 24th International Semantic Web Conference in Nara, Japan. The programme featured 13 papers, comprising 7 long and 6 short contributions, alongside a keynote speech delivered by Dr. Jian Wu entitled “Deep Mining Scholarly Big Data in the Large Language Model Era” [4]. This was followed by a one-hour interactive discussion exploring the key challenges and opportunities surrounding scientific knowledge graphs, which continued through subsequent meetings and led to a submission to AI Magazine. The event also received generous sponsorship of 2,000 USD from Digital Science¹, a world-leading company that serves the global research community by developing integrated solutions that span the entire research lifecycle, from discovery and grant funding to analysis and publishing.

Moreover, the Sci-K series² spawned from the collaboration of two former workshops, namely, the *Scientific Knowledge Graphs Workshop (SKG 2020)*, and the *Workshop on Assessing Impact and Merit in Science (AIMinScience 2020)*, held in conjunction with the 2020 edition of the *International Conference on Theory and Practice of Digital Libraries (TPDL)* on the 25th of August 2020 [5]. Accepted papers from these two workshops have been invited to submit extended versions to a special issue we guest-edited at the open-access journal *Quantitative Science Studies (QSS)*, a leader in the field of *Science of Science* and *Scientometrics*. The special issue, titled “Scientific Knowledge Graphs and Research Impact Assessment”, provides an overview of recent advances in scholarly knowledge representation and research assessment featuring 10 peer-reviewed articles [6]. The synergy between the SKG 2020 and AMinScience 2020 workshops and this QSS special issue helped define the focus and goals of the Sci-K workshop series.

3. Accepted Papers

Here, we report the accepted papers and their abstract, grouping them according to the main themes of the workshop.

3.1. Representation

- **Supporting the Understanding of Ontologies for Scientific Knowledge Graphs with the New Version of LODÉ** (Valentina Pasqual, Ahmadreza Nazari and Silvio Peroni) [7]

¹Digital Science — <https://www.digital-science.com/>

²Sci-K History — <https://sci-k.github.io/history>

Abstract: Scientific Knowledge Graphs (SKGs) rely on increasingly complex and heterogeneous semantic models, whose reuse and interoperability require ontologies and other semantic artefacts to be understandable by both machines and humans. However, existing ontology documentation tools provide limited support for modern reuse-oriented modelling practices, where entities are distributed across imported modules and enriched through cross-module annotations. This paper presents the new version of LODE, a re-engineered and extensible framework for generating human-readable documentation of semantic artefacts. LODE preserves the established single-page, W3C Recommendation-style documentation paradigm of its legacy predecessor, and introduces a modular architecture that separates artefact interpretation (Reader), internal representation (Model), and rendering (Viewer). Served as a web service, LODE supports the documentation of OWL ontologies and enables additional documentation features (e.g., entity stand-alone documentation, RDF entity provenance, Markdown rendering). We demonstrate the applicability of LODE through the documentation of the Scientific Knowledge Graph Interoperability Framework ontology (SKG-O), showing how it improves the accessibility and reuse of modular scientific knowledge graph models.

- **Research Contributions: Towards a Formal and Semantic Representation** (*Simone Angioni, Allard Oelen and Angelo Salatino*) [8]

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.

- **SUDO: From Scientific Publications to Global Scholarly Discourse** (*Nipun Pathirage, Henrique Santos and Deborah L. McGuinness*) [9]

