



# EventKG: A Multilingual Event-Centric Temporal Knowledge Graph

Simon Gottschalk<sup>(✉)</sup>  and Elena Demidova 

L3S Research Center, Leibniz Universität Hannover, Hannover, Germany  
{gottschalk,demidova}@L3S.de

**Abstract.** One of the key requirements to facilitate semantic analytics of information regarding contemporary and historical events on the Web, in the news and in social media is the availability of reference knowledge repositories containing comprehensive representations of events and temporal relations. Existing knowledge graphs, with popular examples including DBpedia, YAGO and Wikidata, focus mostly on entity-centric information and are insufficient in terms of their coverage and completeness with respect to events and temporal relations. EventKG presented in this paper is a multilingual event-centric temporal knowledge graph that addresses this gap. EventKG incorporates over 690 thousand contemporary and historical events and over 2.3 million temporal relations extracted from several large-scale knowledge graphs and semi-structured sources and makes them available through a canonical representation.

**Resource type:** Dataset

**Permanent URL:** <http://eventkg.l3s.uni-hannover.de>

## 1 Introduction

*Motivation:* The amount of event-centric information regarding contemporary and historical events of global importance, such as Brexit, the 2018 Winter Olympics and the Syrian Civil War, constantly grows on the Web, in the news sources and within social media. Efficiently accessing and analyzing large-scale event-centric and temporal information is crucial for a variety of real-world applications in the fields of Semantic Web, NLP and Digital Humanities. In Semantic Web and NLP, these applications include Question Answering [14] and timeline generation [1]. In Digital Humanities, multilingual event repositories can facilitate cross-cultural studies that aim to analyze language-specific and community-specific views on historical and contemporary events (examples of such studies can be seen in [11, 18]). Furthermore, event-centric knowledge graphs can facilitate the reconstruction of histories as well as networks of people and organizations over time [19]. One of the pivotal pre-requisites to facilitate effective analytics of contemporary and historical events is the availability of knowledge repositories providing reference information regarding events, involved entities and their temporal relations (i.e. relations valid over a time period).

*Limitations of the Existing Sources of Event-Centric and Temporal Information:* Currently, event representations and temporal relations are spread across heterogeneous sources. First, large-scale knowledge graphs (KGs) (i.e. graph-based knowledge repositories [7] such as Wikidata [6], DBpedia [16], and YAGO [17]) typically focus on entity-centric knowledge. Event-centric information included in these sources is often not clearly identified as such, can be incomplete and is mostly restricted to named events and encyclopedic knowledge. For example, as it will be discussed later in Sect. 5, of 322,669 events included in EventKG, only 18.70% are classified using the `dbo:Event` class in the English DBpedia. Furthermore, event descriptions in the existing knowledge graphs often lack the key properties such as times and locations. For example, only 33% of the events in Wikidata provide temporal and 11.70% spatial information. Second, a variety of manually curated semi-structured sources (e.g. Wikipedia Current Events Portal (WCEP) [22] and multilingual Wikipedia event lists) contain information on contemporary events. However, the lack of structured representations of events and temporal relations in these sources hinders their direct use in real-world applications through semantic technologies. Third, recently proposed knowledge graphs containing contemporary events extracted from unstructured news sources (such as [19]) are potentially highly noisy (e.g. [19] reports an extraction accuracy of 0.55) and are not yet widely adopted. Finally, the sources of event-centric information that can potentially be explored in future work are Web markup [21] and event-centric Web crawls [8]. Overall, a comprehensive integrated view on contemporary and historical events and their temporal relations usable for real-world applications is still missing. The provision of EventKG will help to overcome these limitations.

*EventKG and Advances to the State of the Art:* EventKG presented in this paper takes an important step to facilitate a global view on events and temporal relations currently spread across entity-centric knowledge graphs and manually curated semi-structured sources. EventKG extracts and integrates this knowledge in an efficient light-weight fashion, enriches it with additional features, such as indications of relation strengths and event popularity, adds provenance information and makes all this information available through a canonical representation. EventKG follows best practices in data publishing and reuses existing data models and vocabularies (such as Simple Event Model [23] and the DBpedia ontology) to facilitate its efficient reuse in real-world applications through the application of semantic technologies and open standards (i.e. RDF and SPARQL). EventKG currently includes data sources in five languages – English (en), German (de), French (fr), Russian (ru) and Portuguese (pt) – and is extensible. The main contributions of EventKG are as follows:

- A multilingual RDF knowledge graph incorporating over 690 thousand events and over 2.3 million temporal relations in V1.1 extracted from several large-scale entity-centric knowledge graphs (i.e. Wikidata, DBpedia in five language editions and YAGO), as well as WCEP and Wikipedia event lists in five languages. In the following, we refer to these sources used to populate EventKG as *the reference sources*. The key features of EventKG include:

- provision of event-centric information (including historical and contemporary events) and temporal relations using a canonical representation;
  - light-weight integration and fusion of event representations and relations originating from heterogeneous reference sources;
  - higher coverage and completeness of event representations compared to the individual reference sources (see Sect. 5);
  - provision of interlinking information, to facilitate e.g. an assessment of relation strength and event popularity;
  - provenance for all information contained in EventKG.
- An open source extraction framework to extract and maintain up-to-date versions of EventKG, extensible to further languages and reference sources.

