SEO: A Scientific Events Data Model

Said Fathalla[1,3(✉)], Sahar Vahdati[1,6(✉)], Christoph Lange[2], and Sören Auer[4,5]

[1] Smart Data Analytics (SDA), University of Bonn, Bonn, Germany
{fathalla,vahdati,langec}@cs.uni-bonn.de
[2] Fraunhofer FIT, Sankt Augustin, Germany
[3] Faculty of Science, University of Alexandria, Alexandria, Egypt
[4] Computer Science, Leibniz University of Hannover, Hanover, Germany
[5] TIB Leibniz Information Center for Science and Technology, Hannover, Germany
soeren.auer@tib.eu
[6] Department of Computer Science, University of Oxford, Oxford, UK

Abstract. Scientific events have become a key factor of scholarly communication for many scientific domains. They are considered as the focal point for establishing scientific relations between scholarly objects such as people (e.g., chairs and participants), places (e.g., location), actions (e.g., roles of participants), and artifacts (e.g., proceedings) in the scholarly communication domain. Metadata of scientific events have been made available in unstructured or semi-structured formats, which hides the interconnected and complex relationships between them and prevents transparency. To facilitate the management of such metadata, the representation of event-related information in an interoperable form requires a uniform conceptual modeling. The Scientific Events Ontology (OR-SEO) has been engineered to represent metadata of scientific events. We describe a systematic redesign of the information model that is used as a schema for the event pages of the *OpenResearch.org* community wiki, reusing well-known vocabularies to make OR-SEO interoperable in different contexts. OR-SEO is now in use on thousands of *OpenResearch.org* events pages, which enables users to represent structured knowledge about events without having to deal with technical implementation challenges and ontology development themselves.

Keywords: Scientific events ontology · Knowledge engineering · Scholarly data · Linked data · Knowledge sharing

1 Introduction

Recent years have witnessed a continual growth in scholarly information: at least 114 million English-language scholarly documents are accessible on the Web [33], thanks to the ease of organizing events and of submitting and publishing manuscripts, in both academia and industry. This information, emanating from scientific events, publishing houses and social networks (e.g., *ResearchGate*) is available online in an unstructured format (e.g., call for papers (CfP) emails) or semi-structured format (e.g., event home pages) which limits the visibility and hampers the discovery of interconnected relationships for humans as well as

C. Ghidini et al. (Eds.): ISWC 2019, LNCS 11779, pp. 79–95, 2019.
https://doi.org/10.1007/978-3-030-30796-7_6

machines. This plethora of scientific literature and heterogeneity of the metadata makes it increasingly difficult to keep an overview of the current state of research. Therefore, establishing knowledge-based representation of information in scholarly communication motivates the development of data models, ontologies and knowledge graphs. Semantically enriched representation of such information makes it easier to efficiently query and process the data [1]. Consequently, collecting, integrating and analyzing the metadata of scientific events, such as association with an event series, important dates, submitted and accepted articles, venue, event type, or the field of research, is of paramount importance for tracking scientific progress [9,11]. An important topic in semantic publishing is the development of semantic models related to various scholarly communication elements in order to describe the meaning and the relationships between data, thus enabling machines to interpret meaning, which is crucial for facilitating the information needs of stakeholders including authors and publishers [22]. Given the heterogeneity of event metadata as input, semantic representation of such information involves modeling event metadata covering different types of entities involved, such as persons, organizations, location, roles of persons before/during/after the event, etc. This article tackles the problem of representing scientific events metadata in a semantic way, i.e., integrating existing events vocabularies and making explicit the relationships and interconnections between event data, thus supporting the transformation of from a "Web of documents" into a "Web of data" in the scientific domain. In this paper, we present OR-SEO (with the namespace prefix `seo`), which enables a semantically enriched representation of scholarly event metadata, interlinked with other datasets and knowledge graphs. OR-SEO does not only represent what happened, i.e., time and place of a scholarly event, but also the roles that each agent played, and the time at which this role was held by a particular agent at a particular event. OR-SEO is in use as the schema of the event pages of *OpenResearch* (OR)[1]. OpenResearch is a semantic wiki platform for crowd-sourcing such metadata and generally facilitating scholarly metadata management and exploitation [31]. We publish event metadata in a semantically structured, machine-comprehensible, and reusable way, i.e., as linked data. Standard methodologies and best practices have been considered when designing and publishing the ontology. OR-SEO has been developed using the Simplified Agile Methodology for Ontology Development (SAMOD) [21], an iterative process that aims at building the final model through a series of small steps. OR-SEO has been designed with a minimum of semantic commitment to guarantee maximum applicability for analyzing event metadata from diverse sources, and maximum reusability by datasets using the ontology for modeling different aspects of scientific events. In accordance with best practices, OR-SEO emphasizes the reuse of events-related vocabularies and the alignment with concepts between them as well as the design and visualization patterns. OR-SEO is available using persistent identifiers (https://w3id.org/seo); future versions can be collaboratively revised on a corresponding Git repository, and it is registered

[1] http://openresearch.org.

and indexed by Linked Open Vocabularies (LOV)[2]. To support knowledge discovery by automated reasoning, a set of SWRL rules has been defined. The validation of the ontology is performed on syntactic and semantic levels using the W3C RDF validation service and description logic reasoners respectively. This step is crucial for making OR-SEO reusable. We shed light on what OR-SEO contributes to the existing literature by reviewing the existing event-related models and pointing out their weaknesses. Furthermore, the ontology is aligned with existing event ontologies. A public SPARQL endpoint to query the ontology is available online (cf. Table 1).

