

ConfIDent – An Open Platform for FAIR Conference Metadata

Stephanie Hagemann-Wilholt, Christian Hauschke, and Margret Plank

TIB, German National Library of Science and Technology –

Leibniz Information Centre for Science and Technology University Library, Germany

Abstract

Currently, information on scientific events such as conferences is often scattered and not available in the long term. With the project ConfIDent we want to develop a service platform for the quality-driven, collaborative curation of semantically structured metadata of scientific events. It will provide reliable and transparent data and workflows for researchers (organizers, speakers, participants) as well as other stakeholders of scientific events such as university administrations, libraries, sponsors, publishers or specialized societies. The sustainability of the service will not only be obtained a user-centered approach but also by connecting it to existing services enabling data exchange, and by the commitment to the FAIR principles. ConfIDent will reach the current desideratum of long-term findable, open, referenceable and reusable metadata on scientific events.

Introduction

Conferences are a central, in some disciplines indispensable element of scholarly communication. They allow a broad and quick overview about new research topics and areas and present an opportunity to

- network with your community;
- get informal and fast feedback from peers independent from long publication lifecycles, and
- publish articles as conference proceedings.

Information on conferences and their resulting outputs such as proceedings, videos, reports and other formats of documentation can be found on numerous platforms that function as disciplinary and interdisciplinary services. However, there are three major challenges with metadata on conferences:

1. *Availability* of scholarly event data: The data is often scattered over service platforms, temporary websites, newsletters, etc. Further services are used to publish and archive proceedings and their metadata. These services are not linked to each other and not a small size of data gets lost when conference websites disappear after a few months.
2. The second problem is closely connected to the first one and deals with the *quality* of scholarly event data. There exists no uniform standard for conference metadata. The data that is provided by service platforms is often insufficient, very little structured and/or not available in the long-term. This lack of sustainable event metadata makes the tracking of conference activities and their output a very time-consuming task. Moreover, non-uniform indexing standards make the disambiguation of conference titles more difficult. In particular, this supports the business models of predatory conference organizers to advertise their events with labels of prominent conferences. Especially for young scholars or scholars from foreign research fields it can be difficult to differentiate between serious and fake conferences. The assessment of the content of conferences and their quality requires the insider knowledge of field experts.
3. The commitment to conferences, e.g. by organizing them, presenting there, accomplishing review tasks, can take up a considerable part of the work of researchers without any *acknowledgement* of these activities as research output. The academic system still only rewards publications as evidence of scientific activity and tries to further condense them with the help of singular indicators.

An analysis of 27 conference platforms has shown that considerable deficits of existing services lie in the long-term findability, availability and accessibility of event metadata and content information. We chose platforms that are frequently used, highly advanced and/or have a broad community approach. The majority of platforms do not use any persistent

identifiers (PIDs) and metadata is often inadequate: In some cases they do not even provide core metadata such as the full title of a conference or a working URL for upcoming events. Often, events are treated as singular entities and are not related to superordinate event series. Links between contributors, contributions, proceedings and affiliations are an exception; most content information (e.g., abstracts, organizers, speakers, research topics) are either unstructured or can only be found on temporary event websites. Best practices can be seen in community driven approaches which seem to attract more users and to encourage better curation of contents.

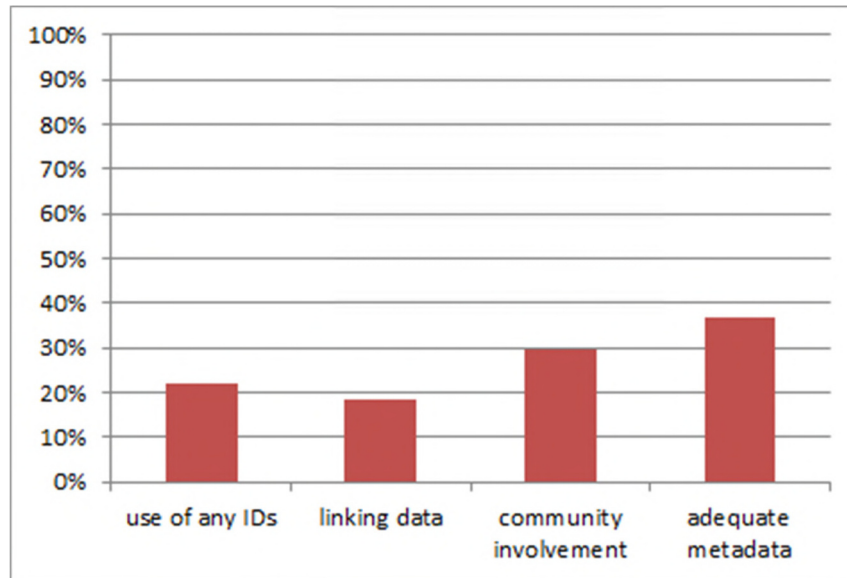


Figure 1: Analysis of 27 conference platforms services and their key components.