*Comparison to Other Existing Resources:* To the best of our knowledge, currently there are no dedicated knowledge graphs aggregating event-centric information and temporal relations for historical and contemporary events directly comparable to EventKG. The heterogeneity of data models and vocabularies for event-centric and temporal information (e.g. [12, 19, 20, 23]), the large scale of the existing knowledge graphs, in which events play only an insignificant role, and the lack of clear identification of event-centric information, makes it particularly challenging to identify, extract, fuse and efficiently analyze event-centric and temporal information and make it accessible to real-world applications in an intuitive and unified way. Through the light-weight integration and fusion of event-centric and temporal information from different sources, EventKG enables to increase coverage and completeness of this information. For example, EventKG increases the coverage of locations and dates for Wikidata events it contains by 14.43% and 17.82%, correspondingly (see Table 6 in Sect. 5 for more detail). Furthermore, existing sources lack structured information to judge event popularity and relation strength as provided by EventKG – the characteristic that gains the key relevance given the rapidly increasing amount of event-centric and temporal data on the Web and the information overload.

## 2 Relevance

*Relevance to the Semantic Web Community and Society:* Our society faces an unprecedented number of events that impact multiple communities across language and community borders. In this context, efficient access to, as well as effective disambiguation of, and analytics of event-centric multilingual information originating from different sources, as facilitated by EventKG, is of utmost importance for several scientific communities, including Semantic Web, NLP and Digital Humanities. In the context of the Semantic Web community, application areas of EventKG include event-centric Question Answering and ranking-based timeline generation that requires assessment of event popularity and relation strength. In Digital Humanities, EventKG as a multilingual event-centric repository can provide a unique source for cross-cultural and cross-lingual event-centric

**Table 1.** All sub-events of the World War II in EventKG that started between February 12 and February 28, 1941.

Start date	Sources	Description
Feb 12	Wikipedia event lists	Erwin Rommel arrives in Tripoli
Feb 17	YAGO, DBpedia	Battle of Trebeshina
Feb 25	YAGO, DBpedia	Operation Abstention
Feb 27	YAGO, DBpedia, Wikidata <sup>†</sup>	Action of 27 February 1941

<sup>†</sup>Wikidata misses the fact that this action is part of the World War II.

analytics (e.g. illustrated in [11, 18]), while reducing barriers of data extraction, integration and fusion.

*Relevance for Question Answering Applications:* In the field of Question Answering (QA) [14], the current focus of research is on the generation of formal query expressions (e.g. in the SPARQL query language) from user queries posed in a natural language as well as interactive approaches for QA and semantic search [4, 24]. Currently, research is mostly performed on questions that can be answered using popular entity-centric knowledge graphs such as DBpedia. With the provision of EventKG, it will become possible to train QA approaches for event-related questions, e.g. “Which events related to Bill Clinton happened in Washington in 1980?” and ranking-based questions, e.g. “What are the most important events related to Syrian Civil War that took place in Aleppo?”.

*Relevance for Timeline Generation Applications:* Timeline generation is an active research area [1], where the focus is to generate a timeline (i.e. a chronologically ordered selection) of events and temporal relations for entities from a knowledge graph. EventKG can facilitate the generation of detailed timelines containing complementary information originating from different sources, potentially resulting in more complete timelines and event representations. For example, Table 1 illustrates an excerpt from the timeline for the query “What were the sub-events of the World War II between February 12 and February 28, 1941?” generated using EventKG. The first event in the timeline in Table 1 (“Erwin Rommel arrives in Tripoli”) extracted from an English Wikipedia event list (“1941 in Germany”) is not contained in any of the reference knowledge graphs used to populate EventKG (Wikidata, DBpedia, and YAGO). The reference sources of the other three events include complementary information. For example, while the “Action of 27 February 1941” is assigned a start date in Wikidata, it is not connected to the World War II in that source.

*Assessing Event Popularity and Relation Strength in Cross-Cultural Event-Centric Analytics:* Event popularity and relation strength between events and entities vary across different cultural and linguistic contexts. For example, Table 2 presents the top-4 most popular events in the Russian vs. the English Wikipedia language editions as measured by how often these events are linked

**Table 2.** Most linked events in the English (en) and the Russian (ru) Wikipedia.

Rank	Event (en)	#Links (en)	Event (ru)	#Links (ru)
1	World War II	189,716	World War II	25,295
2	World War I	99,079	World War I	22,038
3	American Civil War	37,672	October Revolution	7,533
4	FA Cup	20,640	Russian Civil War	7,093