The remainder of this paper is structured as follows: we present an overview of related models in Sect. 2. The development of the OR-SEO ontology and its structure are described in Sect. 3. The main entities in OR-SEO are described in Sect. 4. Two real-world use cases of OR-SEO are presented in Sect. 5. The evaluation of the ontology is presented in Sect. 6. We conclude with an outlook on future work in Sect. 7.

2 Related Data Models

In recent years, several data models have been developed for describing events, such as the Event Ontology (EO) [26], Linking Open Descriptions of Events (LODE) [27], the Simple Event Model (SEM) [32], Wikidata[3], and the Semantic Web Dog Food (SWDF) [19]. Typically, these models differ by focus, i.e., event type, size, and level of abstraction, and they focus on the description of event metadata, including time, location, and topical classifications of events. Early efforts towards events metadata modeling include the metadata projects of the ESWC 2006 and ISWC 2006 conferences [19], but they did not yet provide detailed descriptions of the events. The *Semantic Web Conference (SWC) ontology* is an ontology for describing academic conferences [20]. Semantic Web for Research Communities (SWRC) is an ontology for describing entities involved in research communities [29]. Compared with OR-SEO, SWC and SWRC do not cover several entities related to scientific events, such as awards, registration, Sponsorship and travel information. *Semantic Web Dog Food (SWDF)* dataset and its successor *ScholarlyData*[4] are among the pioneers of comprehensive scholarly metadata. The *Event Ontology (EO)*[5] is a simple ontology centered around four classes (Event, Agent, Factor, and Product) and 17 properties. EO has been designed as a general ontology and therefore does not cover the domain knowledge specific to scientific events. Both EO and OR-SEO reflect the domain of events, but OR-SEO describes more aspects related to scientific events and related entities, such as participants' roles, sponsors and publishers. Similarly, the Scholarly Event Description Ontology (SEDE) [17] describes scholarly events in terms of agents (e.g., persons, committees), places (e.g., cities, venues). The

[2] https://lov.linkeddata.es/dataset/lov/vocabs/seo.
[3] https://www.wikidata.org/wiki/Q1656682.
[4] http://www.scholarlydata.org.
[5] http://motools.sourceforge.net/event/event.html.

SEDE ontology provides a basis to represent, collect, and share scholarly event metadata. Compared with OR-SEO, several aspects were not considered, such as the roles of the organizers, types of events, venue, and proceedings. *Linking Open Descriptions of Events (LODE)*[6] is an ontology for describing historical events and for mapping between other event-related vocabularies and ontologies, such as Time, EO and SKOS. In other words, it links people, places, or things to an event. Compared with EO, it has some restrictions and follows a higher level of abstraction. In the latest version (of 2010), it contains one class (`Event`) and only seven properties: `illustrates`, `inSpace`, `circa`, `atPlace`, `involved`, `involvedAgent` and `atTime`. Furthermore, LODE does not model the connection of agents to events through roles. Compared with OR-SEO, LODE also does not cover entities related to scientific events, such as sponsors, publishers and hosting organization. The *Simple Event Model ontology (SEM)*[7] has a defined core, which is relatively close to EO and LODE, but still far from our ontology, in terms of describing aspects related to scientific events, which do not exist in regular events, such as publishers. SEM is formalized purely in RDFS, describing the fundamental constituents of an event, including their types, roles, temporary validity and the view according to which these constraints hold. SEM has four core classes: `Event`, `Actor`, `Place` and `Time` in addition to three types of constraints: `Role` (the role of an individual in a specific event), `Temporary` (defines the temporal boundary within which a property holds, for example, the type of the place) and `View` (defines points of view).

In the context of publishing metadata of scientific events as Linked Data, Fathalla et al. [7] published EVENTSKG, a knowledge graph featuring a comprehensive semantic description of 73 renowned event series belonging to eight computer science communities since 1969. Notably, EVENTSKG uses the updated version of OR-SEO as a reference ontology for modeling event metadata and connecting related data that was not connected in the previous release. In 2018, Gottschalk and Demidova [15] published a multilingual dataset (EVENTKG) about events and temporal relations. It describes general events at a high level of abstraction. On the contrary, we put a particular focus on scientific events and their related entities. Despite these continuous efforts, there is yet no standard, well-formed ontology covering all those aspects related to scientific events that are covered by OR-SEO, such as types of scientific events, sponsors, publishers and proceedings. OR-SEO is an extended version of the OR ontology [10]; in the comprehensive version presented here, it covers further characteristics of scientific events such as acceptance rate, schedule (submission deadline, notification date, etc.), awards, authors registration types, and social media dissemination (e.g., Twitter account).

[6] http://linkedevents.org/ontology/.
[7] https://semanticweb.cs.vu.nl/2009/11/sem/.

3 OR-SEO Development

OR-SEO is developed to be used as a reference ontology for the conceptualization of scholarly event metadata and capturing the corresponding concepts. It follows the best state-of-the-art practices and design principles for relevant and reusable ontologies. We first point out general design principles, then introduce the terms that we defined for representing the metadata of scientific events.

