

# 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

Workshop

Proceedings

ceur-ws.org

ISSN 1613-0073

CC BY

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

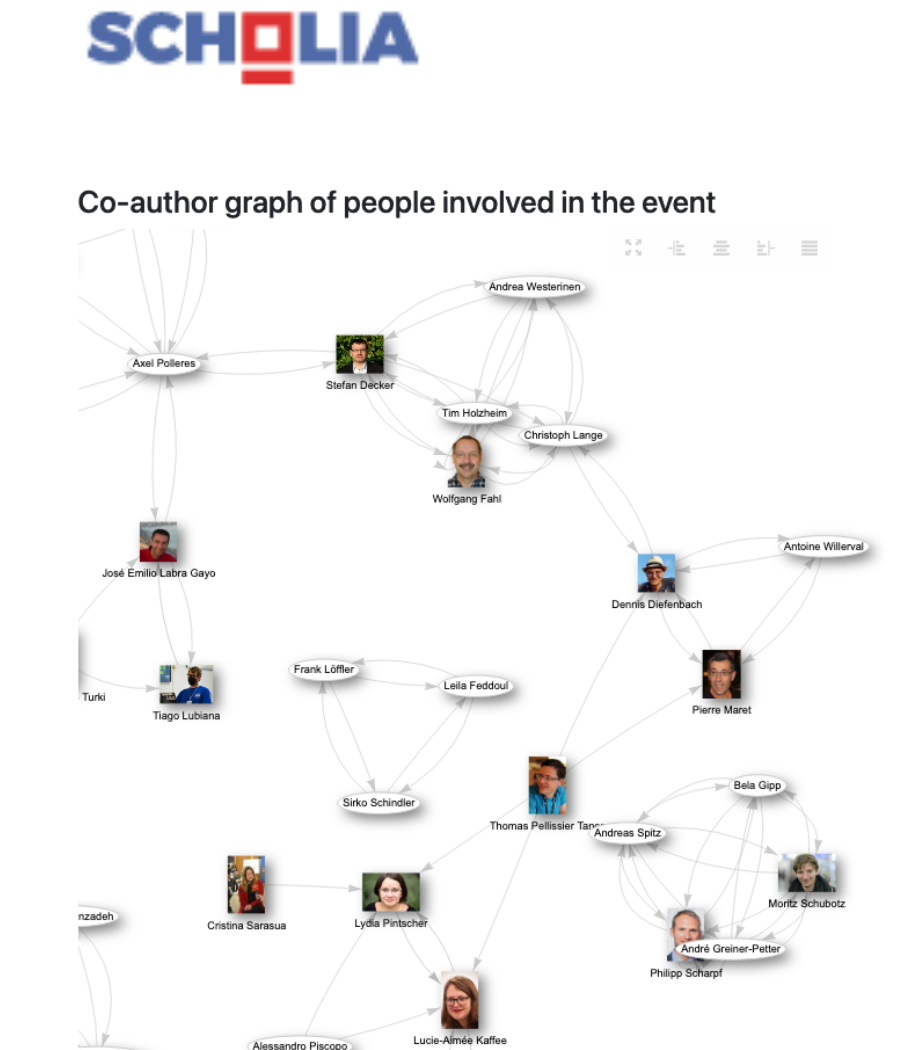
Proceedings of the 3rd Wikidata Workshop 2022  
co-located with the 21st International Semantic Web Conference (ISWC2022)

Virtual Event, Hangzhou, China, October 2022.

