Global persistent identifiers for conferences and Crossmark for conference proceedings.

Authors

This document stems from the ideas discussed in the Crossref and DataCite working group on conference identifiers and project identifiers. Here is the list of people involved in the group:

1. Marcel R. Ackermann, Schloss Dagstuhl / DBLP
2. Aliaksandr Birukou (chair), Springer Nature
3. Geoffrey Bilder, Crossref
4. Nhora Cortes-Comerer, ASME
5. Philip DiVietro, ASME
6. Martin Fenner, DataCite
7. Gerry Grenier, IEEE
8. Christina Hoppermann, Springer Nature
9. Bethan Keall, Elsevier
10. Anna Lacson, ACM
11. Andres Mori, Digital Science
12. Craig Rodkin, ACM
13. Sweitze Roffel, Elsevier
14. Judy Salk, Elsevier
15. Kruna Vukmirovic, The IET
16. Alexander Wagner, DESY

The document has also benefited from comments by …

1. Volha Bryl
2. Maral Dadvar
3. Fleur van Dijl
4. Jan Dvorak
5. Kai Eckert
6. Chris Erdmann
7. Pedro Ferreira
8. Kevin Hawkins
9. Matt Hudson
10. Andrea Mannocci
11. Tasha Mellins-Cohen
12. Mohamad Mostafa, Think Check Attend
13. Natasha Noy

14. Martin Rasmussen
15. Angelo Antonio Salatino
16. John Samuel
17. Rusty Speidel
18. Lars G. Svensson

Introduction

The purpose of this document is to define metadata for conferences and conference series, so that it can be better structured, used for analytics, and linked to the metadata about conference proceedings already stored in Crossref[1].

Metadata documentation

Availability

All conference and conference series metadata will be available under CC0 license.

Basic definitions

Conference series - a series of conferences taking place regularly. For instance, "International Semantic Web Conference" is a conference series. Note that we do not distinguish here among conferences, workshops, symposia, colloquiums, as long as the event has an established identity and takes place (or plans to) regularly.

Conference (instance) - a specific conference (workshop, symposium, etc.) in the series. For instance, "The 16th International Semantic Web Conference".

Colocated conferences - two or more conferences sharing the conference venue but holding separate programs that are open to all attendees. This includes back-to-back conferences, where the attendees can register for one or more of the conferences, which might share the program.

Joint conferences - two or more conferences share the conference venue and also integrate the conference program, which is also open to all attendees.

Umbrella conference - a conference which hosts several conferences. For instance, ETAPS 2018 is an umbrella conference for FOSSACS 2018, TACAS 2018, POST 2018, etc....

---

[1] see http://data.crossref.org/reports/help/schema_doc/4.4.1/schema_4_4_1.html#conference

Umbrella conference series - a conference series linking several other series (e.g., XXX conference which consists of the XXX winter and XXX summer series).

Conference series metadata

A conference series is characterised by the following properties:

Property	Description	Required?	Cardinality	Example	Comments
confSeriesID	The unique ID of the conference series	Yes	1..1	semweb	This might be a DOI or some other ID, currently it is alphanumeric. It stays the same even if the series name changes, to make sure we can trace the conference series history over time.
confSeriesName	The name of the conference series	Yes	1..1	International Semantic Web Conference	
confSeriesAcronym	The acronym of the conference series	No	1..1	ISWC	Not all conferences have an acronym, so we should not force it
OtherIDs	Other IDs in ID_Type + ID_Value format	No	1..n	ID_Type="DBLP", ID_Value="https://dblp.org/db/conf/semweb" ID_Type="ASME", ID_Value="ES"	
ConfSeriesURL	The URL of the conference series	No	1..n	http://swsa.semanticweb.org/content/international-semantic-web-conference-iswc	
ConfSeriesMaintainer	The entity responsible for the conference series, e.g., the society, the steering committee.	No	Most likely 1..n	Semantic Web Science Association, SWSA	Could be a set of people with affiliations or an organization identifier.

Additional information

1. At the implementation level, OtherIDs and ConfSeriesMaintainer can be specified as relations between the conference series ID and other entities (e.g., extra tables, classes)
2. Changes in the conference series name, acronym or the conference series maintainer can be related via "successor-predecessor" relations, which are described in the following sections. Successor and predecessor are inverse relations, i.e. if A is a predecessor of B then B is a successor of A.
3. At the implementation level, it is not necessary to have both successor and predecessor field, just having predecessor(s) for s enough, as then one can always deduce successors. If the conference series is new, it should have an empty predecessor.