3.1 Design Principles and Requirements

The best practices within the Semantic Web community have been followed from the initial steps of the OR-SEO development [2]. The paramount intention behind our decision to develop an ontology for scholarly events is that, to the best of our knowledge, there is a need for a well-formed ontology in this domain to describe scholarly events. In particular, aspects related particularly to scholarly events are not covered by existing ontologies, such as roles of organizers, e.g., proceedings chair, sponsors, event proceedings, and quality metrics such as acceptance rate and the ranking of the event. Inspired by Linked Data principles [16], the following design decisions have been made while developing OR-SEO:

- *Addressing different stakeholders:* OR-SEO is developed to be used in the OpenResearch platform, supporting, e.g., authors to find high-impact events to submit their work to, and event chairs and proceedings publishers to derive useful facts to assess the impact of their events and the competing ones.
- *Broad coverage of the relevant concepts:* Events, according to OR-SEO, comprise everything that happens, no matter whether there is a specific place or time, or agents involved.
- *Flexibility and ease of changes:* The use of any class and their corresponding properties is optional, i.e., there are no property or cardinality restrictions such as `owl:allValuesFrom`.
- *Reusability:* We only use `rdfs:domain` and `rdfs:range` to indicate where to use properties. This facilitates the reuse of OR-SEO by other ontologies.
- *Efficient reasoning:* In the development of OR-SEO, several logic rules have been taken into consideration in order to facilitate efficient reasoning.
- *Availability:* The ontology has been published under a persistent URL (cf. Table 1) under the open CC-BY 3.0 license. OR-SEO is published according to the best practices of the Linked Data community [2]; its source is available from a *GitHub* repository (cf. Table 1). The ontology has been made discoverable through LOV, a high-quality catalog of well-documented vocabularies for data on the Web.
- *Validation:* Two types of validation have been performed: *syntactic* and *semantic* validation. We syntactically validated OR-SEO to conform with the W3C RDF standards using the online RDF validation service[8]. The dereferenceability of the URIs of the OR-SEO terms over the HTTP protocol

[8] https://www.w3.org/RDF/Validator/.

(cf. [18]) has been validated using Vapour[9]. We semantically validated OR-SEO using Protégé reasoners such as FaCT++[10] , and the OOPS! Ontology Pitfall Scanner[11], for detecting inconsistencies.
- *Documentation:* The documentation for the ontology is available online through its PURL. Detailed information about entities and properties are also included in the ontology, i.e., as `rdfs:comment`s.
- *Adoption and Sustainability:* OR-SEO is maintained and used by the editors of OR to represent metadata of scientific events so far mainly in computer science but also some other fields including physics and chemistry. OR-SEO also has an issue tracker on its GitHub repository in order to make it easier to request new features, e.g., re-using related ontologies that may appear in future, and to report any problems.
- *Metadata completion*: We followed the best practices for completing the vocabulary metadata proposed in [13].

Table 1. OR-SEO-related resources

Resource	URL
PURL	https://w3id.org/seo
Turtle file	http://kddste.sda.tech/SEOontology/SEO.ttl
RDF/XML file	http://kddste.sda.tech/SEOontology/SEO.rdf
GitHub repository	https://github.com/saidfathalla/SEOontology
Issue Tracker	https://github.com/saidfathalla/SEOontology/issues
SPARQL endpoint	http://kddste.sda.tech/SEOontology/sparql
VoID	http://kddste.sda.tech/SEOontology/VoID.nt

3.2 Challenges and Requirements

Towards the development of an ontology for scholarly events, challenges started with identifying the pitfalls in the state-of-the-art model. In addition, the scholarly events domain itself relates entities from diverse information sources including bibliographical information, spatial and temporal data. Therefore, data models necessitates an effective integration of concepts and their semantics. After studying the domain and the state-of-the-art model, the diversity of information representation and large amount of data pose high requirements to be addressed by OR-SEO. The ontology should be maintainable with respect to the evolution of linked data vocabularies and adaptable to other domains of science. A part of these requirements will be represented as a set of competency questions related

[9] http://linkeddata.uriburner.com:8000/vapour?.
[10] https://github.com/ethz-asl/libfactplusplus.
[11] http://oops.linkeddata.es/.

to different use cases that the ontology should be able to answer. Some of these questions are: (1) *Which events related to the target domain X, e.g., "semantic web", took place in country Y over a particular time span, with an acceptance rate less than a value Z*? (2) *What are the top-X countries hosting most of the events belonging to "Security and Privacy" in the past decade?*", and (3) *In which events did person X participate in the organization committee*? More competency questions and the corresponding SPARQL queries are available at OpenResearch.org.[12]

3.3 Reuse of Existing Ontological Knowledge

Techniques for efficient and effective reuse of ontological knowledge are key factors in developing ontology-based systems [28]. A challenging task for ontology engineers is to decide in advance, which of the available vocabularies are the most useful ones for reuse, especially because the Web allows reuse across domains. By its nature, the scientific events domain involves entities from various other domains, including location, agents, time, and scholarly data, as shown in Fig. 1. Therefore, the first step in building our ontology is reusing terms from related ontologies, since, the more vocabularies a model reuses, the higher the value of its semantic data is. We have selected the most closely related ontologies listed in the Linked Open Vocabularies directory (LOV). The reuse of these vocabularies by explicitly linking to them brings OR-SEO its richness. We reuse several well-known ontologies to make OR-SEO interoperable in different contexts:

- The *Semantic Web Conference (SWC) ontology*, one of the vocabularies of choice for describing academic conferences [20], is used to represent, e.g., `Conference`s and `ConferenceSeries`.
- *Time-indexed Value in Context (TVC)*, a standard ontology design pattern to describe a time-indexed situation that expresses a particular role held by an agent at an event [22], is used to represent, e.g., `Duration` and `Interval`.
- *Dublin Core Metadata Initiative (DCMI)* is used to describe metadata of typical entities in scientific events, e.g., of `Agent`s or `Proceeding`s,
- The *Friend-of-a-Friend (FOAF)* ontology describes involved persons and their social network profiles,
- *Semantically-Interlinked Online Communities (SIOC)* describes information from online communities such as `Role` and `Site` [3]
- SPAR ontologies [23] describe research papers type (fabio), publications identifiers (datacite) and document parts (doco).
- SemSur ontology describes research findings based on an explicit semantic representation of the knowledge contained in scientific publications [8], and
- DBpedia Ontology (dbo)[13] is used to represent geographical data, such as `dbo:Country` and `dbo:City`.