**Table 3.** Top-3 persons mentioned jointly with the World War II per language.

	fr	de	ru	pt
1	Adolf Hitler	Adolf Hitler	Adolf Hitler	Adolf Hitler
2	Charles de Gaulle	Winston Churchill	Franklin D. Roosevelt	Getúlio Vargas
3	Winston Churchill	Franklin D. Roosevelt	Joseph Stalin	Joseph Stalin

to in the respective Wikipedia edition. Whereas both Wikipedia language editions mention events of global importance, here the two World Wars, most frequently, other most popular events (e.g. “October Revolution” and “American Civil War”) are language-specific. The relation strength between events and entities in specific language contexts can be induced by counting their joint mentions in Wikipedia. For example, Table 3 lists the persons most related to the World War II in different language editions. Information regarding event popularity and relation strength can enable the selection of the most relevant timeline entries given the layout constraints (e.g. EventKG contains 2,816 sub-events of the World War II). An EventKG application to cross-lingual timeline generation is presented in [10]. EventKG-empowered interfaces can be used as a starting point to identify controversial events for more detailed analysis using tools such as MultiWiki [9].

*Impact in Supporting the Adoption of Semantic Web Technologies:* EventKG follows best practices in data publishing and relies on open data and W3C standards to make the data reusable for a variety of real-world applications. We believe that researchers using EventKG outside the Semantic Web community, e.g. in the fields of NLP and Digital Humanities, will profit from the adoption of the W3C standards such as RDF, SPARQL and re-use of established vocabularies, thus stimulating adoption of Semantic Web technologies, e.g. in the context of Information Extraction, media analytics and cross-cultural studies.

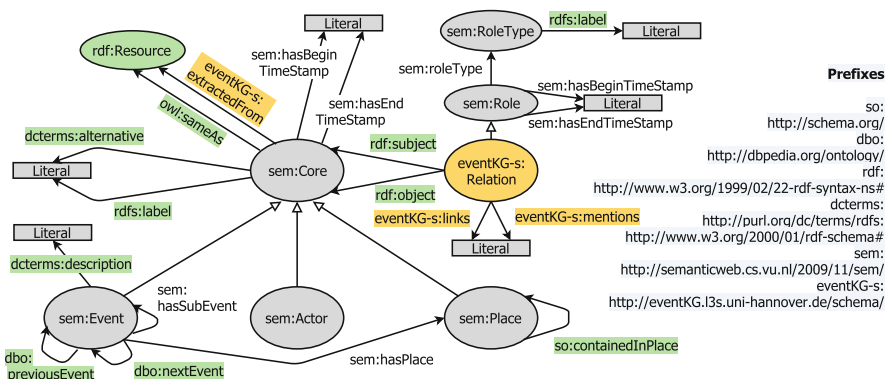
### 3 EventKG Data Model

*The goal of the EventKG data model* is to facilitate a light-weight integration and fusion of heterogeneous event representations and temporal relations extracted

from the reference sources, and make this information available to real-world applications. The EventKG data model is driven by the following goals:

- Define the key properties of events through a canonical representation.
- Represent temporal relations between events and entities (including event-entity, entity-event and entity-entity relations).
- Include information quantifying and further describing these relations.
- Represent relations between events (e.g. in the context of event series).
- Support an efficient light-weight integration of event representations and temporal relations originating from heterogeneous sources.
- Provide provenance for the information included in EventKG.

*EventKG Schema and the Simple Event Model:* In EventKG, we build upon the Simple Event Model (SEM) [23] as a basis to model events. SEM is a flexible data model that provides a generic event-centric framework. Within the EventKG schema (namespace `eventKG-s`<sup>1</sup>), we adopt additional properties and classes to adequately represent the information extracted from the reference sources, to model temporal relations and event relations as well as to provide provenance information. The schema of EventKG is presented in Fig. 1.



**Fig. 1.** The EventKG schema based on SEM. Arrows with an open head denote `rdfs:subClassOf` properties. Regular arrows visualize the `rdfs:domain` and `rdfs:range` restrictions on properties. Terms from other reused vocabularies are colored green. Classes and properties introduced in EventKG are colored orange. (Color figure online)

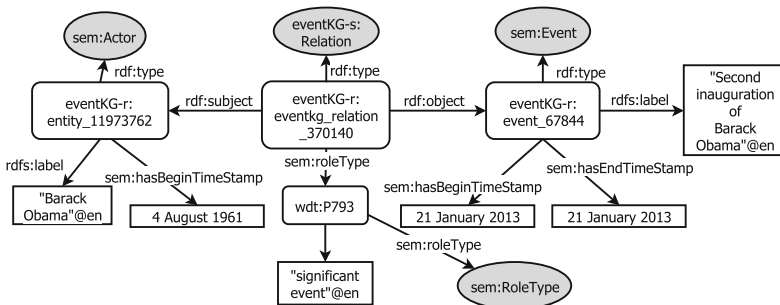
*Events and Entities:* SEM provides a generic event representation including topical, geographical and temporal dimensions of an event, as well as links to its actors (i.e. entities participating in the event). Such resources are identified within the namespace `eventKG-r`<sup>2</sup>. Thus, the key classes of SEM and of the

<sup>1</sup> <http://eventkg.l3s.uni-hannover.de/schema/>.