Conference instance metadata

A conference instance is characterised by the following fields. Note that we do not need to duplicate the conference acronym and name here, as their changes lead to the changes of the confSeriesID.

Property	Description	Required?	Cardinality	Value	Comments
confSeriesID	The ID of the conference series	Yes	1..1	semweb	We assume every conference is a part of a series, even if it is a series of one conference.
confID	The ID of the conference	Yes	1..1	long_and_mnemonic_identifier_12354	This might be a DOI or some other ID, currently it is alphanumeric
confSeriesName	The name of the conference series	Yes	1..1	International Semantic Web Workshop	The snapshot of the current conference series name
confSeriesAcronym	The acronym of the conference series	Yes	1..1	ISWW	The snapshot of the current conference series acronym

confNumber	The number of the conference in the conference series.	No	1..1	13	
StartDate	The start date of the conference.	No	1..1	2014-10-19	The dates must be in the ISO 8601 format: https://en.wikipedia.org/wiki/ISO_8601 You can do year only or year-month OR XSD datatypes[2]
EndDate	The end date of the conference	No	1..1	2014-10-23	See above
City	The city in which the conference was held	No	1..n	Riva del Garda	We will use an existing vocabulary here, like http://www.getty.edu/research/tools/vocabularies/tgn/index.html
State	The state, if the conference was held in the USA, Australia, Canada. Could be dropped if we use the	No	1..n	NC	To use a standard list of states

---

[2] Alternatives (to consider by the technical implementation group) include the W3C Location vocabulary (https://www.w3.org/ns/locn) that ties in nicely with INSPIRE metadata or Dublin Core Date Property. For W3C Location, the combination of year-month is xsd:gYearMonth and a year alone is xsd:gYear. Cf. https://www.w3.org/TR/xmlschema-2/#gYearMonth and https://www.w3.org/TR/xmlschema-2/#gYear. I'd suggest to use XSD datatypes instead since those are not ambiguous in terms of e. g. hyphenation (ISO 8601 allows both 2018-06-13 and 20180613)

	thesaurus of locations				
Countr y	The country in which the conference was held	No	1..n	Italy	
Place	The location of the conference, represented in schema.org vocabulary. An alternative is to use the W3C Location vocabulary (https://www.w3.org/ns/locn) that ties in nicely with INSPIRE metadata?	No	1..n	"location":{"@type":"Place","address":"National Culture Palace, \nBoulevard \"Bulgaria\", 1463 NDK, Sofia, Bulgaria","name":"Sofia, Bulgaria"}	To be used together or as an alternative to City/State/Country. See http://schema.org/Place
URL	The URL of the conference website	No	1..1	http://iswc2014.semanticweb.org/	

Additional information

1. For the list of proceedings see Conference Proceedings section
2. A conference can occur in multiple locations, e.g., " IFIP Congress 1980: Tokyo, Japan & Melbourne, Australia", so we should support multiple locations
3. Additional metadata, which could be included in the future:
   a. Sponsors - as usually conferences have Sponsors especially with industry-related fields such as Engineering & Biomedical research
   b. Organizing committee
   c. Keynote speakers (see more in the Scholarly Data ontology document: http://www.scholarlydata.org/ontology/doc/). We can also send such info to ORCID
   d. Venue details (University, Hotel etc.)
   e. Papers stats: submitted, accepted, rejected
   f. Conference topics
   g. Hashtag, links to social media accounts - keywords.

Proceedings metadata

The proceedings metadata, is defined by Crossref, see, for instance,
[http://data.crossref.org/reports/help/schema_doc/4.4.1/schema_4_4_1.html#proceedings_meta](http://data.crossref.org/reports/help/schema_doc/4.4.1/schema_4_4_1.html#proceedings_metadata)
[data](http://data.crossref.org/reports/help/schema_doc/4.4.1/schema_4_4_1.html#proceedings_metadata)

Relations between conferences and conference series

Conference series evolution

In case conference series changes name, acronym, maintainer or merged with other conference series (or splits), it is modelled by the predecessor-successor relation.

Property	Description	Required?	Example	Comments
predecessor ConfSeriesID	The ID of the "old" conference series which has changed / was merged	Yes	atal	AAMAS is a merger of ATAL, AA and ICMAS conference series, so this relation will repeat 3 times - for atal (this example), AA and ICMAS
successorConfSeriesID	The ID of the "new" conference series	Yes	aamas	

Umbrella conferences

ETAPS 2018 is an umbrella conference, composed of ESOP 2018, FASE 2018, FoSSaCS 2018, POST 2018 and TACAS 2018. We can model the fact that ETAPS 2018 is an umbrella conference for these 5 conferences.