[12] https://www.openresearch.org/wiki/Sparql_endpoint/Examples.
[13] http://dbpedia.org/ontology/.

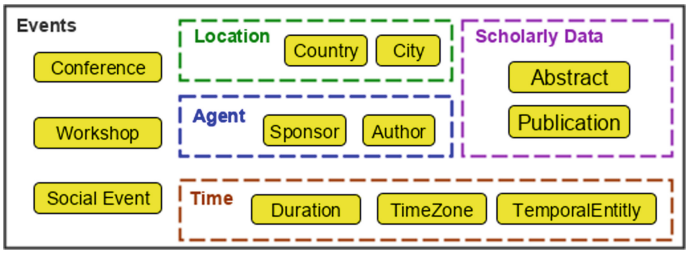

Fig. 1. A layered view of the domains, with corresponding entities of SEO.

4 Ontology Description

The SAMOD agile methodology is used for developing OR-SEO ontology. SAMOD takes into consideration various issues when developing ontologies to achieve a "data-centric" model, such as avoiding inconsistencies, being self-explanatory, and giving examples of usage. This section describes the main entities in the scientific events ontology. We focus on core classes and properties, and reasoning support provided by the ontology. More details can be found in the online documentation of the ontology.

4.1 Core Classes

The OR-SEO ontology imports some of the main classes from the ontologies introduced in Subsect. 3.3. For the ones not explicitly matching with the concepts addressed by OR-SEO, new definitions have been developed. The core entities of the scholarly events in OR-SEO are: (1) *Event*, as the entity of main interest, including metadata such as event type (e.g., conference or workshop), bibliographic and retrospective information (the numbers of submitted and accepted articles, information about the attendees, tracks), (2) *Agents*, including the *Organizations* hosting or sponsoring the event and *Persons* involved in the organization of the events in different roles, (3) *Role during event* of such stakeholders and persons, (4) *Location*, the city and country in which the event was held, (5) *Proceedings*, the proceedings including the publications produced by the event, and (6) *Time*, to describe the duration of events. Concretely, these entities are represented in OR-SEO as follows (see Figs. 2 and 3): `OrganizedEvent` represents the event itself and all the sub-events of those which are about the topic or theme of the main event, such as academic or non-academic events. `Agent` represents a person, group, company or organization, which can be a sponsor or a publisher of the proceedings of the event. `RoleDuringEvent` represents a time indexed situation that expresses a role held by an agent in the context of the event. `Country/City` represents the physical location of the event. `Proceedings` represents proceedings produced by academic events. `TemporalDuration` is a time interval representing the duration of the event. Agents, i.e., persons and organizations, play a key role in the scholarly events domain. Agents hold different roles (`RoleDuringEvent`) in participating in scholarly events, including

Publishing Role During Event, *Organizational Role During Event* and *Chair Role During Event*.

Class Specialization: Because of the complexity and diversity of the concepts, some of the defined or reused classes need more specialization, so we created respective subclasses. For instance, `Symposium` has been added as a subclass of the `AcademicEvent` class, and another subclass to represent the series of such events as `SymposiumSeries` to the super class `EventSeries`. In addition, a set of classes missing from other ontologies, for example, to describe agents and their roles more specifically, such as `Publisher` and `Sponsor`, have been added.

Class Disjointness: We assert pairwise disjointness, where applicable, between any of the classes in the ontology. For instance, the `IrregularRegistration` class is disjoint with `RegularRegistration` and `LateRegistration` is disjoint with `EarlyBirdRegistration`.

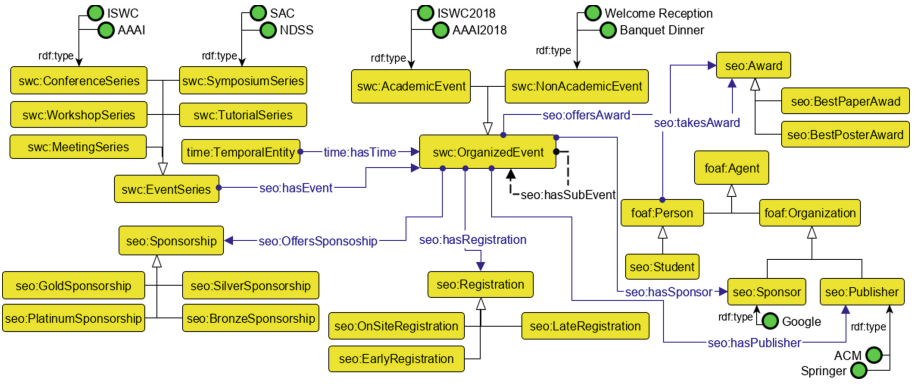

Fig. 2. Core concepts in OR-SEO and their relationships. Arrows with open arrow heads denote `rdfs:subClassOf` properties between the classes.