Abstract: Scholarly Knowledge Graphs (SKGs) support literature search, scientific discovery, and agent-assisted research, yet most pre-specify the knowledge unit to be represented—typically named scholarly entities or contribution-level statements—and retain only the publication content expressible through that unit. This choice produces two dominant but partial paradigms: concept-centric SKGs preserve entities and local relations, but not their roles within the paper’s research narrative; statement-centric SKGs capture high-level claims, but not the entity-level detail and source grounding needed to connect knowledge across publications. In both cases, the graph reflects a chosen extraction target rather than how a paper constructs and communicates knowledge. We introduce SUDO, the Scholarly Upper Discourse Ontology, which changes what the ontology specifies: rather than prescribing which domain knowledge to extract, it defines how knowledge elements are organized and related to form a paper’s research narrative—its scholarly discourse. Using linguistic principles that generalize across fields and tasks, SUDO enables each publication’s scholarly discourse to be constructed from the bottom up as a coherent, source-grounded representation. Applied consistently across papers, the same structure connects their research narratives into a global scholarly discourse without discarding local context or provenance. We demonstrate SUDO’s capabilities across a comprehensive suite of scholarly knowledge modeling use cases and performance improvements on a downstream LLM-based

scientific QA task.

- **Representing Evidence of Emotional Introspection in Large Language Models: A Scientific Knowledge Graph Research Agenda** (*Rashid Hmaidan, Antonio Puertas Gallardo, Sergio Consoli, Alessandro Zani, Mario Ceresa and Diego Reforgiato*) [10]

Abstract: Research on emotional introspection in Large Language Models (LLMs) — whether a model can detect, report, and regulate the internal emotion-related states that shape its outputs — is growing rapidly but is strikingly fragmented. Findings are reported through heterogeneous terminology, model families, units of internal analysis, intervention techniques, evaluation protocols, and standards of causal evidence, and studies routinely conflate emotionally appropriate output, latent representation, causal influence, introspective access, and self-regulation, although these rest on fundamentally different levels of evidence. This position paper argues that progress here requires not only new experiments but a semantic representation and assessment framework for the evidence itself. We treat a structured review of the literature as a domain analysis, identifying the concepts, entities, relationships, and evidence types to be represented, and propose an Emotional Scientific Knowledge Graph (E-SKG): a provenance-aware, statement-centric graph of emotion constructs, internal model components, interventions, reported states, behavioural effects, experiments, datasets, evaluation methods, claims, and publications. Building on existing scientific knowledge graphs, our contribution is not an implementation but a research agenda: a conceptual semantic model, seven representation requirements, competency questions, and future directions, closing with an outlook on such a resource as infrastructure for trustworthy, emotionally-aware AI in healthcare and surgical robotics.

- **Open-Schema Metadata Extraction from GitHub READMEs: A Cross-Model Agreement Study** (*Genet Asefa Gesese, Zongxiong Chen, Sonja Schimmler and Harald Sack*) [11]

Abstract: GitHub README files contain information needed to discover and reuse scientific artifacts, yet much of it is absent from conventional scholarly metadata. Large language models (LLMs) can convert this heterogeneous content into structured metadata, but open-schema extraction is difficult to evaluate without document-level gold annotations and consistent field names. To address this challenge, we applied an agreement-based evaluation methodology for assessing open-schema metadata extraction without requiring document-level gold metadata. The methodology is applied to metadata independently generated by six representative LLMs from 200 README files in the NERdME corpus using a zero-shot, evidence-grounded prompt. Output well-formedness, extraction cardinality, field-name diversity, and agreement over fields, values, and field–value pairs are evaluated under exact and embedding-based semantic matching. JSON validity ranges from 0.990 to 1.000 and record completeness is at least 0.998, although extraction cardinality varies substantially across models. Semantic matching raises multi-model field–value Jaccard similarity from 0.0000 to 0.1053 and Krippendorff’s alpha from 0.1784 to 0.5237. Agreement differences between open- and closed-model groups are limited, whereas within-family agreement consistently exceeds cross-family agreement. These results show that models often extract semantically similar metadata while organizing it differently, demonstrating that inter-model agreement is a useful diagnostic for evaluating open-schema metadata extraction, although it is not a substitute for correctness validation.

