

Semantification of CEUR-WS with Wikidata as a target Knowledge Graph

Wolfgang Fahl, Tim Holzheim, Christoph Lange and Stefan Decker

TEXT2KG@ESWC 2023, Hersonissos, Greece, 2023-05-29

Semantification of CEUR-WS

Agenda

- CEUR-WS.org
- Core Entities
- FAIRness of CEUR-WS to be improved
- Workflow of the CEUR-WS Semantification
- Phases of the Semantification
- Semantification Challenges / Semantic Publishing Challenge
- Implementation and Demos
- Future Work / TODOs
- Tools used

See also <https://cr.bitplan.com/index.php/SoC>

- **Free open-access publication service for academic workshops**
 - Since 1995
 - >3,400 volumes
 - >4,000 editors
 - >50,000 papers
- **semi-structured data about**
 - proceedings
 - events
 - papers
 - editors
 - authors



ceur-ws.org
ISSN 1613-0073

Vol-3262
urn:nbn:de:0074-3262-0

Copyright © 2022 for the individual papers by the papers' authors. Copyright © 2022 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).

Wikidata 2022 Wikidata Workshop 2022

Proceedings of the 3rd Wikidata Workshop 2022
co-located with the 21st International Semantic Web Conference (ISWC2022)

Virtual Event, Hangzhou, China, October 2022.

Edited by
Lucie-Aimée Kaffee*
Simon Razniewski**
Gabriel Amaral***
Kholoud Saad Alghamdi***

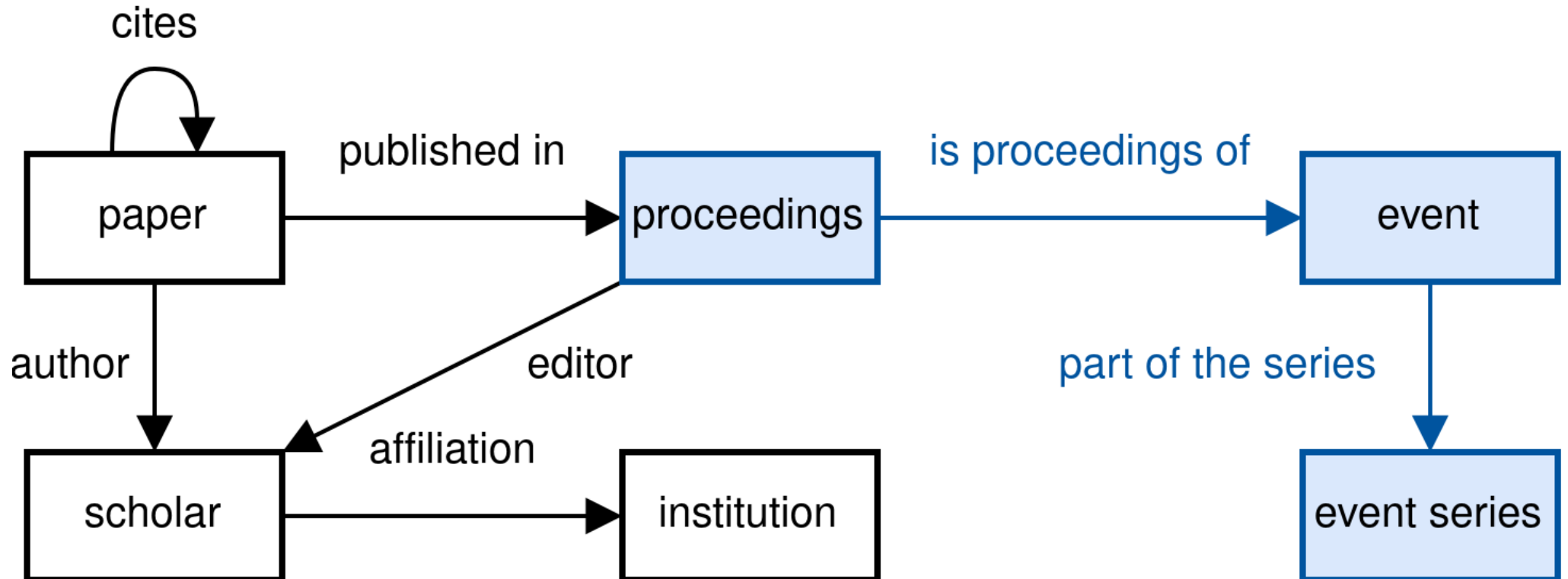
* University of Copenhagen, Denmark
** Max Planck Institute for Informatics, Germany
*** King's College London, UK