4.2 Properties

OR-SEO's properties are divided into two categories: newly defined and directly reused properties. We indicate the classes to be used with several data and object properties by defining domain and range using `rdfs:domain` and `rdfs:range` respectively. For instance, we capture the domain of newly-defined data properties for describing abstract and submission deadline, i.e., `seo:abstractDeadline` and `seo:submissionDeadline`, to be `swc:AcademicEvent` and the range to be `xsd:dateTime`. In addition, OR-SEO defines its own object properties, such as `seo:belongsToSeries`, `seo:hasTrack`, `seo:colocatedWith`, `seo:hasPublisher`. Some properties have complex ranges, e.g., `seo:hasRegistrationType` has range (`LateRegistration`

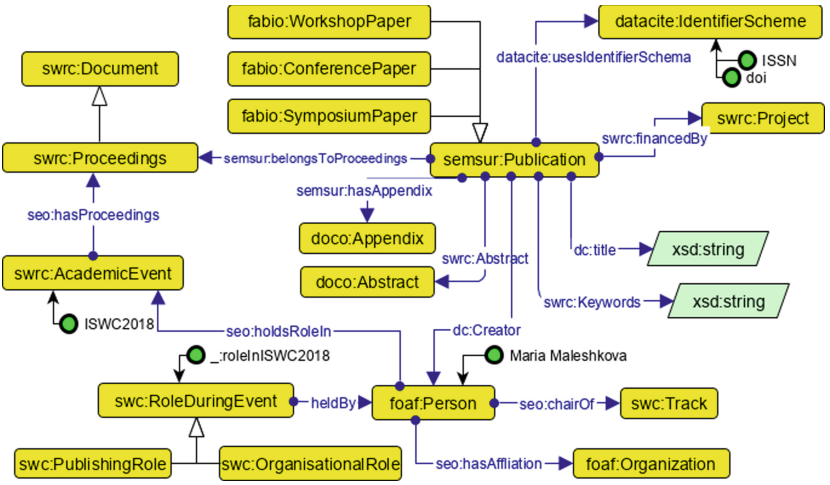

Fig. 3. Publications and roles of Agents during a scientific event.

⊔ `EarlyBirdRegistration`) because these two classes are disjoint. Ontology design patterns are applied, e.g., the OWL patterns of Gangemi [12], to capture notions such as inverse relations and composition of relations. There are some inverse relations, e.g., `seo:isTrackOf` is the inverse of `seo:hasTrack` and `seo:isSponsorOf` is the inverse of `seo:hasSponsor`. Thus, if an event E `seo:hasTrack` T, then it can be inferred that T `seo:isTrackOf` E. Also, some symmetric relations are defined, such as `seo:colocatedWith`, e.g., if an event E_1 is co-located with another one E_2, then it could be inferred that E_2 is also co-located with E_1. Furthermore, it is a property whose domain is the same as its range, which provides the information that an organized event can only be co-located with another organized event, and a reflexive relation, i.e., each event is co-located with itself. Such definitions allow to reveal implicit information and increase the coherence and thus the value of event metadata.

Representation Pattern for n-ary Relations. One common representation of n-ary relations is to represent the relation as a class rather than a property, and using n properties to point to the related entities. Instances of such classes are instances of the n-ary relation and additional properties can provide binary links to each argument of the relation, i.e., an instance of the relation linking the n individuals. For more illustration, consider the case of *Maria Maleshkova*, the sponsorship chair in the ISWC conference in 2018. As shown in Fig. 4, the individual *:roleInISWC2018* represents a single object encapsulating both the event, the person that had a role there, and the type of the role in that event.

4.3 Reasoning

Inference on the Semantic Web is additionally used to improve the quality of data integration in the ontology by combining rules and ontologies to discover new

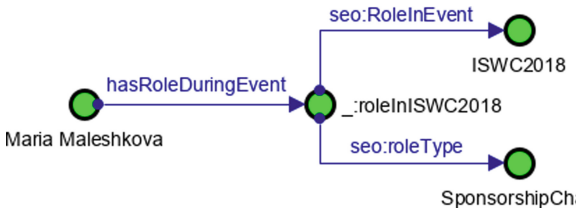

Fig. 4. Representation pattern for n-ary relations in OR-SEO.

relationships, detect possible inconsistencies and infer logical consequences from a set of asserted facts or axioms. Drools reasoner [25] is one of the reasoners that the Protégé ontology development environment uses for performing rule-based inference. Our goal is to define a rule set for discovering new relationships and inferring new knowledge that did not explicitly exist in a knowledge graph. Therefore, a set of rules following the Semantic Web Rule Language (SWRL) [24] has been defined and written using the SWRLtab plugin for Protégé 5.2. SWRL was designed based on a combination of the OWL DL and OWL Lite sub-languages of OWL Full. SWRL allows users to write Horn-like rules expressed in terms of OWL classes and properties to reason about OWL individuals. A set of rules to support the inference in OR-SEO have been defined. These rules have been semantically validated using Drools reasoner. The rule set in OR-SEO includes the following SWRL rules (for readability, we omitted namespaces). Using Formula 1, participants in a specific event can be easily inferred, while using Formula 2, the location of one event can be determined from a co-located event.

$$Agent\,(?a) \wedge holdsRole\,(?a, ?e) \rightarrow participatesIn\,(?a, ?e) \quad (1)$$

$$colocatedWith\,(?e1, ?e2) \;\wedge\; hasLocation\,(?e1, ?l) \rightarrow hasLocation\,(?e2, ?l) \quad (2)$$

5 Real-World Use Cases

This section presents two real-world use cases for the OR-SEO ontology: OpenResearch.org and the EVENTSKG dataset.