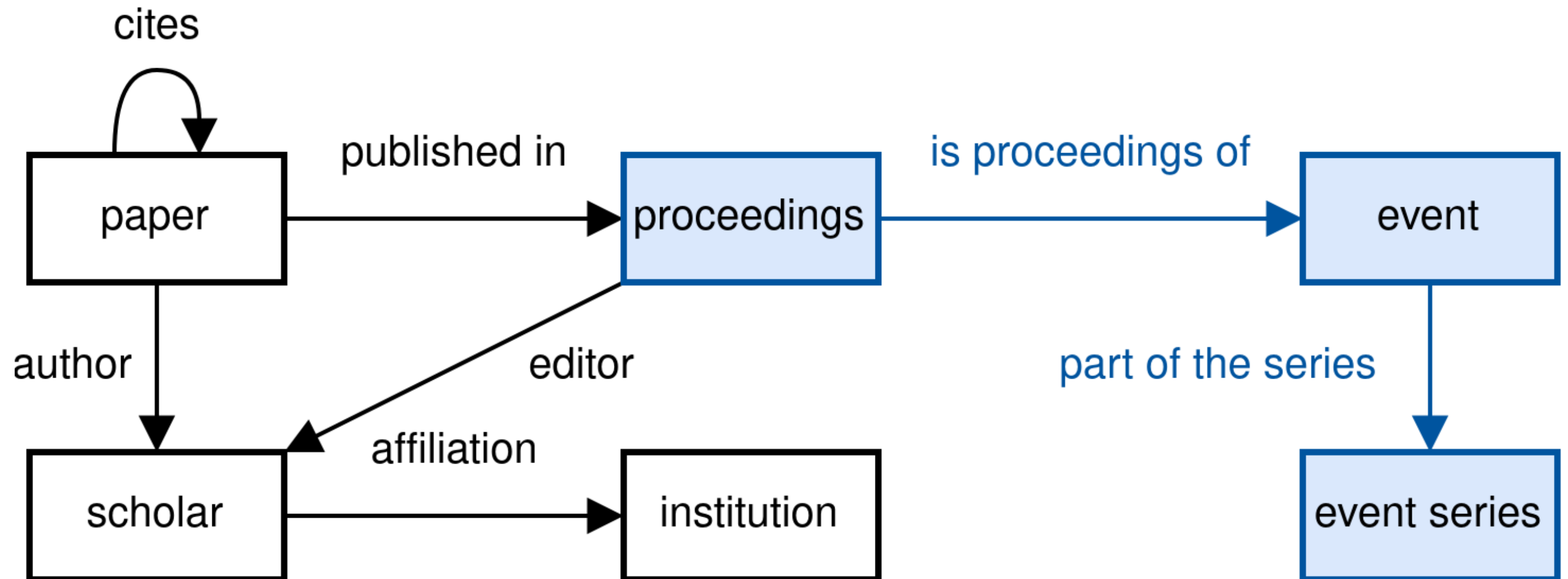
Edited by

**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



## 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 

#### Interoperable:

HTML  SPARQL 