Colocated conferences

ESOP 2018 and TACAS 2018 have different PCs and different conference programs, therefore they are colocated.
Colocation has an implication on the Crossmark (see below) - each of the conferences should have separate information about the peer review process.

Similarly, for Euro-Par 2016 is conference, colocated with the following workshops:
1. Euro-EDUPAR (Second European Workshop on Parallel and Distributed Computing Education for Undergraduate Students)

2. HeteroPar 2016 (the 14th International Workshop on Algorithms, Models and Tools for Parallel Computing on Heterogeneous Platforms)
3. IWMSE (5th International Workshop on Multicore Software Engineering)
4. LSDVE (Fourth Workshop on Large-Scale Distributed Virtual Environments)
5. PADABS (Fourth Workshop on Parallel and Distributed Agent-Based Simulations)
6. PBio (Fourth International Workshop on Parallelism in Bioinformatics)
7. PELGA (Second Workshop on Performance Engineering for Large-Scale Graph Analytics)
8. REPPAR (Third International Workshop on Reproducibility in Parallel Computing)
9. Resilience (9th Workshop in Resilience in High Performance Computing in Clusters, Clouds, and Grids)
10. ROME (Fourth Workshop on Runtime and Operating Systems for the Many-Core Era)
11. UCHPC (9th Workshop on UnConventional High-Performance Computing).

Joint conferences

ASME 2017 Power Conference was held jointly with ICOPE-17:

ASME Power Conference (POWER)

ASME 2017 Power Conference Joint With ICOPE-17 collocated with the ASME 2017 11th International Conference on Energy Sustainability, the ASME 2017 15th International Conference on Fuel Cell Science, Engineering and Technology, and the ASME 2017 Nuclear Forum (POWER2017-ICOPE17)

Conference proceedings

For each conference we can link several proceedings, not necessarily published as conference proceedings, but also as journal special issues, edited books, etc.

Property	Description	Required?	Example	Comments
confID	The ID of the conference	Yes	ecml2017	ECML PKDD 2017
DOI	The DOI of the proceedings volumes / journal issue / individual articles / book	No	https://doi.org/10.1007/978-3-319-71249-9 https://doi.org/10.1007/978-3-319-71246-8 https://doi.org/10.1007/978-3-319-71273-4	Three books with conference proceedings and 2 special issues with the journal Track of ECML PKDD 2017 in the DMKD and ML journals.

URL	The URL of the proceedings / journal issue / individual articles / book	No	https://link.springer.com/journal/10618/31/5/page/1 https://link.springer.com/journal/10994/106/9/page/1	To be filled in case the DOI is not available

Additional Info

1. The Conference-Proceedings relation can be many-to-many. A Conference may have multiple proceedings volumes (e.g.: http://dblp.org/db/conf/eusflat/ 2017, Vol 1-3), a proceedings volume may combine multiple conference instances (e.g.: http://dblp.org/db/conf/fg/2015 + 2016).
2. We should support also URLs for proceedings, in case DOI is missing. In the example one can see that journal issues do not always have DOIs….Also, CEUR-WS proceedings do not have DOIs

Conference PCs

The list of ORCIDs and affiliations, using Crossref model for contributors, wherever possible.

Property	Description	Required?	Example	Comments
confID	The ID of the conference	Yes	ecml2017	ECML PKDD 2017
PCRole	The role of the person in the PC	No	pc_chair	We can use one of the standard ontologies like http://www.scholarlydata.org/ontology/doc/#d4e1265
ContributorID	The ID of the contributor. The contributor element in the Crossref schema contains person name, affiliation, role	Yes		See http://data.crossref.org/reports/help/schema_doc/4.4.1/schema_4_4_1.html#contributors For the roles we might need to extend the Crossref taxonomy of roles with PC Chair, PC member, etc. Alternatively, we can use PRO ontology from SPAR ontologies: http://www.sparontologies.net/ontologies/pro

Modeling granularity

In the example with Euro-Par 2016, it is up to the publisher/conference if they would like to model each individual workshop. We think it should be possible to model such workshops (conference instances), even retrospectively, if it is important for the community. We should not mandate modelling all possible conference instances, however.