3.2. Discovery

- **ArticleMiner: Ontology-Guided Knowledge Graph Construction from Scientific Publications** (*Md Abrar Jahin, Craig A. Knoblock and Jay Pujara*) [12]

Abstract: Scientific papers keep much of their quantitative content in tables and supplementary files, where a number means something only through its header, caption, unit, analytical method, and the conventions of its field. Recovering the rows and columns of a table is therefore not the

same as recovering the scientific fact it reports. Most semantic table-interpretation methods assume that a clean table is already available and subsequently map its cells or columns to ontology terms, whereas most publication-level extraction systems are designed for a single domain. We study a middle path: a shared process that reads a paper and its supplementary files, gathers evidence from several parsers and a language model, and reconciles that evidence, while a bounded human-authored task module for each task supplies the domain meaning. The module lists the canonical names the graph may use, the surface forms that map to them, a small set of derivation rules and validity constraints, an identity key, and the bindings used to write RDF. It defines what a task is allowed to emit; it does not try to list every convention of a field. We build four such modules (for drug-discovery chemistry, materials science, machine learning, and mineral geochemistry) in the ArticleMiner framework, and evaluate them on 163 papers, including a new geochemistry benchmark with expert-curated ground truth. On matched comparisons in which a same-LLM few-shot baseline receives the same model and input, the estimates favor ArticleMiner on all four tasks; the paired differences are statistically significant on the two larger tasks (materials science and geochemistry) and remain uncertain on the two smaller ones.

- **Paper2ModelCard: A Framework for the Automated Generation of Model Cards from Research Papers** (*Mateo Campaya-Pinto, Elvira Amador-Domínguez, Erick Cedeño, Daniel Garijo and Sergio Zaballo-Herrera*) [13]

Abstract: Model Cards were introduced in 2019 as a means to provide a human-readable, clear summarization of machine learning models, including key metadata such as license, biases or evaluation results. While the adoption of these cards has grown by the scientific community (being adopted by open platforms like HuggingFace), model cards are usually manually filled by authors when uploading their models, thus leading to missing or inaccurate information. Moreover, model cards are not semantically interoperable, since they are not compliant with any ontology or data model. This paper presents Paper2ModelCard (P2MC), a framework for end-to-end generation of model cards from research papers. P2MC integrates processing tools for correctly parsing the input PDF into a structured format (using existing tools like SciPDF and LightOnOCR-1B), which is then processed by an ensemble of models, each targeted towards efficiently extracting specific information from the input. The FAIR4ML ontology is used as the backbone to generate the model cards, in order to ensure their interoperability. Moreover, P2MC links the generated model cards to existing scientific knowledge graphs, such as SemOpenAlex or LinkedPapersWithCode.

- **SciGraphChat: Retrieval-Augmented Generation over Scientific Knowledge Graphs** (*Angelika Kaplan, Jan Keim, Xiyu Zhong and Anne Koziolk*) [14]

Abstract: Effective retrieval and synthesis of scholarly knowledge remain challenging due to fragmented publication practices, LLM hallucinations, and the limitations of keyword-based search. We present SciGraphChat, an operational system deployed as an AIAssistant within the NFDIxCS consortium, enabling natural language exploration of scholarly knowledge graphs (KGs) through an integrated conversational and graph interface. The system combines two complementary, schema-agnostic and training-free retrieval methods for KG question answering, supporting source-aware and semantically informed access to structured scholarly data. By coupling KG-based Retrieval-Augmented Generation with a dialogue-oriented interface, SciGraphChat produces transparent, provenance-aware responses grounded in FAIR scholarly metadata, while a dedicated graph panel allows interactive exploration of the underlying KG structure. We evaluate the system with NFDIxCS stakeholders and external users via the System Usability Scale (SUS) and the User Experience Questionnaire–Short (UEQ-S), complemented by qualitative feedback. Results show good perceived usability (SUS mean 69.4), clearly positive pragmatic quality (UEQ-S mean 0.93), and neutral hedonic quality (UEQ-S mean 0.70). Qualitative feedback indicates that users value the combination of conversational access and graph-based

inspection, while also requesting tighter integration with scholarly workflows such as reference export, corpus scoping, and adaptive result presentation. This paper reports the system design and deployment-grounded evaluation results as evidence for the practical usability of Semantic Web technologies in research infrastructures.