Co-author graph of people involved in the event



Core Entities



Clickable version with links to Wikidata at <https://cr.bitplan.com>

FAIRness of CEUR-WS to be improved

FAIR Principles

Findable:

URL  URN  DOI 
DBLP  National Library 
Wikidata 

Accessible:

HTML  JSON 

Interoperable:

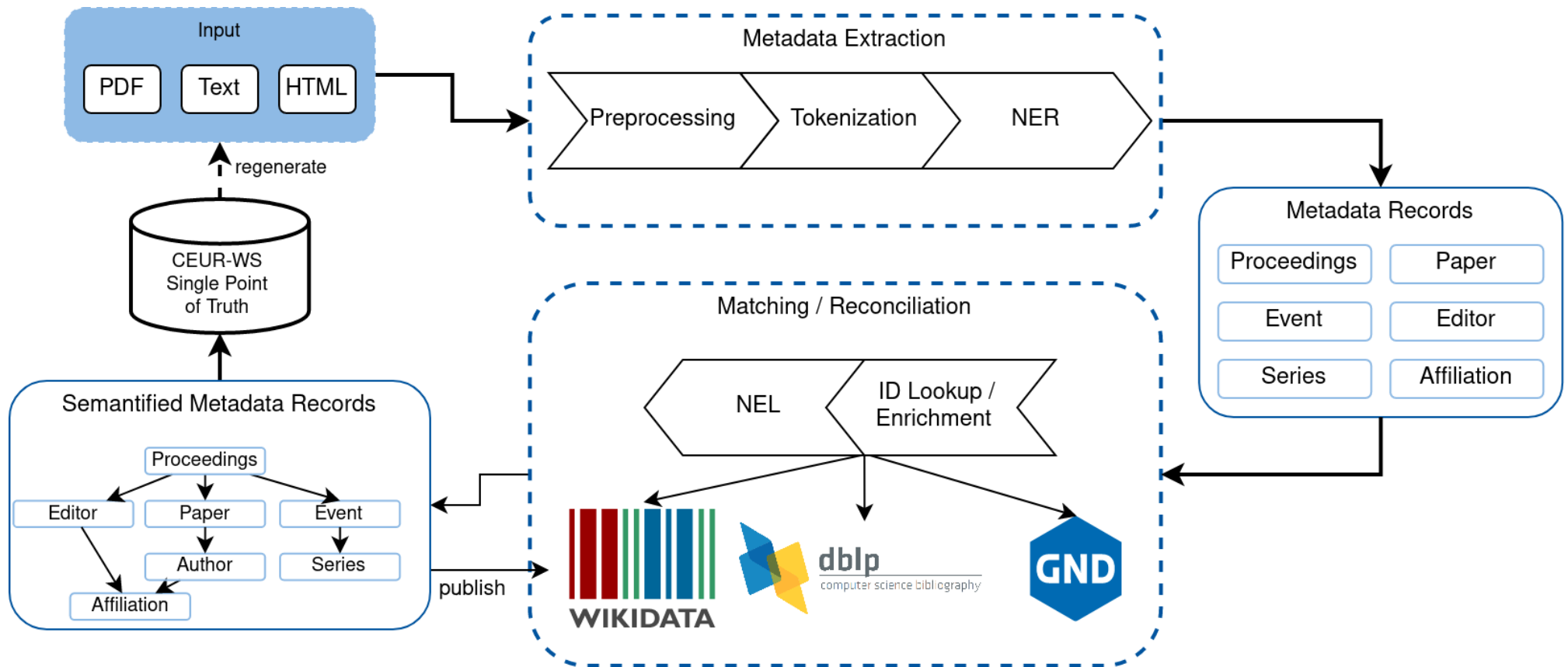
HTML  SPARQL 

Reusable:

HTML  OpenAPI 

Legend:  available  incomplete  in progress/focus of this work

Workflow of the CEUR-WS Semantification



Phases of the Semantification

Numbers as of 2023-05-26

Phase 1:

created/matched:

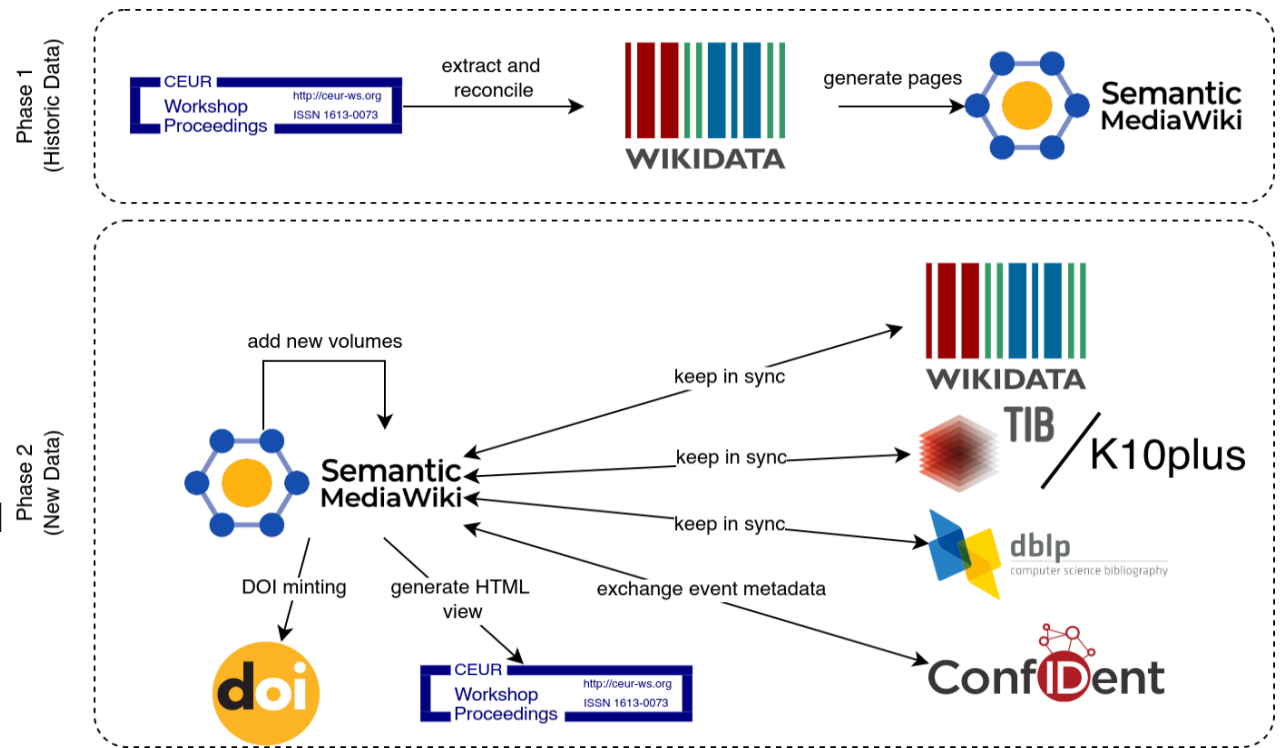
- 3,381 proceedings items
- 3,501 event items

next:

- editor
- paper
- author

Phase 2

- direct semantification of new volumes
- DOI minting for event, proceeding and papers
- API for proper access
- synchronization with wikidata



Semantification Challenges / Semantic Publishing Challenge

Example Challenges

- 2014/2015 Challenge by Christoph Lange et al. [List of Queries](#)
 - NER with HTML parsers for index and volumes
 - NEL and Disambiguation via identifier match
- Wikidata
 - property P11633: [colocated with](#)
 - CEUR-WS bot approval
- Red tape
 - The Devil in the [details](#)
 - GDPR ([DSVGO Article 6](#)) conformance
 - Change of publishing Workflow ([HOWTOSUBMIT/PUT](#))
 - DOI Minting