Illustrating examples

Summer and Winter conference series

One can also define a "common" conference series which groups Summer and Winter conference series by using umbrella conference series relation. E.g., for http://dblp.org/db/conf/vtc/ one has:
1. VTC Spring conference series
2. VTC Fall conference series
3. VTC conference series - an umbrella conference for both VTC Spring and VTC Fall

Merger of several conference series

International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS, is a merger of three very successful conference series, AA (the International Conference on Autonomous Agents), ICMAS (the International Conference on Multi-Agent Systems), and ATAL (the International Workshop on Agent Theories, Architectures, and Languages).
In this case, we have three relations between the conference series:

Predecessor	Successor
aa	aamas
icmas	aamas
atal	aamas

Split of a conference

We are not aware of such an example, please contact us if you have one.

Limitations

In the current version of the metadata model we are not addressing several relations or properties, listed below.

1. "Also known as" - other names under which the conference is known in the community (formally or informally). Here is an example from Geoffrey:
   a. I've also seen conferences with a formal name and an informal name. For instance the (fictitious) "International Psychoceramics Conference" of 2016 was known as informally known as the "Beyond cracked pots" conference, while the "International Psychoceramics Conference" of 2017 was called "The pot calling the kettle black" conference.
   b. ….and Marcel: the ACM Symposium on Computational Geometry has been abbreviated "SCG" for years by its publisher, although the community almost always called it "SoCG". Only recently, the conference "formally" adopted that acronym.
2. Links to the Call for Papers (CfP), with structured information about topics[3] and submission deadlines, etc. This can support high level tasks such as classification and exploration.
3. Conference structure (tracks, program, etc). See, for instance the CfPs of the ISWC 2017: https://iswc2017.semanticweb.org/calls/ with multiple tracks: research, in-use, resource, industry, etc. Some of those track have separate PCs.
4. Data about peer review (see Crossmark for proceedings proposal below). In theory, each conference could have several "peer review processes" in the future, e.g., one for selecting research papers and another for industry track papers.
5. People around the conference
   a. Steering Committee
   b. Organizers
   c. Chairs
   d. Speakers

Other relevant examples

The lifecycle of a conference (series) is similar to that of other entities. Below we list a couple of such examples.

GND relations for organizations

Models both hierarchical relations (part-of) as well as temporal relations (predecessor/successor). For instance, http://d-nb.info/gnd/1102188522 http://d-nb.info/gnd/2029304-5 lists a superior organization ("Beziehungen zu Organisationen, Administrativ übergeordnet"), which organizes the conference and a predecessor conference ("Vorgänger").

---

[3] See http://2018.jcdl.org/topics_of_interest for an example

JuSER Institute collection

Lists authority data about names of institutes and their history. For instance, http://juser.fz-juelich.de/record/98641 has changed name from "Erdöl und Geochemie (ICG-4)" to "Agrosphäre (ICG-IV)" and then to "Agrosphäre (ICG-4)", currently being called "Agrosphäre (IBG-3)".

Microsoft Academic Knowledge API

Provides one with the acces to the data from the Microsoft Academic Graph (MAG), which contains also info about conference series and conference instances: https://docs.microsoft.com/en-us/azure/cognitive-services/academic-knowledge/conferenceinstanceentityattributes

Use cases

- Linking research outputs (presentations, videos, etc) to a conference without proceedings - like FORCE 2017 or PIDapalooza
- Reviewer recognition for ORCID, Publons, ...
- Computing metrics for a conference series - like Google Scholar
- Help in research evaluation - CORE, QUALIS, CCF, Danish conference ratings
- Recognizing fake conferences
- Assigning IDs in submission systems

Action items

Here we list the ideas suggested during the public commenting round we'd like to consider:
1. Compare our metadata with Indico metadata (Jose Benito Gonzalez Lopez)
2. Provide PID data to *"Think. Check. Attend"* https://thinkcheckattend.org/ (Mohamad Mostafa)
3. Use OSF Meetings data https://osf.io/meetings/
4. Use Microsoft Academic Knowledge Graph (not sure about the license): https://www.microsoft.com/en-us/research/project/academic/articles/april-2018-graph-update/

Crossmark for conference proceedings and papers

The purpose of Crossmark for conference proceedings is to structure and preserve the information about the peer review process of a conference as declared by the general or program chairs. Even if in principle, such info could be manipulated/faked by the chairs, this

would constitute a misconduct. In the future we can imagine such info coming directly from the submission systems.