Use Case 1. As populating ontologies with instances is a time-consuming and error-prone task, OR-SEO is in use on 6,800+ event pages on *OpenResearch* [31], which facilitates the creation of instances of events and events series as wiki pages, without having to go into the details of the implementation of the ontology. It is an extended version of the original ontology of *OpenResearch*, which has been redesigned and systematically validated. Data acquisition in *OpenResearch* follows an approach that combines manual/crowd-sourced contribution and semi-automated methods. *OpenResearch* provides semantic descriptions of scientific events, publications, tools and organizations using ontologies for each such entity type. Semantic MediaWiki (SMW), a semantically enhanced wiki

engine, is the core software for *OpenResearch* that serves as data curation interface employing semantic forms. OpenResearch employs one semantic form per core class of OR-SEO; combined with properties, they enable semantic annotations in the wiki markup. Semantic forms enable users to create and modify the knowledge graph via forms, without the need for actual programming. Listing 1 shows an example of an individual event (ISWC 2018) created on *OpenResearch*[14]. Furthermore, semantically annotated text is found at the bottom of the corresponding wiki page of the ISWC series[15], which represents the metadata of the event using corresponding terms of the ontology, such as chairs, country, or Twitter account. For instance, the *info box* on the right contains the metadata of the events series, including full title, bibliography, CORE 2017 and 2018 ranks, and the average acceptance rate. Semantically annotated metadata can be exported as RDF triples using the *"RDF feed"* feature. Several interesting information can be exposed from OpenResearch, such as a list of upcoming events in a Calendar view[16], and top-ranked events along with their ranking and average acceptance rate[17]. Finally, such ontologies and events metadata added by the community extend *OpenResearch*'s distributed data collection by embedding markup in conference websites aligned with schema.org, and links to other portals and services.

Use Case 2. The second use case of OR-SEO is the representation of a comprehensive dataset (EVENTSKG) of scholarly events sourced from several resources and curated semi-automatically [5–7]. Going beyond existing work (cf. Sect. 2), it comprises metadata of 73 renowned events in eight computer science communities using OR-SEO as its schema. EVENTSKG is not only able to answer quantitative questions, but it also provides qualitative information, such as which countries hosted most events in a particular community.

Listing 1. Use case 1. Representation of metadata on OpenResearch.org in its markup language.

```
{{Event
| Title      = 17th International Semantic Web Conference
| Series     = ISWC            | Type       = Conference
| Field      = Linked Data     | Start date = 2018/10/08
| End date   = 2018/10/12      | Homepage   = iswc2018.semanticweb.org/
| Twitter    = @iswc2018       | City       = Monterey
| Country    = USA
}}
```

[14] http://openresearch.org/wiki/ISWC_2018.
[15] https://www.openresearch.org/wiki/ISWC.
[16] https://www.openresearch.org/mediawiki/index.php?title=Events_Calendar&field=Science.
[17] https://www.openresearch.org/mediawiki/index.php?title=Series&field=Science.

Listing 2. Use case 2. Using OR-SEO in metadata representation for ISWC 2015 in EVENTSKG, in Turtle.

```
### https://w3id.org/seo#ISWC2015
ekg:ISWC2015  rdf:type  owl:NamedIndividual ,
conference-ontology:Conference;
 seo:belongsToSeries      ekg:ISWC ;
 seo:acceptanceRate       "0.22"^^xsd:decimal;
 seo:submittedPapers      "172"^^xsd:int;
 seo:acceptedPapers       "38"^^xsd:int;
 seo:city                 <http://dbpedia.org/page/Bethlehem>;
 seo:country              <http://dbpedia.org/page/United_States>;
 seo:field                seo:InformationSystem ;
 conference-ontology:startDate "2015-10-11"^^xsd:date;
 conference-ontology:endDate  "2015-10-15"^^xsd:date;
 seo:eventWebsite         "http://iswc2015.semanticweb.org/"^^xsd:anyURI.
```

The aim is to transform event metadata, distributed across different sources, to Linked Open Data, which can be interpreted by machines to create innovative event-related services. Listing 2 shows the metadata of ISWC 2015 in EVENTSKG. Three major prefixes are used in metadata representation namely: `ekg`, `seo` and `conference-ontology` according to http://prefix.cc/.

6 Evaluation

Evaluating ontologies is the process of measuring the quality of the ontology content, ensuring that its definitions satisfy the requirements or perform correctly in the real world [14]. In other words, the quality of ontologies can be assessed using metrics that evaluate the success of the ontology in modeling a real-world domain (as illustrated in Sect. 5). Ontologies can be evaluated against a gold standard, or using a criteria-based or task-based evaluation [34]. This is majorly a manual task because it is difficult to construct automated tests to compare ontologies using such criteria [4]. We assess OR-SEO using a criteria-based evaluation as proposed by Tartir et al. [30]. They proposed an ontology evaluation model, called OntoQA, which evaluates the ontology using schema metrics and instance metrics. We evaluate the ontology design by comparing to the related work (with the best coverage of the domain, i.e., SWC, SEDE, and SWRC).

- *Attribute richness (AR)* refers to the average number of attributes per class. Formally, $AR = A/C$, the number of attributes for all classes (A) divided by the number of classes (C). The more attributes are defined, the more knowledge the ontology provides.
- *Relationship richness (RR)* refers to the diversity of relations and the placement of them in the ontology. Formally, $RR = R/(S + R)$, the number of relationships (R) defined in the schema, divided by the sum of the number of sub-classes (S) and the number of relationships. The more relations, except is-a relations, the ontology has, the richer it is.