- **Knowledge Graph Generation from Biodiversity Catalogs with LLM Annotation** (*Sefika Efeoglu and Jana Hoffmann*) [15]

Abstract: Type specimen (insect individuals) catalogs in the entomology domain contain a large amount of biodiversity data, including changes and losses in species numbers. Yet these catalogs lack machine readability because they are published as journal articles without annotations. Therefore, to analyze and trace how type specimens have changed in either collections or the domain over years, they must be represented in a structured format. This work constructs a domain-specific schema from competency questions (CQs) of four user groups in the entomology domain and its library while leveraging Darwin Core terms where needed. Then, collaborative large language model annotation is applied to the type specimen catalogs using the domain-specific schema to construct a knowledge graph (KG). Finally, we evaluate this KG using SPARQL queries for these CQs.

- **Local Citation Recommendation via Heterogeneous Graph Neural Networks and Contrastive Learning** (*Bahaeddine Hantous, Bruno Yun and Nadia Yacoubi-Ayadi*) [16]

Abstract: The exponential growth of scientific literature makes it extremely difficult for researchers to identify the most relevant prior work to cite at a given point in a manuscript. Local citation recommendation (LCR) addresses this problem by suggesting, for a given citation context, the papers that best support the underlying claim. Existing approaches to LCR are typically confined to single-stage retrieval, prefetch-then-rerank pipelines, or generative formulations that only weakly exploit the structural relationships between papers, authors, and citation contexts. In this paper, we propose a Scholarly Knowledge Graph (SKG)-based approach to LCR that jointly exploits the semantic structure of scholarly data and neural representation learning. We first construct ArgKG, a domain-specific SKG for the argumentation-theory literature, comprising 1,576 collected corpus papers, 24,438 additional referenced papers (26,014 in total), and 82,868 citation contexts, modelled as an RDF graph that reuses established scholarly ontologies and vocabularies to represent papers, authors, citation links, and citation-context semantics explicitly. Building on this representation, we introduce a two-tower architecture that learns a shared embedding space for citation contexts and candidate papers through contrastive learning. Candidate papers are represented by metapath-based features extracted from the SKG – their title and abstract, reference neighbourhood, and previously observed citation contexts – fused using a transformer-based encoder. We report experimental results across several feature-set and architecture configurations on ArgKG and on the established ACL-200 benchmark, showing that combining a paper’s own title and abstract with its citation neighbourhood yields the best-performing configuration and outperforms existing LCR baselines on ACL-200. To support reproducibility, the code, trained models, and SKG construction pipeline are publicly released at <https://github.com/Ba-aH/SciteKG>.

- **The CIDOC CRM Periodic Table: Documentation You Can Test to Support the Adoption of Scientific Research Ontologies for the Humanities** (*Remo Grillo*) [17]

Abstract: The CIDOC CRM family is not only a schema for museum collections. Its extensions formalise the apparatus through which research knowledge is produced and justified: CRMsci models scientific observation and measurement, CRMdig the provenance of digital processes, CRMinf inference and belief. Together they constitute a schema for research data, and the obstacle to their adoption is documentary rather than theoretical: four normative documents, four code conventions, and eighty-nine typed links crossing document boundaries. This paper presents the CIDOC CRM Periodic Table, a single-page browser rendering 131 classes and 480

property entries from CIDOC CRM 7.1.3, CRMsci 3.2, CRMdig 5.0 and CRMinf 1.2.1 as one navigable surface, with two inference facilities: a one-step triple builder admitting only the class–property combinations licensed by declared domains, ranges and the subclass closure, and a multi-step path finder enumerating property chains between two classes. Because all loaded models are merged into one code-keyed dictionary before the hierarchy closures are computed, both resolve chains that cross model boundaries; we document the addition of CRMsci and the three cross-model tests it made possible, and report preliminary uptake signals from 56 months of analytics, with the reading they support.