<sup>2</sup> <http://eventkg.l3s.uni-hannover.de/resource/>.

EventKG schema are `sem:Event` representing events, `sem:Place` representing locations and `sem:Actor` to represent entities participating in events. Each of these classes is a subclass of `sem:Core`, which is used to represent all entities in EventKG. (Note that entities in EventKG are not necessarily actors in the events; temporal relations between two entities are also possible). Events are connected to their locations through the `sem:hasPlace` property. A `sem:Core` instance can be assigned an existence time denoted via `sem:hasBeginTimeStamp` and `sem:hasEndTimeStamp`. In addition to the SEM representation, EventKG provides textual information regarding events and entities extracted from the reference sources including labels (`rdfs:label`), aliases (`dcterms:alternative`) and descriptions of events (`dcterms:description`).

*Temporal relations* are relations valid over a certain time period. In EventKG, they include event-entity, entity-event and entity-entity relations. Temporal relations between events and entities typically connect an event and its actors (as in SEM). A typical example of a temporal relation between two entities is a marriage. Temporal relations between entities can also indirectly capture information about events [19]. For example, the DBpedia property <http://dbpedia.org/property/acquired> can be used to represent an event of acquisition of one company by another. Temporal relations in SEM are limited to the situation where an actor plays a specific role in the context of an event. This yields two limitations: (i) there is no possibility to model temporal relations between events and entities where the entity acts as a subject. For example, it is not possible to directly model the fact that “Barack Obama” participated in the event “Second inauguration of Barack Obama”, as the entity “Barack Obama” plays the subject role in this relation; and (ii) a temporal relation between two entities such as a marriage can not be modeled directly. To overcome these limitations, EventKG introduces the class `eventKG-s:Relation` that links two `sem:Core` instances (each representing an event or an entity). This relation can be annotated with a validity time and a property `sem:RoleType` that characterizes the relation. This way, arbitrary temporal relations between entity pairs or relations involving an entity and an event can be represented. Figure 2 visualizes the example mentioned above using the EventKG data model.



**Fig. 2.** Example of the event representing the participation of Barack Obama in his second inauguration as a US president in 2013 as modeled in EventKG.

*Relations with Indirect Temporal Information:* The temporal validity of a relation is not always explicitly provided, but can often be estimated based on the existence times of the participating entities or events. For example, the validity of a “mother” relation can be determined using the birth date of the child entity. Therefore, in addition to temporal relations with known validity times, EventKG also includes relations connected to events as well as relations connected to entities as long as the existence time of both entities is provided.

*Other Event and Entity Relations:* Relations between events (in particular sub-event, previous and next event relations) play an important role in the context of event series (e.g. “Summer Olympics”), seasons containing a number of related events (e.g. in sports), or events related to a certain topic (e.g. operations in a military conflict). Sub-event relations are modeled using the `so:hasSubEvent` property. To interlink events within an event series such as the sequence of Olympic Games, the properties `dbo:previousEvent` and `dbo:nextEvent` are used. A location hierarchy is provided through the property `so:containedInPlace`.

*Towards Measuring Relation Strength and Event Popularity:* Measuring relation strength between events and entities and event popularity enables answering question like “Who was the most important participant of the event  $e$  ?” or “What are the most popular events related to  $e$  ?”. We include two relevant factors in the EventKG schema:

1. *Links:* This factor represents how often the description of one entity refers to another entity. Intuitively, this factor can be used to estimate the popularity of the events and the strength of their relations. In EventKG the links factor is represented through the predicate `eventKG-s:links` in the domain of `eventKG-s:Relation`. `eventKG-s:links` denotes how often the Wikipedia article representing the relation subject links to the entity representing the object.
2. *Mentions:* `eventKG-s:mentions` represents the number of relation mentions in external sources. Intuitively, this factor can be used to estimate the relation strength. In EventKG, `eventKG-s:mentions` denotes the number of sentences in Wikipedia that mention both, the subject and the object of the relation.

*Provenance Information:* EventKG provides the following provenance information: (i) provenance of the individual resources; (ii) representation of the reference sources; and (iii) provenance of statements.

*Provenance of the Individual Resources:* EventKG resources typically directly correspond to the events and entities contained in the reference sources (e.g. an entity representing Barack Obama in EventKG corresponds to the DBpedia resource <http://dbpedia.org/page/Barack.Obama>). In this case, the `owl:sameAs` property is used to interlink both resources. EventKG resources can also be extracted from a resource collection. For example, philosophy events in 2007 can be extracted from the Wikipedia event list at [https://en.wikipedia.org/wiki/2007\\_in\\_philosophy](https://en.wikipedia.org/wiki/2007_in_philosophy). In this case, the EventKG property



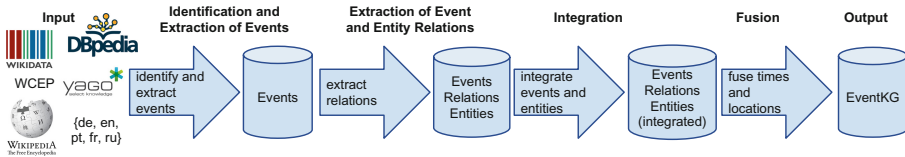
`eventKG-s:extractedFrom` is utilized to establish the link between the EventKG resource and the resource collection from which it was extracted. Through the provenance URIs, background knowledge contained in the reference sources can be accessed.