– *Inheritance richness (IR)* refers to the average number of sub-classes per class. Formally, $IR = S/C$, the number of sub-classes divided by the sum of the number of classes. A high IR means that ontology represents a wide range of general knowledge, i.e., is of a horizontal nature.

Table 2. Evaluation of OR-SEO using OntoQA model

Ontology	Classes	Sub-classes	Attributes	Relations	AR	RR	IR
SWC	390	351	118	189	0.30	0.40	0.90
SEDE	122	46	47	56	0.39	0.60	0.38
SWRC	248	221	51	57	0.21	0.21	0.89
OR-SEO	165	197	93	177	**0.57**	**0.61**	**1.19**

As shown in Table 2, OR-SEO has a moderate size but an overall beneficial knowledge structure. Among similar domain ontologies it has the largest *AR* which enables the provision of more knowledge per instance. Regarding *RR*, OR-SEO has moderate diversity of relations and has much richer relations in comparison with SWC and SWRC, and slightly richer than SEDE. Regarding *IR*, OR-SEO has the highest value of all ontologies (1.19), which means that it represents a wider range of knowledge than the state of the art. In terms of usability evaluation, most of the users of OpenResearch found it easy to populate the ontology via a user-friendly interface, i.e., SMW semantic forms. For instance, event organizers, or even any researcher interested in an event, can add event series or an individual event metadata using "*Add event series*"[18] and "*Add event*"[19] semantic forms, respectively. As mentioned before in Sect. 5, the produced data are wiki pages presenting events metadata in a user-friendly way.

7 Conclusions and Future Work

We presented OR-SEO, a reference ontology for capturing metadata of scientific events. Its real-world instantiation in the OpenResearch platform is discussed with some inference rules to discover new relationships, detect possible inconsistencies and infer logical consequences from a set of asserted facts. We shed light on what OR-SEO contributes to the existing literature by reviewing the existing event-related models, pointing out their weaknesses. Actually, OR-SEO covers issues closely related to scholarly events, which are not covered by other scholarly communication domain ontologies, such as types of scholarly events, sponsor, publisher and proceedings. Furthermore, OR-SEO models scholarly events characteristics, such as acceptance rate, submission deadline, and notification date, and Twitter account. The ontology is publicly available online, following

[18] https://www.openresearch.org/wiki/Special:FormEdit/EventSeries.
[19] https://www.openresearch.org/wiki/Special:FormEdit/Event.

ontology resource publication best practices. We showed that it fits well for a heterogeneous set of existing metadata covered by the OpenResearch platform. The ontology will continue to be maintained and extended in the context of the *OpenResearch* effort, aiming at large scale event data acquisition and analysis through applying semi-automated and crowd-sourcing methods. We hope that OR-SEO will thus contribute to facilitating the representation and analysis of the currently not yet well-structured space of scholarly event information, thus supporting all stakeholders of events, particularly including organizers and potential authors.

Regarding future work in the context of the maintenance plan of OR-SEO we envision to: (1) model event evolution considering property changes such as type, e.g., from symposium to conference, or events re-scheduled, or events whose chairs changed, (2) adapt the ontology to cover events in other research fields, such as Physics, Mathematics, and Engineering, where scholarly events take a different shape, (3) improve the coverage, by including more concepts related to sponsorship, event's program, social events within the event itself and events' calls for papers, (4) model other publishing venues such as journals, and (5) develop a smart data analytics tool in order to assess events' progress and recommend relevant events to potential authors and a SPARQL endpoint and a Linked Data navigator to browse the ontology and its instances.

Acknowledgement. This work is supported by EPSRC grant EP/M025268/1, WWTF grant VRG18-013, EC Horizon 2020 grants LAMBDA (#809965) and the ERC project ScienceGRAPH (#819536).

References

1. Auer, S., Kovtun, V., Prinz, M., Kasprzik, A., Stocker, M., Vidal, M.E.: Towards a knowledge graph for science. In: 8th International Conference on Web Intelligence, Mining and Semantics. ACM (2018)
2. Berrueta, D., Phipps, J., Miles, A., Baker, T., Swick, R.: Best practice recipes for publishing RDF vocabularies. In: Working draft, W3C (2008)
3. Breslin, J.G., Decker, S., Harth, A., Bojars, U.: SIOC: an approach to connect web-based communities. Int. J. Web Based Commun. **2**(2), 133–142 (2006)
4. Brewster, C., Alani, H., Dasmahapatra, S., Wilks, Y.: Data driven ontology evaluation. In: 4th International Conference on Language Resources and Evaluation, LREC 2004 (2004)
5. Fathalla, S., Lange, C.: EVENTS: a dataset on the history of top-prestigious events in five computer science communities. In: González-Beltrán, A., Osborne, F., Peroni, S., Vahdati, S. (eds.) SAVE-SD 2017-2018. LNCS, vol. 10959, pp. 110–120. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01379-0_8
6. Fathalla, S., Lange, C.: EVENTSKG: a knowledge graph representation for top-prestigious computer science events metadata. In: Nguyen, N.T., Pimenidis, E., Khan, Z., Trawiński, B. (eds.) ICCCI 2018. LNCS (LNAI), vol. 11055, pp. 53–63. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98443-8_6