The ConfIDent Project

Objectives

In December 2019 ConfIDent started, a joined project of Technische Informationsbibliothek (TIB) Hannover and RWTH Aachen University which is funded by German Research Foundation (DFG – Deutsche Forschungsgemeinschaft).¹ The project aims to establish a service that provides data on scientific events which enables researchers to find, promote and archive information. It is designed as a pilot project developing a prototype for two research communities – a) computer and data science and b) research on transport and mobility – considering the specific relevance of scientific events within these communities. A user-centred approach supports the development of a service that reflects the information needs of the scientific communities. In addition, we want to empower the users to curate the data on their own in order to use their field expertise. A standardisation of event metadata will not only allow data exchange with existing databases but will also foster the assignment of PIDs to improve the quality of metadata. This is a prerequisite to meet the FAIR data requirements and make conference metadata findable, accessible, interoperable and re-usable in the long-term.²

User-centred approach

In order to develop and operate a service platform for event related scholarly metadata that is geared to community needs of specific target groups a user-centred design approach³ is applied. This ensures that the platform will be easy and intuitive to use and that the user can effectively and efficiently achieve the desired result during the interaction with the system. First of all, information is collected with potential users from the target groups regarding

¹ Project Website ConfIDent: <https://projects.tib.eu/en/confident> (access: 20/01/14).

² FORCE11 – FAIR Data Principles: <https://www.force11.org/group/fairgroup/fairprinciples> (access: 20/01/14).

³ ISO-Standard 9241-210:2019: Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems: <https://www.iso.org/standard/77520.html> (access: 20/01/14).

their typical usage behaviour, tasks, needs and expectations regarding the platform. In particular, the domain-specific characteristics of conferences will be taken into account. The target groups will be surveyed by typical user centred design methods such as brainstorming, task analysis, user stories, personas, focus groups and interviews. The analysis results of the user requirements form the basis for the identification of the functionalities and the visualisation of the platform. In an iterative approach potential users from the identified user groups will work together with the developers already in the early phases of specification. Two iterations of evaluations are planned during the development process of the platform. In that way errors and difficulties of the different user groups and stakeholders can be identified at early stages of the process. Central points of the evaluation are the effectiveness, the usability, and the user experience of the platform. Measuring and testing these aspects is achieved by a combination of different methods like task-based usability testing combined with eye tracking, thinking aloud, user interviews and questionnaires. We plan on a first prototyped version after six months of development. A revised and optimized version will be created in the first project year.

In order to become a long-lasting, reliable and accepted platform for science conferences and other scientific events and to reach a critical mass of users the development of a scholarly and a technical community around the ConfIDent project is an important objective. In order to accomplish this, we will establish connections of the ConfIDent platform to existing services including library catalogues, research information systems such as VIVO,⁴ publishing platforms like Open Journal Systems e.g. Copernicus Publications⁵ or the TIB AV-Portal⁶ for the provision of conference recordings. By connecting the resources of various existing services, ConfIDent ensures permanent links between different resource types (such as proceedings, recordings, contributor profiles, organizers etc.). The provision of conference metadata and the assignment of persistent identifiers allow both researchers and infrastructure providers an improved disambiguation and quality assessment of scientific events.

Based on the requirements of users and researchers from the target communities as well as their usage behaviour, various business models will be developed within the framework of the project. These will be discussed and further developed in user workshops. Moreover, the project partners will raise awareness in the target communities for the benefits of ConfIDent by attending events, sending announcements and invitations. In particular, this effort will be supported by the German Informatics Society (GI)⁷ in computer science.

The communities will also be engaged into contributing to the ConfIDent software and platform by inviting research communities to use the ConfIDent platform as a test bed for their developments and tools such as recommender systems, graph partitioning, and clustering methods.