*Representation of the Reference Sources:* EventKG and each of the reference sources are represented through an instance of `void:Dataset`<sup>3</sup>. Such an instance in the namespace `eventKG-g`<sup>4</sup> includes specific properties of the source (e.g. its creation date).

*Provenance Information of Statements:* A statement in EventKG is represented as a quadruple, containing a triple and a URI of the named graph it belongs to. Through named graphs, EventKG offers an intuitive way to retrieve information extracted from the individual reference sources using SPARQL queries.

## 4 EventKG Generation Pipeline

The EventKG generation pipeline is shown in Fig. 3.



**Fig. 3.** EventKG generation pipeline.

*Input:* First, the dumps of the reference sources are collected.

*Identification and Extraction of Events:* Event instances are identified in the reference sources and extracted, as follows:

*Step Ia: Identification and extraction of events.*

- **Wikidata** [6]: We identify events as subclasses of Wikidata’s “event” and “occurrence”. The “occurrence” instances are added to increase recall. Some of the identified subclasses are blacklisted manually.
- **DBpedia** [16]: For each language edition, we identify DBpedia events as instances of `dbo:Event` or its subclasses.

<sup>3</sup> The VoID vocabulary <https://www.w3.org/TR/void/>.

<sup>4</sup> <http://eventkg.l3s.uni-hannover.de/graph/>.

- **YAGO** [17]: We do not use the YAGO ontology for event identification due to the noisy event subcategories (e.g. **event** > **act** > **activity** > **protection** > **self-defense** > **martial.art**). YAGO events are identified in Step Ib.
- **Wikipedia Event Lists**: For each language, we extract events from the Wikipedia event lists whose titles contain temporal expressions, such as “2007 in Science” and “August 11”, using methods similar to [13].
- **WCEP**: In the Wikipedia Current Events Portal, events are represented through rather brief textual descriptions and refer to daily happenings. We extract WCEP events using the WikiTimes interface [22].

We manually evaluated a random sample of the events identified in this step in DBpedia and Wikidata including 100 events per KG and language edition, achieving precision of 98% on average.

*Step Ib: Using Additional Event Identification Heuristics to Increase Recall:* First, we propagate the information regarding the identified events across the reference sources using existing `owl:sameAs` links. Second, we use Wikipedia category names that match a manually defined language-dependent regular expression (e.g. English category names that end with “events”) as an indication that a KG entry linked to such an article is an event. We manually evaluated this heuristic on a random sample of 100 events linked to the English and the Russian Wikipedia, respectively, achieving 94% and 88% precision.

In EventKG V1.1, we do not explicitly distinguish between single events such as “Solar eclipse of August 10, 1915”, seasons with a number of related events such as “2008 Emperor’s Cup” and event series like “Mario Marathon”.

*Extraction of Event and Entity Relations:* We extract the following types of relations: (1) *Temporal relations* are identified based on the availability of temporal validity information. Temporal relations are extracted from YAGO and Wikidata, as DBpedia does not provide such information. (2) *Relations with indirect temporal information:* we extract all relations involving events as well as relations of entities with known existence time. (3) *Other event and entity relations:* we use a manually defined mapping table to identify predicates that represent event relations in EventKG such as `so:hasSubEvent` (e.g. we map Wikidata’s `part of` property (P361) to `so:hasSubEvent` in cases where the property is used to connect events), `dbo:previousEvent` and `dbo:nextEvent` as well as `so:containedInPlace` to extract location hierarchies. We extract information that quantifies relation strength and event popularity based on the Wikipedia interlinking for each pair of interlinked entities containing at least one event. Entities are extracted only if they participate in an extracted relation.

*Integration:* The statements extracted from the reference sources are included in the named graphs, each named graph corresponding to a reference source. In addition, we create a named graph `eventKG-g:event_kg`. Each `sem:Event` and `sem:Core` instance in the `eventKG-g:event_kg` integrates event-centric and entity-centric information from the reference sources related to equivalent real-world instances. For the instances extracted from the KGs, known `owl:sameAs`

links are used. Events extracted from the semi-structured sources are integrated using a rule-based approach based on descriptions, times and links.