7. Fathalla, S., Lange, C., Auer, S.: EVENTSKG: a 5-star dataset of top-ranked events in eight computer science communities. In: Hitzler, P., et al. (eds.) ESWC 2019. LNCS, vol. 11503, pp. 427–442. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21348-0_28
8. Fathalla, S., Vahdati, S., Auer, S., Lange, C.: SemSur: a core ontology for the semantic representation of research findings. Proc. Comput. Sci. **137**, 151–162 (2018)
9. Fathalla, S., Vahdati, S., Auer, S., Lange, C.: Metadata analysis of scholarly events of computer science, physics, engineering, and mathematics. In: Méndez, E., Crestani, F., Ribeiro, C., David, G., Lopes, J.C. (eds.) TPDL 2018. LNCS, vol. 11057, pp. 116–128. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00066-0_10
10. Fathalla, S., Vahdati, S., Auer, S., Lange, C.: The scientific events ontology of the openresearch.org curation platform. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (2019)
11. Fathalla, S., Vahdati, S., Lange, C., Auer, S.: Analysing scholarly communication metadata of computer science events. In: Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L., Karydis, I. (eds.) TPDL 2017. LNCS, vol. 10450, pp. 342–354. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67008-9_27
12. Gangemi, A., Presutti, V.: Ontology design patterns. In: Staab, S., Studer, R. (eds.) Handbook on Ontologies. IHIS, pp. 221–243. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-92673-3_10
13. Garijo, D., Poveda-Villalón, M.: A checklist for complete vocabulary metadata (2017). https://w3id.org/widoco/bestPractices
14. Gómez-Pérez, A.: Evaluation of ontologies. Int. J. Intell. Syst. **16**(3), 391–409 (2001)
15. Gottschalk, S., Demidova, E.: EventKG: a multilingual event-centric temporal knowledge graph. In: Gangemi, A., et al. (eds.) ESWC 2018. LNCS, vol. 10843, pp. 272–287. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-93417-4_18
16. Heath, T., Bizer, C.: Linked data: evolving the web into a global data space. Synth. Lect. Semant. Web Theory Technol. **1**(1), 1–136 (2011)
17. Jeong, S., Kim, H.-G.: SEDE: an ontology for scholarly event description. J. Inf. Sci. **36**(2), 209–227 (2010)
18. Lewis, R.: Dereferencing HTTP URIs. In: Draft Tag Finding (2007). http://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14.html
19. Möller, K., Heath, T., Handschuh, S., Domingue, J.: Recipes for semantic web dog food—The ESWC and ISWC metadata projects. In: Aberer, K., et al. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 802–815. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_58
20. Nuzzolese, A.G., Gentile, A.L., Presutti, V., Gangemi, A.: Semantic web conference ontology - a refactoring solution. In: Sack, H., Rizzo, G., Steinmetz, N., Mladenić, D., Auer, S., Lange, C. (eds.) ESWC 2016. LNCS, vol. 9989, pp. 84–87. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47602-5_18
21. Peroni, S.: A simplified agile methodology for ontology development. In: Dragoni, M., Poveda-Villalón, M., Jimenez-Ruiz, E. (eds.) OWLED/ORE -2016. LNCS, vol. 10161, pp. 55–69. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54627-8_5
22. Peroni, S.: The semantic publishing and referencing ontologies. Semantic Web Technologies and Legal Scholarly Publishing. LGTS, vol. 15, pp. 121–193. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-04777-5_5

23. Peroni, S., Shotton, D.: The SPAR ontologies. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11137, pp. 119–136. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_8
24. Plinere, D., Borisov, A.: SWRL: rule acquisition using ontology. Sci. J. Riga Tech. Univ. Comput. Sci. **40**(1), 117–122 (2009)
25. Proctor, M.: Drools: a rule engine for complex event processing. In: Schürr, A., Varró, D., Varró, G. (eds.) AGTIVE 2011. LNCS, vol. 7233, pp. 2–2. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34176-2_2
26. Raimond, Y., Abdallah, S.: The event ontology (2007)
27. Shaw, R., Troncy, R., Hardman, L.: LODE: linking open descriptions of events. In: Gómez-Pérez, A., Yu, Y., Ding, Y. (eds.) ASWC 2009. LNCS, vol. 5926, pp. 153–167. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10871-6_11
28. Simperl, E.: Reusing ontologies on the semantic web: a feasibility study. Data Knowl. Eng. **68**(10), 905–925 (2009)
29. Sure, Y., Bloehdorn, S., Haase, P., Hartmann, J., Oberle, D.: The SWRC ontology – semantic web for research communities. In: Bento, C., Cardoso, A., Dias, G. (eds.) EPIA 2005. LNCS (LNAI), vol. 3808, pp. 218–231. Springer, Heidelberg (2005). https://doi.org/10.1007/11595014_22
30. Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P., Aleman-Meza, B.: OntoQA: metric-based ontology quality analysis. In: IEEE ICDM Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources (2005)
31. Vahdati, S., Arndt, N., Auer, S., Lange, C.: OpenResearch: collaborative management of scholarly communication metadata. In: Blomqvist, E., Ciancarini, P., Poggi, F., Vitali, F. (eds.) EKAW 2016. LNCS (LNAI), vol. 10024, pp. 778–793. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49004-5_50
32. Van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the simple event model (SEM). Web Seman. Sci. Serv. Agents World Wide Web **9**(2), 128–136 (2011)
33. Wu, Z., Wu, J., Khabsa, M., Williams et al.: Towards building a scholarly big data platform: Challenges, lessons and opportunities. In: 14th ACM/IEEE- CS Joint Conference on Digital Libraries. IEEE Press (2014)
34. Yu, J., Thom, J.A., Tam, A.: Ontology evaluation using Wikipedia categories for browsing. In: 16th ACM Conference on Conference on Information and Knowledge Management (2007)