Based on the result of the user needs analysis and in close cooperation with the communities, guidelines for potential content deliverers will be developed, which clearly define the portal's profile. Further user workshops, webinars and training materials such as explanatory videos will be developed to support the communities when using the portal.

Metadata quality, PIDs, and scientometric indicators

One of the core aspects of the project is to define a framework for high quality of metadata and content. Rich metadata helps to disambiguate event data and provides context information on events. PIDs strongly support the standardization of metadata, and are a prerequisite for FAIR data including long-term availability of information.⁸

⁴ Conlon et al. (2019).

⁵ Copernicus Publications: <https://publications.copernicus.org> (access: 20/01/14).

⁶ TIB AV-Portal: <https://av.tib.eu> (access: 20/01/14).

⁷ Gesellschaft für Informatik: <https://gi.de> (access: 20/01/14).

⁸ Demeranville (2018).

ConfIDent maps existing machine readable metadata schemes as used e.g. by DataCite, Crossref, ORCID or ROR for PID registration as well as schema.org., the RDA Framework of the Integrated Authority File (GND - Gemeinsame Normdatei) used by the German National Library and the cataloguing system of the Common Library Network (GBV - Gemeinsamer Verbundkatalog).⁹ In this way, interoperability with existing services should be achieved. Furthermore, we are actively engaged in the international Working Group initiated by DataCite and Crossref to develop a Conference PID.¹⁰ This PID will provide a response to the specific metadata requirements for events as a resource type and will support a standardization of conference information which is still a desideratum. Standardization will help to support a quality assessment approach to conference information both on the metadata and the content level. ConfIDent will have a tiered metadata concept and differentiate between mandatory, recommended and optional fields. On the one hand, this modular approach offers a minimum, generic set of metadata that is necessary to identify events unambiguously; on the other hand it allows subject-specific adjustments as the event related information needs may differ from research community to research community. Content quality criteria for conferences can to a certain extent be represented by metadata as indicators. However, it must always be considered that

1. each indicator has a limited value in itself; and
2. quality criteria for conferences are extremely dependent on the professional culture of each research field or community.

Therefore, we want to define the metadata requirements together with experts from the pilot communities in order to take into account their information and quality requirements. Rich metadata offers more options to assess the quality of an event, but it also means more effort for data providers and curators. These users should also be recruited from the scientific communities, but we want to balance information requirements with the workload for the individual user.

Technical base

Building upon the experiences of both project partners with OpenResearch.org¹¹ (OR) – an experimental platform for the analysis of research information on events, papers, projects and other entities – the project will start with a prototype based on the open source software Semantic MediaWiki (SMW)¹² fostering openness and extensibility. The SMW prototype will be customized to metadata and user requirements identified in an iterative process as described by the user-centred approach. The software allows collaborative data curation and options for a high degree of transparency to display the provenance of data. Nevertheless, this collaborative approach challenges the objective of high quality metadata. ConfIDent aims to provide a mature rights and roles managements as well as guidance for data ingest and curation to allow easy use but prevent misuse. We want to provide an open platform with interfaces to allow data exchange with existing services and an attractive platform for individual users and their information needs.

Extensional use of data

The use of scientometric indicators to measure the impact of scientific output is heavily discussed, with some key documents defining the outlines of what is called altmetrics

⁹ Crossref Metadata Schema 4.4.2 on conferences:

http://data.crossref.org/reports/help/schema_doc/4.4.2/schema_4_4_2.html#conference; DataCite Metadata Schema 4.3: https://schema.datacite.org/meta/kernel-4.3/doc/DataCite-MetadataKernel_v4.3.pdf; GND Erfassungshilfen für Körperschaften und Konferenzen:

<https://wiki.dnb.de/pages/viewpage.action?pageId=90411359>; ORCID Metadata Schema 3.0:

https://github.com/ORCID/ORCID-Source/tree/master/orcid-model/src/main/resources/record_3.0; schema.org event: <https://schema.org/Event> (access: 20/01/14).

¹⁰ Birukou (2018).

¹¹ Vahdati et al. (2016).

¹² Semantic MediaWiki: https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki (access: 20/01/14).