#### Accessible:

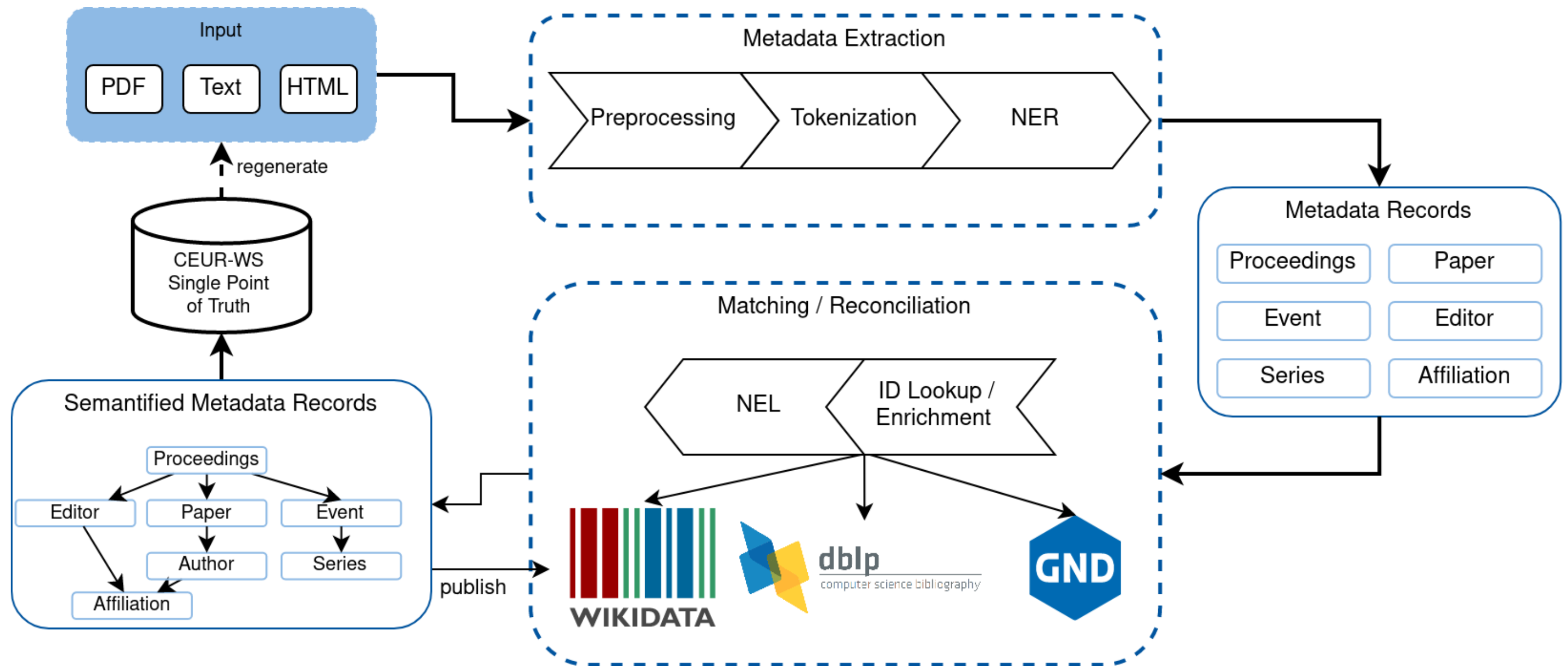
HTML  JSON 

#### 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:

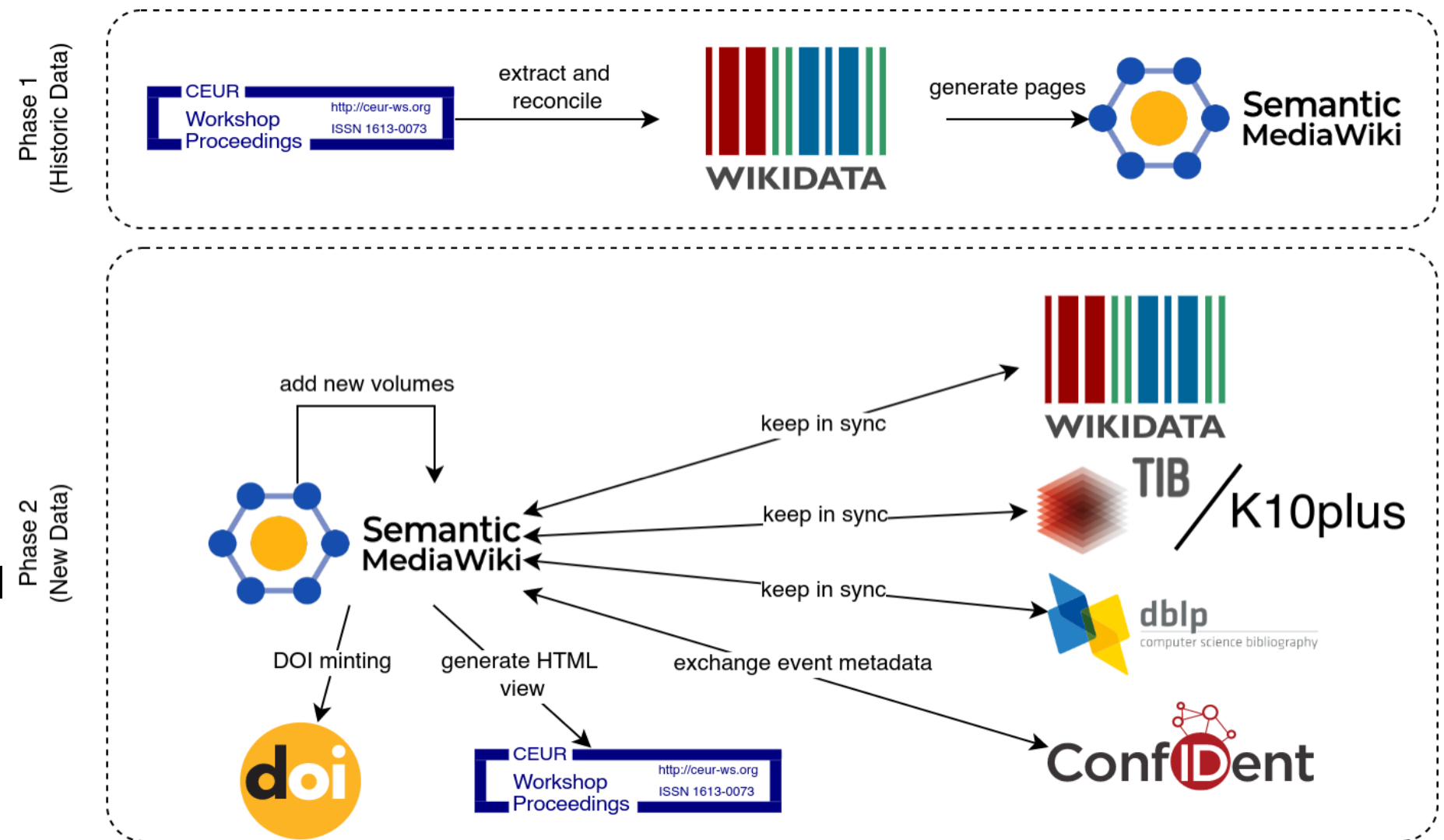
- 3,381 proceedings items
- 3,501 event items