*Fusion:* In the fusion step, we aggregate temporal, spatial and type information of `eventKG-g:event_kg` events using a rule-based approach. *Location fusion:* For each event in `eventKG-g:event_kg`, we take the union of its locations from the different reference sources and exploit the `so:containedInPlace` relations to reduce this set to the minimum (e.g. the set {Paris, France, Lyon} is reduced to {Paris, Lyon}). *Time fusion:* For each entity, event or relation with a known existence or a validity time stamp, the integration is done using the following rules: (i) ignore the dates at the beginning or end of a time unit (e.g. January, 1st), if alternative dates are available; (ii) apply a majority voting among the reference sources; (iii) take the time stamp from the trusted source (in order: Wikidata, DBpedia, Wikipedia, WCEP, YAGO).

*Type fusion:* We provide `rdf:type` information according to the DBpedia ontology (dbo), using types and `owl:sameAs` links in the reference sources.

*Output:* Finally, extracted instances and relations are represented in RDF according to the EventKG data model (see Sect. 3). As described above, the information extracted from each reference source and the results of the fusion step are provided in separate named graphs.

## 5 EventKG Characteristics

In EventKG V1.1, we extracted event representations and relations in five languages from the latest available versions of each reference source as of 12/2017. Table 4 summarizes selected statistics from the EventKG V1.1, released in 03/2018. Overall, this version provides information for over 690 thousand events and over 2.3 million temporal relations. Nearly half of the events (46.75%) originate from the existing KGs; the other half (53.25%) is extracted from semi-structured sources. The data quality in the individual named graphs directly corresponds to the quality of the reference sources. In `eventKG-g:event_kg`, the majority of the events (76.21%) possess a known start or end time. Locations are provided for 12.21% of the events. The coverage of locations can be further increased in the future work, e.g. using NLP techniques to extract locations from the event descriptions. Along with over 2.3 million temporal relations, EventKG V1.1 includes relations between events and entities for which the time is not available. This results in overall over 88 million relations. Approximately half of these relations possess interlinking information.

### 5.1 Comparison of EventKG to its Reference Sources

We compare EventKG to its reference sources in terms of the number of the identified events and completeness of their representation. The results of the event identification Step Ia are shown in Table 5. EventKG with 690,247 events

**Table 4.** Number of events and relations in eventKG-g:event\_kg.

	#Events	Known time	Known location
Events from KGs	322,669	163,977	84,304
Events from semi-structured sources	367,578	362,064	Not extracted
Relations	88,473,111	2,331,370	Not extracted

**Table 5.** Number of events extracted from the reference sources (Step Ia).

Wikidata	DBpedia					Wikipedia event lists					WCEP
	en	fr	de	ru	pt	en	fr	de	ru	pt	
266,198	60,307	43,495	9,383	5,730	14,641	131,774	110,879	21,191	44,025	18,792	61,382

contains a significantly higher number of events than any of its reference sources. This is especially due to the integration of KGs and semi-structured sources.

Table 6 presents a comparison of the event representations in EventKG and its reference knowledge graphs (Wikidata, YAGO, DBpedia). As we can observe, through the integration of event-centric information, EventKG: (1) enables better event identification (e.g. we can map 322,669 events from EventKG to Wikidata, whereas only 266,198 were identified as events in Wikidata initially - see Table 5), and (2) provides more complete event representations (i.e. EventKG provides a higher percentage of events with specified temporal and spatial information compared to Wikidata, that is the most complete reference sources). The most frequent event types are source-dependent (see Table 7).

## 5.2 Relation and Fusion Statistics

Over 2.3 million temporal relations are an essential part of EventKG. The majority of the frequent predicates in EventKG such as “member of sports team” (882,398 relations), “heritage designation” (221,472), “award received” (128,125), and “position held” (105,333) originate from Wikidata. The biggest fraction of YAGO’s temporal relations have the predicate “plays for” (492,263), referring to football players. Other YAGO predicates such as “has won prize” are less frequent. Overall, about 93.62% of the temporal relations have a start time from 1900 to 2020. 81.75% of events extracted from KGs are covered by multiple sources. At the fusion step, we observed that 93.79% of the events that have a known start time agree on the start times across the different sources.

## 5.3 Textual Descriptions

EventKG V1.1 contains information in five languages. Overall, 87.65% of the events extracted from KGs provide an English label whereas only a small fraction (4.49%) provide labels in all languages. Among the 367,578 events extracted

**Table 6.** Comparison of the event representation completeness in the source-specific named graphs (after the Step Ib).

	EventKG	Wikidata	YAGO	DBpedia				
				en	fr	de	ru	pt
#Events with	322,669	322,669	222,325	214,556	78,527	62,971	47,304	35,682
Location (L)	26.13%	11.70%	26.61%	6.21%	8.32%	4.03%	10.60%	6.15%
Time (T)	50.82%	33.00%	39.02%	7.00%	17.21%	2.00%	1.35%	0.08%
L&T	21.97%	8.83%	19.02 %	4.29%	0.00%	4.84%	1.18%	0.08%

**Table 7.** The most frequent event types extracted from the references sources and the percentage of the events in that source with the respective type.