3.3. Assessment

- **From Data Footprints to Researcher Fingerprints: Temporal Ego-centric Research Group Detection in Scientific Knowledge Graphs** (*Simone Angioni, Miriam Baglioni, Andrea Mannocci and Paolo Manghi*) [18]

Abstract: Understanding who researchers collaborate with, when and how their research groups form, and how these evolve over time is a foundational prerequisite for constructing multidimensional, evidence-based researcher fingerprints – a requirement that has become increasingly urgent in the context of ongoing reforms in research evaluation advocated by initiatives such as DORA, the Leiden Manifesto, and CoARA. In this paper, we propose an egocentric approach for fine-grained research group detection, operating at the scale of individual researchers’ ego networks, as extracted from Scientific Knowledge Graphs (SKGs), and requiring the ego node to be present in all identified research groups – a constraint that is not merely technical but reflects a commitment to researcher-centred analysis. Our approach builds on the Leiden algorithm as its computational core and is grounded in SKGs, which provide the heterogeneous, open scholarly metadata necessary to trace collaboration patterns at scale in a reproducible and transparent manner. We evaluate the proposed method through a user evaluation involving 10 researchers with diverse backgrounds and career stages, demonstrating its ability to surface meaningful and interpretable collaboration structures. The research groups identified constitute the structural scaffold upon which richer characterisations of scholarly identity – including roles, seniority, contribution types, and publication patterns – can be built in subsequent work, contributing to the broader goal of fairer and more transparent research assessment.

- **ReCitE: Evaluating Citation Context Quality and Relevance with Large Language Models** (*Ahmar Kamal Hussain, Bernhard Sabel and Andreas Nürnberger*) [19]

Abstract: The growing number of fraudulent publications poses a serious threat to scientific integrity; however, a substantial number of these publications remain undetected in the published literature. To help address this problem, we developed ReCitE (Reliable Citation Evaluation): a framework to identify fraudulent papers by analyzing citation relevance, that is, whether a reference actually supports the claim to which it is attached. The framework includes parsing a PDF, extracting references, finding the referenced papers in open bibliographic databases, then prompting a large language model to identify the claims tied to each citation, score their relevance, and provide a brief justification for each assessment. We manually evaluate the quality of citation context extraction and evaluate citation relevance with five open-source, locally deployable LLMs on a public dataset of 22,381 labeled citations. The best-performing model for citation relevance was Gemma 2-27b, with an F1 score of 0.981 and an ROC AUC of 0.981, comparable to the state-of-the-art results based on SBERT embeddings. An error analysis shows that misclassifications occur due to generic citation contexts, uninformative abstracts and citations used to establish background rather than support a direct claim. This method is not intended to replace human reviewers, but rather to serve as an automated tool that helps reviewers, editors and integrity sleuths flag papers for closer inspection.

- **Retraction without Traction: Towards Handling Post-publication Invalidation in Open Science Infrastructure** (*Christoph Stade and Dexter Williams*) [20]

Abstract: The process of retraction can invalidate the results and claims of a scientific publication. Therefore, to maintain the integrity of past and future research, it is critical for human researchers and research agents to know when a work has been retracted. Yet, scholarly knowledge graphs like OpenAIRE fail to consistently communicate retraction status. We argue for the inclusion of an explicit retraction indicator in OpenAIRE's schema by extending it with scholarly ontologies that represent retraction information; this retraction information is already available through a trusted aggregation source: Crossref. Further, we borrow the concept of forgetting from early knowledge representation and cognitive psychology literature to conceptualize retraction within scientific knowledge graphs. Finally, we argue that retracted publications propagate uncertainty through a citation network and propose that the PageTrust algorithm can be used to compute this effect.