Depending on how much information is available from the conference organizers, one can fill basic or extended version of Crossmark.

Crossmark basics

Label	Required?	Meaning	Example
peerReviewType	Yes	single blind, double blind, open, other	single blind
confManagementSystem	No	EasyChair, CMT, etc.	OCS
submissionsSentForReview	Yes	The number of papers sent for peer review. Does not include straightforward rejects by the PC chairs due to out-of-scope or other reasons	100
PapersAcceptedForPublication	Yes	The number of full papers accepted for publication.	30
accRatePerc	Yes	The number of papers/The number of submissions sent for review * 100	30
avgNumReviewsPerPaper	No	The number of reviews / the number of submissions sent for review.	3.25
additionalInfoOnReviewProcess	No	Any additional information provided about the peer review process by the organizers.	"Short papers underwent shepherding process and 5 out of 10 were accepted as full papers."2.

Crossmark extended

Label	Required?	Meaning	Example

peerReviewType	Yes	single blind, double blind, open, other	single blind
confManagementSystem	No	EasyChair, CMT, etc.	OCS
submissionsSentForReview	Yes	The number of papers sent for peer review. Does not include straightforward rejects by the PC chairs due to out-of-scope or other reasons	100
fullPapersAccepted	Yes	The number of full papers accepted.	30
shortPapersAccepted	No	The number of short papers accepted.	15
accRateFullPapersPerc	Yes	The number of full papers/The number of submissions sent for review * 100	30
avgNumReviewsPerPaper	No	The number of reviews / the number of submissions sent for review.	3.25
avgNumPapersPerReviewer	No	The number of papers each reviewer has to review on average.	5.5
extReviewersInvolved	No	Were external reviewers (people not included in the Program Committee) involved?	yes
additionalInfoOnReviewProcess	No	Any additional information provided about the peer review process by the organizers.	"Short papers underwent shepherding process and 5 out of 10 were accepted as full papers."2.

Open questions

1. We might want to specify that the review was only based on conference abstracts.
2. Currently we support only "full", "short" paper types, need to support "abstracts", may be "demo", "talk", "poster", "data", "software"

Revision history

2018-06-01. Replying to the community comments
1. Conference event metadata -> changed to conference instance metadata to be consistent with https://docs.microsoft.com/en-us/azure/cognitive-services/academic-knowledge/conferenceinstanceentityattributes and get rid of other conflicting meaning of the "event" - e.g., Nature/MacMillan Linked Open Data used to have "Events" referring to production events
2. Clarified when a new conference series is created (changes in the Acronym, Name or Maintainer)
3. Clarified that successor-predecessor are inverse relations
4. Introduced cardinality of properties for conference series and conferences

2018-01-08. First draft, based on the group documents created previously.
1. Conference metadata:
   a. TopLevelID renamed to ConferenceSeriesID
   b. Publications removed as they'll be handled separately in Conference Proceedings
   c. PC removed and added as relations - see Conference PCs
   d. Removed "ConfYear" as Start/End dates might only contain year if the complete date is not available
   e. Suggesting the use of Getty TGN
   f. Introduced the concept of relations for colocated, joint and umbrella conferences
2. Conference series:
   a. Introduced the concept of relations with predecessor and successor

2017-09-19.

Conference Series Metadata
1. ConfSeriesEntities is called "ConfSeriesMaintainer" now
2. DBLP ID has been removed, as we'd like support various IDs - from other publishers or WikiData IDs, etc. So, now there is "OtherIDs" field instead
3. **IMPORTANT - done after the call** Added "confSeriesUniqueID" as at the conference event level, we'd like to link to the unique combination of the conference acronym and conference series name (e.g., ACAL + "Australian Conference on Artificial Life and Computational Intelligence" or ACAL + "Australian Conference on Artificial Life"). This is

NOT possible if all conference series have the same ConfSeriesID. The latter, however, allows one to know that all such conference series belong together.

Conference Event Metadata

1. Removed ConfAcronym and ConfName as they are represented in the series
2. TopLevelID is now mandatory - in other words, we create a conference series always
3. Added a field for listing the PC members (a simple list of ORCIDs and affiliations + using Crossref model for authors). We can add organization identifiers later. See https://www.crossref.org/categories/organization-identifiers/
4. At the conference event level we can mode co-located and joint conferences. ECML - **succeeded_by** ECML/PKDD; PKDD - **succeeded_by** ECML/PKDD (multiple succeeded_by in case of split)