	Wikidata	DBpedia				
		en	fr	de	ru	pt
dbo:type	Season	Military conflict	Sports event	Tennis tournament	Military conflict	Soccer tournament
Events, %	11.37%	6.31%	21.86%	33.00%	11.87%	16.17%

from the semi-structured sources, just 115 provide a description in all five languages, e.g. the first launch of a Space Shuttle in 1981. This indicates potential for further enrichment of multilingual event descriptions in future work.

## 6 Reusability Aspects

In order to facilitate an efficient reuse of EventKG, we provide the resource for download, as well as through a SPARQL endpoint. The homepage of EventKG provides a comprehensive documentation of the resource including example queries. A schema diagram of EventKG is presented in Fig. 1. EventKG is modeled in RDF and is highly extensible. For example, it is possible to include further languages and customize the selection of the reference data sources. Recent studies indicate that interlinking is an important factor of dataset reuse [5]. To this extent, EventKG provides substantial interlinking with its reference sources.

At the moment, the intended use of EventKG includes collaboration with EU-projects such as ALEXANDRIA (for enrichment of Web Archives with event-centric data)<sup>5</sup>, and WDAqua ITN<sup>6</sup>, in the context of innovative event-centric Question Answering applications. We believe that due to its unique nature and general applicability, EventKG will be widely reused by third parties in a number of communities in the future, as also discussed in Sect. 2.

<sup>5</sup> <http://alexandria-project.eu/>.

<sup>6</sup> WDAqua (Answering Questions using Web Data) <http://wdaqua.eu/>.

EventKG follows best practices in data publishing. It uses RDF W3C standard to model and interlink the data it contains. EventKG adopts the open data and open source approach to make it available to a wide audience and facilitates data and software reuse. EventKG supports multilinguality of the data, provides dereferencable URIs and implements a persistent strategy to maintain its URIs across the versions, ensuring that the same URIs are consistently reused for the same real-world objects.

EventKG reuses and extends an established event model, which is SEM [23] to describe event-related information it includes, and reuses existing vocabularies (e.g. DBpedia ontology, Dublin Core). The EventKG metadata is provided using the VoID<sup>7</sup> vocabulary. EventKG follows FAIR principles<sup>8</sup> to make it findable, accessible, interoperable and reusable. The EventKG description is available in human and machine readable formats at the EventKG homepage.

## 7 Availability and Sustainability

*Availability Aspects:* EventKG uses open standards and is publicly available under a persistent URI<sup>9</sup> under the CC BY 4.0 license<sup>10</sup>. The EventKG homepage<sup>11</sup> provides information on citing the resource. Our extraction pipeline is available as open source software on GitHub<sup>12</sup> under the MIT License<sup>13</sup>.

*Sustainability Plan:* The sustainability of EventKG is ensured through three building blocks: (1) *Open source architecture and software:* The software developed for the creation of EventKG is available as open source and can be re-used by the community to extract a new version of the knowledge graph, or extend the resource to include more reference sources, languages, or event properties. (2) *Integration of existing publicly available data:* The reference sources that serve as a basis for the data within the EventKG are publicly available and many of them are maintained by the community, so that it is possible to maintain a fresh version of the resource, in particular to include new events. (3) *Maintenance of EventKG:* The authors plan to perform regular EventKG updates. The URIs of EventKG resources will be maintained and remain stable across versions.

## 8 Related Work

*Data Models and Vocabularies for Events:* Several data models and the corresponding vocabularies (e.g. [12, 19, 20, 23]) provide means to model events. For example, the ECKG model proposed by Rospocher et al. [19] enables fine-grained

<sup>7</sup> <https://www.w3.org/TR/void/>.

<sup>8</sup> <https://www.force11.org/group/fairgroup/fairprinciples>.

<sup>9</sup> <https://doi.org/10.5281/zenodo.1112283>.

<sup>10</sup> <https://creativecommons.org/licenses/by/4.0/>.

<sup>11</sup> <http://eventKG.l3s.uni-hannover.de>.

<sup>12</sup> <https://github.com/sgottsch/eventkg>.

<sup>13</sup> <https://opensource.org/licenses/MIT>.

textual annotations to model events extracted from news collections. The Simple Event Model (SEM) [23], schema.org [12] and the Linking Open Descriptions of Events (LODE) ontology [20] provide means to describe events and interlink them with actors, times and places. In EventKG, we build upon SEM and extend this model to represent a wider range of temporal relations and to provide additional information regarding events.

*Extracting Event-Centric Information:* Most approaches for automatic knowledge graph construction and integration focus on entities and related facts rather than events. Examples include DBpedia [16], Freebase [2], YAGO [17] and YAGO+F [3]. In contrast, EventKG is focused on events and temporal relations. In [22], the authors extract event information from WCEP. EventKG builds upon this work to include WCEP events.

*Extraction of Events and Facts from News:* Recently, the problem of building knowledge graphs directly from plain text news [19], and extraction of named events from news [15] have been addressed. These approaches apply Open Information Extraction methods and develop them further to address specific challenges in the event extraction in the news domain. State-of-the-art works that automatically extract events from news potentially obtain noisy and unreliable results (e.g. the state-of-the-art extraction approach in [19] reports an accuracy of only 0.551). In contrast, contemporary events included in EventKG originate from manually curated sources such as WCEP and Wikipedia event lists.