(alternative metrics).¹³ These metrics track the use of research outcomes online, such as on social media, news sites, blogs and policy papers. Compared to conventional metrics such as citation counts, altmetrics provide among others a time advantage and the possibility to include mentions outside the scientific communities.¹⁴

Conclusion

The central objective of the project is to develop a sustainable service platform that provides reliable data on scientific events. ConfIDent will not only facilitate quality assessment of data with regard to a wide range of criteria and stakeholders' perspectives, taking into account broad context information. The platform will also foster a cultural change in science by providing a higher visibility of scientific events as an independent achievement of research beyond counting article citations and by promoting their impact. We see the initial community oriented approach as starting point for the development of a generic service that serves the scientific community as a whole. The service is supposed to be connectable to numerous projects and initiatives that aim to better capture the heterogeneity of scientific outputs and making them accessible.¹⁵

Literature

- Auer, Sören; Blümel, Ina; Ewerth, Ralph; Garatziogianni, Alexandra; Heller, Lambert; Hoppe, Anett; Kasprzik, Anna; Koepler, Oliver; Nejd, Wolfgang; Plank, Margret (2018): Towards An Open Research Knowledge Graph (Zenodo). DOI: <https://doi.org/10.5281/zenodo.1157185>.
- Birukou, Aliaksandr (2018): PIDs for conferences - your comments are welcome! (DataCite Blog). DOI: <https://doi.org/10.5438/prq9-ed42>.
- Collins, Sandra; Genova, Françoise; Harrower, Natalie; Hodson, Simon; Jones, Sarah; Laaksonen, Leif et al. (2018): Turning FAIR into reality. Edited by European Commission Expert Group on FAIR Data. Directorate-General for Research and Innovation. DOI: <https://doi.org/10.2777/1524>.
- Conlon, Michael; Woods, Andrew; Triggs, Graham; O'Flinn, Ralph; Javed, Muhammad; Blake, Jim et al. (2019): VIVO: a system for research discovery. In JOSS 4 (39), p. 1182. DOI: <https://doi.org/10.21105/joss.01182>.
- Demeranville, Tom (2018): Building a robust research infrastructure, one PID at a time (ORCID Blog). Available online at <https://orcid.org/blog/2018/08/08/building-robust-research-infrastructure-one-pid-time>.
- Fraumann, Grischa (2018): The Values and Limits of Altmetrics. In New Directions for Institutional Research 2018 (178), pp. 53–69. DOI: <https://doi.org/10.1002/ir.20267>.
- Hauschke, Christian; Cartellieri, Simone; Heller, Lambert (2018): Reference implementation for open scientometric indicators (ROSI). In RIO 4, p. 59. DOI: <https://doi.org/10.3897/rio.4.e31656>.
- Hicks, Diana; Wouters, Paul; Waltman, Ludo; Rijcke, Sarah de; Rafols, Ismael (2015): Bibliometrics: The Leiden Manifesto for research metrics. In Nature 520 (7548), pp. 429–431. DOI: <https://doi.org/10.1038/520429a>.
- Tunger, Dirk; Clermont, Marcel; Meier, Andreas (2018): Altmetrics: State of the Art and a Look into the Future. In Mari Jibu, Yoshiyuki Osabe (Eds.): Scientometrics: InTech. DOI: <https://doi.org/10.5772/intechopen.76874>.
- Vahdati, Sahar; Arndt, Natanael; Auer, Sören; Lange, Christoph (2016): OpenResearch: Collaborative Management of Scholarly Communication Metadata. In Eva Blomqvist, Paolo Ciancarini, Francesco Poggi, Fabio Vitali (Eds.): Knowledge Engineering and Knowledge Management. 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings, vol. 10024. Cham, s.l.: Springer International Publishing (Lecture Notes in Computer Science, 10024), pp. 778–793. DOI: https://doi.org/10.1007/978-3-319-49004-5_50.
- Wilsdon, James; Bar-Ilan, Judit; Frodeman, Robert; Lex, Elisabeth; Peters, Isabella; Wouters, Paul (2017): Next-Generation Metrics: Responsible Metrics and Evaluation for Open Science. Report of the European Commission Expert Group on Altmetrics. Handle: <http://hdl.handle.net/1887/58254>.

¹³ Fraumann (2018); Tunger et al. (2018).

¹⁴ Hauschke et al. (2018); Hicks et al. (2015); Wilsdon et al. (2017).

¹⁵ Auer et al. (2018); Collins et al. (2018); FORCE11 – The Fair Data Principles: <https://www.force11.org/group/fairgroup/fairprinciples>; PID Graph at FREYA Project: <https://www.project-freya.eu/en/about/mission> (access: 20/01/14).