4. Keynote

For this 6th edition, we are honoured to have Dr. Silvio Peroni, Associate Professor at the Department of Classical Philology and Italian Studies, University of Bologna. Director of OpenCitations³ as our keynote speaker. Here is how Silvio introduces himself.

I hold a Ph.D. degree in Computer Science, and I am an Associate Professor at the Department of Classical Philology and Italian Studies, University of Bologna, where I teach 'Basic Informatics', 'Introduction to Programming', 'Computational Management of Data' and 'Open Science'. I am an expert in document markup and semantic descriptions of bibliographic entities using OWL ontologies. I am Director of the Research Centre for Open Scholarly Metadata, one of the main developers of the SPAR (Semantic Publishing and Referencing) Ontologies, Director of OpenCitations, and founding member of the Initiative for Open Citations (I4OC) and the Initiative for Open Abstract (I4OA). I have been Principal Investigator of the Spoke 4 of the Project CHANGES (Cultural Heritage Active Innovation For Next-Gen Sustainable Society) and of the H2020 project SPICE – Social cohesion, Participation, and Inclusion through Cultural Engagement. Also, I have participated in several EU projects as Local Unit (UNIBO) Coordinator, including LUMEN, GRAPHIA, GraspOS, and OpenAIRE-Nexus. I am an Editorial Board member of Journal of Documentation (Emerald) and PeerJ Computer Science (PeerJ), among others, and I have been involved in the organisation and programme committee of several Semantic Web venues, including ISWC and ESWC. Among my research interests are Semantic Web technologies, markup languages for complex documents, design patterns for digital documents and ontology modelling, and automatic processes of analysis and segmentation of documents. My works concern theoretical studies and technical implementation of tools to foster semantic interoperability of Open Science services and infrastructures, the empirical analysis of the nature of scholarly citations, bibliometrics and scientometrics studies, visualisation and browsing interfaces for semantic data, and the development of ontologies to manage, integrate and query bibliographic information and cultural heritage data. Recently, I have become the UNIBO representative for the Barcelona Declaration on Open Research Information (DORI) and a member of several international working groups, including the CoARA Working Group Towards Open Infrastructures for Responsible Research Assessment (OI4RRA) and the RDA Working Group Scientific Knowledge Graphs Interoperability Framework (SKG-IF).

³OpenCitations – <https://opencitations.net/>

Title: There and Back Again: A Journey about Interoperability, Open Scholarly Infrastructures, and Scientific Knowledge Graph

Abstract: While it is difficult to formally define what they are precisely, Open Scholarly Infrastructures (OSIs) are a fundamental building block for the implementation of the Open Science goals as introduced by UNESCO, since they play roles in reforms, address critical gaps, and, more generally, enable more inclusive, equitable, and transparent science in all disciplines and academic communities. Even if the OSI community broadly agrees that the sum of their endeavours could surpass the outreach of proprietary systems, presenting such OSIs as a unique decentralised commodity requires first addressing one of the most crucial and delicate problems in information systems: interoperability. In this talk, I will introduce recent efforts from the past few years to provide standards and tools that support the implementation of interoperability, particularly at the technical and semantic levels, with a focus on the OSIs maintaining Scientific Knowledge Graphs (SKGs), which is one of the main topics of the Sci-K workshop series.

5. Partnership and OpenAIRE Graph Award



Sci-K 2026 is proud to partner with OpenAIRE to offer a €300 prize for the best paper making significant use of the OpenAIRE Graph – one of the largest open scholarly knowledge graphs available, covering millions of publications, datasets, software, projects, and organisations. Whether you use it as your sole data source or alongside others, if the OpenAIRE Graph plays a central role in your methodology or findings, your paper is automatically in the running. The winner will be announced on the workshop day.