## 9 Conclusion

In this paper we presented EventKG – a multilingual knowledge graph that integrates and harmonizes event-centric and temporal information regarding historical and contemporary events. EventKG V1.1 includes over 690 thousand event resources and over 2.3 million temporal relations. Unique features of EventKG include light-weight integration and fusion of structured and semi-structured multilingual event representations and temporal relations in a single knowledge graph, as well as the provision of information to facilitate assessment of relation strength and event popularity, while providing provenance. The light-weight integration enables to significantly increase the coverage and completeness of the included event representations, in particular with respect to times and locations.

**Acknowledgements.** This work was partially funded by the ERC (“ALEXANDRIA”, 339233) and BMBF (“Data4UrbanMobility”, 02K15A040).

## References

1. Althoff, T., Dong, X.L., Murphy, K., Alai, S., Dang, V., Zhang, W.: TimeMachine: timeline generation for knowledge-base entities. In: Proceedings of SIGKDD 2015 (2015)
2. Bollacker, K., et al.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of SIGMOD 2008, pp. 1247–1250 (2008)

3. Demidova, E., Oelze, I., Nejdil, W.: Aligning freebase with the YAGO ontology. In: Proceedings of CIKM 2013, pp. 579–588. ACM (2013)
4. Demidova, E., Zhou, X., Nejdil, W.: Efficient query construction for large scale data. In: Proceedings of SIGIR 2013, pp. 573–582 (2013)
5. Endris, K.M., Giménez-García, J.M., Thakkar, H., Demidova, E., Zimmermann, A., Lange, C., Simperl, E.: Dataset reuse: an analysis of references in community discussions, publications and data. In: Proceedings of the K-CAP 2017, pp. 5:1–5:4 (2017)
6. Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., Vrandečić, D.: Introducing Wikidata to the linked data web. In: Mika, P., et al. (eds.) ISWC 2014. LNCS, vol. 8796, pp. 50–65. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-11964-9\\_4](https://doi.org/10.1007/978-3-319-11964-9_4)
7. Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semant. Web* 1–53 (2016)
8. Gossen, G., Demidova, E., Risse, T.: iCrawl: improving the freshness of web collections by integrating social web and focused web crawling. In: JCDL 2015 (2015)
9. Gottschalk, S., Demidova, E.: MultiWiki: interlingual text passage alignment in Wikipedia. *TWEB* 11(1), 6:1–6:30 (2017)
10. Gottschalk, S., Demidova, E.: EventKG+TL: creating cross-lingual timelines from an event-centric knowledge graph (2018, inpress)
11. Gottschalk, S., Demidova, E., Bernacchi, V., Rogers, R.: Ongoing events in Wikipedia: a cross-lingual case study. In: Proceedings of WebSci 2017, pp. 387–388 (2017)
12. Guha, R.: Introducing schema.org: Search engines come together for a richer web. Google Official Blog (2011)
13. Hienert, D., Wegener, D., Paulheim, H.: Automatic classification and relationship extraction for multi-lingual and multi-granular events from Wikipedia. In: Proceedings of DeRiVe 2012 (2012)
14. Höffner, K., et al.: Survey on challenges of question answering in the semantic web. *Semant. Web* 8(6), 895–920 (2017)
15. Kuzey, E., Vreeken, J., Weikum, G.: A fresh look on knowledge bases: distilling named events from news. In: Proceedings of CIKM 2014, pp. 1689–1698 (2014)
16. Lehmann, J., et al.: DBpedia - a large-scale, multilingual knowledge base extracted from Wikipedia. *Semant. Web J.* 6(2), 167–195 (2015)
17. Mahdisoltani, F., Biega, J., Suchanek, F.: YAGO3: a knowledge base from multilingual Wikipedias. In: Proceedings of CIDR 2014 (2014)
18. Rogers, R.: *Digital Methods*. MIT Press, Cambridge (2013)
19. Rospocher, M., et al.: Building event-centric knowledge graphs from news. *Web Semant.* 37, 132–151 (2016)
20. Shaw, R., Troncy, R., Hardman, L.: LODÉ: 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](https://doi.org/10.1007/978-3-642-10871-6_11)
21. Tempelmeier, N., Demidova, E., Dietze, S.: Inferring missing categorical information in noisy and sparse web markup. In: Proceedings of WWW 2018 (2018)
22. Tran, G.B., Alrifai, M.: Indexing and analyzing Wikipedia’s current events portal, the daily news summaries by the crowd. In: Proceedings of WWW 2014 (2014)
23. Van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the simple event model (SEM). *Web Semant.* 9(2), 128–136 (2011)
24. Zheng, W., Cheng, H., Zou, L., Yu, J.X., Zhao, K.: Natural language question/answering: let users talk with the knowledge graph. In: CIKM 2017 (2017)