Implementation and Demos

CEUR-WS Volume Browser

The screenshot shows the CEUR-WS Volume Browser interface. At the top, there is a navigation bar with icons for home, search, settings, and other functions. Below the navigation bar, the text "CEUR-WS Volume Browser" is displayed. A search bar contains the text "Volume 3184". Below the search bar, there are two buttons: a blue button with a refresh icon and a light blue button with a globe icon. Below these buttons, there are links for "wikidataitem", "dblp", "k10plus", "scholia", and "event series". The main title of the volume is "TEXT2KG 2022 & MK 2022 First International Workshop on Knowledge Graph Generation From Text and First International Workshop on Modular Knowledge". Below the title, there is a link for "TEXT2KG 2022" and a note: "Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text and the 1st International Workshop on Modular Knowledge". Below this, there is a note: "? published: 2022-08-11 00:00:00 submitted By: Sanju Tiwari".

CEUR
Workshop
Proceedings

ceur-ws.org
ISSN 1613-0073

Copyright © 2022 for the individual papers by the papers' authors. Copyright © 2022 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).

Vol-3184
urn:nbn:de:0074-3184-1

TEXT2KG 2022 & MK 2022
First International Workshop on Knowledge Graph Generation From Text and First International Workshop on Modular Knowledge

Implementation and Demos

CEUR-WS Semantic MediaWiki

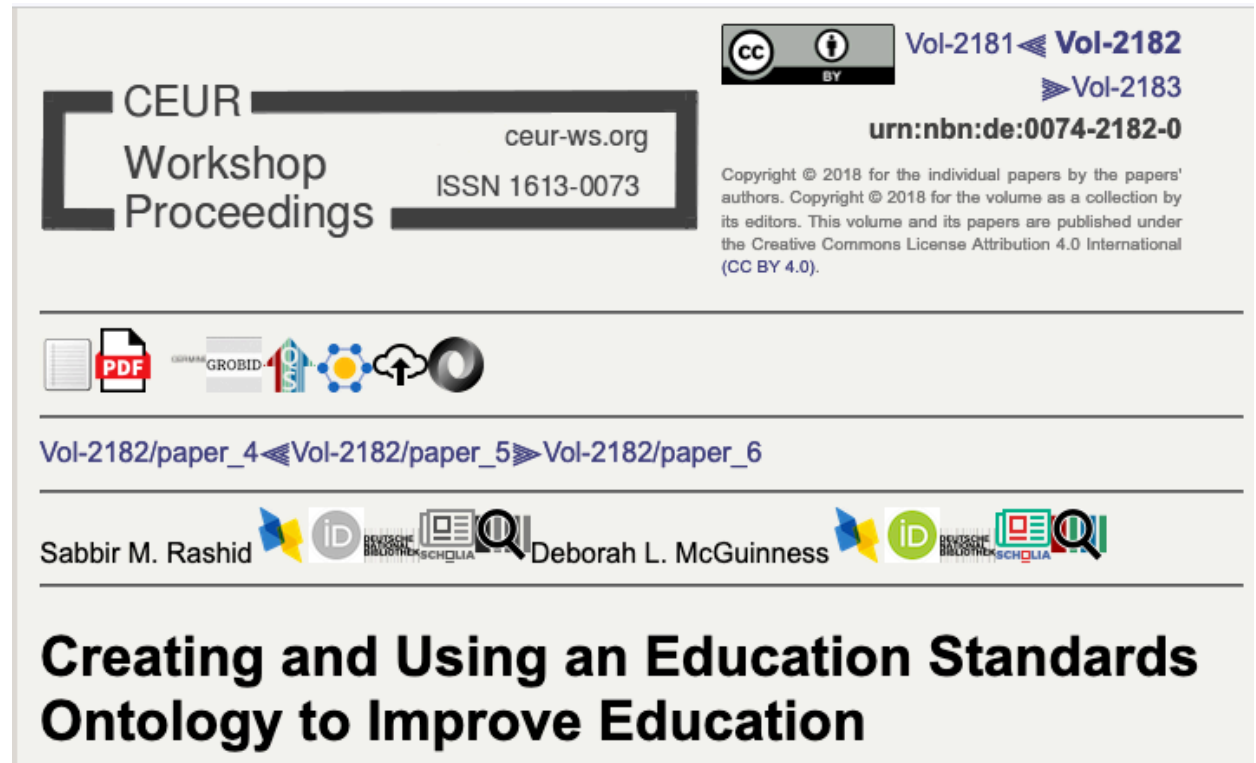
The screenshot shows the CEUR-WS Semantic MediaWiki interface. On the left is a sidebar with navigation links: Main page, CEUR-WS, Volumes, Sessions, Papers, Scholars, Institutions, Events, EventsSeries, Wiki, Search, HowTo, Random page, Help about MediaWiki, Browse categories, Browse files, Recent changes, Tools, What links here, Related changes, and Special pages. The main content area is titled "List of Papers" and includes a "Discussion" tab, "View source", "View history", and "More" options, along with a search bar for "Search BITPlan ceur-ws Wiki". Below the title is a table with columns: icon, Topic, name, pluralName, and documentation. The first row shows a paper icon, "Paper", "Paper", "Papers", and "A paper is e.g. a scholarly article". Below this is a "Papers" section with a "10" count and an "add Paper" button. The main table has columns: Paper, description, id, wikidataid, title, pdfUrl, and authors. It lists several papers with their respective details.