6. Program Committee

We would like to express our heartfelt gratitude to the members of the programme committee, who, in the middle of August, dedicated time to reviewing papers.

- Bradley Allen (University of Amsterdam)
- Simone Angioni (ISTI CNR)
- Arshad Beg (Maynooth University, Ireland.)
- Grégoire Burel (The Open University)
- Davide Buscaldi (LIPN, Université Paris 13, Sorbonne Paris Cité)
- Ezgi Cavus (Middle East Technical University)
- Serafeim Chatzopoulos (Athena Research Center)
- Auriol Degbelo (Chair of Geoinformatics, TU Dresden)
- Giorgio Maria Di Nunzio (University of Padua)
- Hang Dong (University of Exeter)
- Yuanxi Fu (University of Illinois at Urbana-Champaign)
- Sasi Kiran Gaddipati (TIB – Leibniz Information Centre for Science and Technology)
- Daniel Garijo (Universidad Politécnica de Madrid)
- Genet Asefa Gesese (FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur)

- Esteban González Guardia (Ontology Engineering Group – Universidad Politecnica de Madrid)
- Md Abrar Jahin (USC Information Sciences Institute)
- Mohamad Yaser Jaradeh (L3S Research Center, Leibniz University Hannover)
- Alireza Javadian Sabet (University of Pittsburgh)
- Lena John (TIB – Leibniz Information Centre for Science and Technology)
- Angelika Kaplan (Karlsruhe Institute of Technology)
- Jan Keim (Karlsruhe Institute of Technology (KIT))
- Mutahira Khalid (TIB – Leibniz Information Centre for Science and Technology)
- Soo-Yon Kim (RWTH Aachen University)
- Craig Knoblock (USC Information Sciences Institute)
- Giacomo Magnifico (Tartu University)
- Tomasz Miksa (SBA Research)
- Diarmuid O’Donoghue (Maynooth University)
- Nipun Pathirage (Rensselaer Polytechnic Institute)
- Jay Pujara (University of Southern California)
- Enayat Rajabi (Dalhousie University)
- Rohitha Ravinder (ZB MED – Information Center for Life Sciences)
- Diego Reforgiato Recupero (Università degli studi di Cagliari)
- Tomer Sagi (Aalborg Universitet)
- Giuseppe Scarpi (University of Cagliari)
- Jodi Schneider (University of Illinois Urbana Champaign)
- Mako Sorensen (Carleton University)
- Christoph Stade (University of Wisconsin-Madison)
- Sanju Tiwari (Sharda University, Delhi NCR, India)
- Nadia Yacoubi-Ayadi (Université Claude Bernard (Lyon 1))
- Dilek Yargan (FIZ Karlsruhe – Leibniz Institute for Information Infrastructure)
- Giacomo Zamprogno (Vrije Universiteit Amsterdam)

Declaration on Generative AI

During the preparation of this preface, the authors used Gemini in order to: Grammar and spell checking.

References

- [1] P. Manghi, A. Mannocci, F. Osborne, D. Sacharidis, A. Salatino, T. Vergoulis, Sci-k 2022-international workshop on scientific knowledge: Representation, discovery, and assessment, in: Companion Proceedings of the Web Conference 2022, 2022, pp. 735–738.
- [2] A. A. Salatino, Y. Bu, Y. Ding, Á. Horvát, Y. Huang, M. Liu, P. Manghi, A. Mannocci, F. Osborne, D. Romero, et al., 3rd international workshop on scientific knowledge representation, discovery, and assessment (sci-k 2023), in: Companion Proceedings of the ACM Web Conference 2023, 2023, pp. 783–783.
- [3] A. Salatino, A. Mannocci, F. Osborne, G. Rehm, S. Schimmler, 4th international workshop on scientific knowledge: Representation, discovery, and assessment, volume Vol-3780, CEUR, 2024. URL: <https://ceur-ws.org/Vol-3780/preface.pdf>.
- [4] A. M. Jacyszyn, A. Mannocci, F. Osborne, G. Rehm, A. Salatino, S. Schimmler, L. Stork, 5th international workshop on scientific knowledge: Representation, discovery, and assessment, volume Vol-4065, CEUR, 2025. URL: <https://ceur-ws.org/Vol-4065/preface.pdf>.