#### 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



### 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




## CEUR-WS Volume Browser

9 Semantification of CEUR-WS with Wikidata as a target Knowledge Graph | Wolfgang Fahl | i5 | 2023-05-29

# Implementation and Demos

## CEUR-WS Semantic MediaWiki



CEUR Workshop Proceedings

Main page

CEUR-WS

Volumes

Sessions

Papers

Scholars

Institutions

Events

Events

EventSeries

Wiki

Search

HowTo

Random page

Help about MediaWiki

Browse categories

Browse files

Recent changes

Tools

What links here


Related changes

Special pages

Page **Discussion**

Read [View source](#) [View history](#) [More](#)

### List of Papers

| icon  | Topic | name  | pluralName | documentation                       |
|---|-------|-------|------------|-------------------------------------|
|  | Paper | Paper | Papers     | A paper is e.g. a scholarly article |

### Papers

10


[add Paper](#)

| Paper            | description | id                      | wikidataid                 | title  | pdfUrl  | authors   |
|------------------|-------------|-------------------------|----------------------------|--|---|---|
| Vol-1878         |             | Vol-1878/article-05.pdf |                            | Scholarly Social Machines  | <a href="http://ceur-ws.org/Vol-1878/article-05.pdf">http://ceur-ws.org/Vol-1878/article-05.pdf</a> | David De Roure                                    |
| Vol-2535/paper10 |             | Vol-2535/paper10        | <a href="#">Q117032134</a> | Towards a Knowledge Graph Lifecycle: A pipeline for the population of a commercial Knowledge Graph       | <a href="https://ceur-ws.org/Vol-2535/paper%2010.pdf">https://ceur-ws.org/Vol-2535/paper 10.pdf</a> | Jürgen Umbrich<br>Dieter Fensel<br>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                            | <a href="http://ceur-ws.org/Vol-2644/paper35.pdf">http://ceur-ws.org/Vol-2644/paper35.pdf</a>       | Gerhard Kober, Adrian Paschke                     |
| Vol-2644         |             | Vol-2644/paper36        |                            | Action Rules: Counterfactual Explanations in Python (winner of the 14th Rule Challenge 2020 competition) | <a href="http://ceur-ws.org/Vol-2644/paper36.pdf">http://ceur-ws.org/Vol-2644/paper36.pdf</a>       | 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).



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!