Paper	description	id	wikidataid	title	pdfUrl	authors
Vol-1878		Vol-1878/article-05.pdf		Scholarly Social Machines	http://ceur-ws.org/Vol-1878/article-05.pdf	David De Roure
Vol-2535/paper10		Vol-2535/paper10	Q117032134	Towards a Knowledge Graph Lifecycle: A pipeline for the population of a commercial Knowledge Graph	https://ceur-ws.org/Vol-2535/paper_10.pdf	Jürgen Umbrich Dieter Fensel Umutcan Şimşek
Vol-2599/paper5		Vol-2599/paper5		Private Digital Identity on Blockchain		
Vol-2644		Vol-2644/paper35		Using PROVA-Rule Engine as Dispatching-Service for FHIR-Observation-Resources	http://ceur-ws.org/Vol-2644/paper35.pdf	Gerhard Kober, Adrian Paschke
Vol-2644		Vol-2644/paper36		Action Rules: Counterfactual Explanations in Python (winner of the 14th Rule Challenge 2020 competition)	http://ceur-ws.org/Vol-2644/paper36.pdf	Lukas Sykora, Tomas Kliegr

Implementation and Demos

Single Point of Truth Server

- OpenAPI:
<http://ceurspt.wikidata.dbis.rwth-aachen.de/docs>
- Generated HTML
- Content Negotiation – Computer and Human Readable
- Federated Queries



CEUR Workshop Proceedings
ceur-ws.org
ISSN 1613-0073

Vol-2181 ◀ Vol-2182
▶ Vol-2183
urn:nbn:de:0074-2182-0

Copyright © 2018 for the individual papers by the papers' authors. Copyright © 2018 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).

PDF GROBID-1

Vol-2182/paper_4 ◀ Vol-2182/paper_5 ▶ Vol-2182/paper_6

Sabbir M. Rashid Deborah L. McGuinness

Creating and Using an Education Standards Ontology to Improve Education

Future Work / TODOs

What needs to be done

- Automated Editor/Author/Paper addition
- Plausibility Checks
- Use and monitoring of Wikidata in production
- Use of Single Point of Truth metadata in production
- Modified Workflow
- DOI minting in cooperation with RWTH Aachen Library

Tools Used

Open source python libraries

- [BeautifulSoup4](#)
- [Geograpy3](#)
- [fastapi, uvicorn, linkml-runtime, tqdm, orjson](#)
- [justpy, lxml, pylodstorage, pyparsing, pyJustpyWidgets, pyGenericSpreadSheet, py-3rdparty-mediawiki, xmltodict](#)

Thank you for your attention!