- [5] L. Bellatreche, F. Bentayeb, M. Bielíková, O. Boussaid, B. Catania, P. Ceravolo, E. Demidova, M. Halfeld Ferrari, M. T. G. Lopez, C. S. Hara, et al., Databases and information systems in the ai era: Contributions from adbis, tpd and eda 2020 workshops and doctoral consortium, in: ADBIS, TPD and EDA 2020 Common Workshops and Doctoral Consortium: International Workshops: DOING, MADEISD, SKG, BBIGAP, SIMPDA, AIMinScience 2020 and Doctoral Consortium, Lyon, France, August 25–27, 2020, Proceedings 24, Springer, 2020, pp. 3–20.
- [6] P. Manghi, A. Mannocci, F. Osborne, D. Sacharidis, A. Salatino, T. Vergoulis, New trends in scientific knowledge graphs and research impact assessment, *Quantitative Science Studies* 2 (2021) 1296–1300.
- [7] V. Pasqual, A. Nazari, S. Peroni, Supporting the understanding of ontologies for scientific knowledge graphs with the new version of lode, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper1.pdf>.
- [8] S. Angioni, A. Oelen, A. Salatino, Research contributions: Towards a formal and semantic representation, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper2.pdf>.
- [9] N. Pathirage, H. Santos, D. L. McGuinness, Sudo: From scientific publications to global scholarly discourse, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper3.pdf>.
- [10] R. Hmaidan, A. Puertas Gallardo, S. Consoli, A. Zani, M. Ceresa, D. Reforgiato, Representing evidence of emotional introspection in large language models: A scientific knowledge graph research agenda, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper4.pdf>.
- [11] G. A. Gesese, Z. Chen, S. Schimmler, H. Sack, Open-schema metadata extraction from github readmes: A cross-model agreement study, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper5.pdf>.
- [12] M. A. Jahin, C. A. Knoblock, J. Pujara, Articleminer: Ontology-guided knowledge graph construction from scientific publications, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper6.pdf>.
- [13] M. Campaya-Pinto, E. Amador-Domínguez, E. Cedeño, D. Garijo, S. Zaballos-Herrera, Paper2modelcard: A framework for the automated generation of model cards from research papers, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper7.pdf>.
- [14] A. Kaplan, J. Keim, X. Zhong, A. Koziol, Scigraphchat: Retrieval-augmented generation over scientific knowledge graphs, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper8.pdf>.
- [15] S. Efeoglu, J. Hoffmann, Knowledge graph generation from biodiversity catalogs with llm annotation, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper9.pdf>.
- [16] B. Hantous, B. Yun, N. Yacoubi-Ayadi, Local citation recommendation via heterogeneous graph neural networks and contrastive learning, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper10.pdf>.
- [17] R. Grillo, The cidoc crm periodic table: Documentation you can test to support the adoption of scientific research ontologies for the humanities, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper11.pdf>.
- [18] S. Angioni, M. Baglioni, A. Mannocci, P. Manghi, From data footprints to researcher fingerprints: Temporal ego-centric research group detection in scientific knowledge graphs, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper12.pdf>.
- [19] A. K. Hussain, B. Sabel, A. Nürnberger, Recite: Evaluating citation context quality and relevance with large language models, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper13.pdf>.
- [20] C. Stade, D. Williams, Retraction without traction: Towards handling post-publication invalidation in open science infrastructure, *CEUR Workshop Proceedings Vol-XXXX* (2026). URL: <https://ceur-ws.org/Vol-XXXX/paper14.pdf>.