

# Leveraging Knowledge Graph Technologies to Assess Journals and Conferences at Springer Nature

Simone Angioni<sup>1</sup>, Angelo Salatino<sup>2</sup>, Francesco Osborne<sup>2,3</sup>, Aliaksandr Birukou<sup>4</sup>, Diego Reforgiato Recupero<sup>1</sup>, and Enrico Motta<sup>2</sup>

<sup>1</sup> Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy {simone.angioni, diego.reforgiato}@unica.it

<sup>2</sup> Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom {angelo.salatino, francesco.osborne, enrico.motta}@open.ac.uk

<sup>3</sup> Department of Business and Law, University of Milano Bicocca, Milan, Italy

<sup>4</sup> Springer-Verlag GmbH, Tiergartenstrasse 17, 69121 Heidelberg, Germany aliaksandr.birukou@springer.com

**Abstract.** Research publishing companies need to constantly monitor and compare scientific journals and conferences in order to inform critical business and editorial decisions. Semantic Web and Knowledge Graph technologies are natural solutions since they allow these companies to integrate, represent, and analyse a large quantity of information from heterogeneous sources. In this paper, we present the AIDA Dashboard 2.0, an innovative system developed in collaboration with Springer Nature to analyse and compare scientific venues, now also available to the public. This tool builds on a knowledge graph which includes over 1.5B RDF triples and was produced by integrating information about 25M research articles from Microsoft Academic Graph, Dimensions, DBpedia, GRID, CSO, and INDUSO. It can produce sophisticated analytics and rankings that are not available in alternative systems. We discuss the advantages of this solution for the Springer Nature editorial process and present a user study involving 5 editors and 5 researchers, which yielded excellent results in terms of quality of the analytics and usability.

**Keywords:** Scholarly Data · Knowledge Graphs · Scholarly Ontologies · Science of Science · Scholarly Analytics · Scholarly Knowledge

## 1 Introduction

Springer Nature (SN) is one of the main publishers of research in Computer Science and manages a vast catalogue of about 162 journals in this field and several series of proceedings books (e.g., LNCS, LNAI, IFIP-AICT, CCIS, LNBIP) for a total of about 800 volumes per year. Their data analysts have to regularly integrate and analyse a large quantity of information regarding these venues<sup>5</sup> for supporting crucial business and editorial decisions. In particular, SN editorial team needs to compare all journals and conferences in a field according to several metrics, be aware of which venues are rising and attracting more attention

---

<sup>5</sup> In this paper, we use the term ‘venue’ to denote both journals and conferences.

in the community, monitor how they change over time in terms of researchers and topics distribution, and assess the involvement of commercial organisations. However, bibliometric systems and academic search engines provide a limited support for analysing scientific venues. This led to the creation of the AIDA Dashboard, an innovative tool for supporting editors in performing advanced analysis of these dynamics.

In this paper, we present the AIDA Dashboard 2.0, the last version of the system developed in collaboration with SN to analyse and compare journals and conferences, which we are now releasing to the wider scientific community. This tool builds on the *Academia/Industry DynAmics Knowledge Graph* (AIDA KG) [3], a knowledge graph which includes over 1.5B RDF triples and was produced by integrating information about 25M papers from Microsoft Academic Graph, Dimensions, DBpedia, and the Global Research Identifier Database (GRID). Journals and conferences are categorised according to the *Focus Areas Taxonomy*<sup>6</sup>, a new ontology detailing the 124 most prominent research fields within Computer Science venues. The specific research topics in these venues are instead represented according to 14K research topics from the *Computer Science Ontology*<sup>7</sup> (CSO) [24], whereas the industrial sectors of the organisations (e.g., education, energy, financial, technology) are represented with the *Industrial Sectors Ontology*<sup>8</sup> (INDUSO). The last version of the AIDA knowledge graph (ver. 3.0) is publicly available via a dump and a SPARQL endpoint at <https://w3id.org/aida>.

The main novelties with respect to the earlier version of AIDA Dashboard reported in previous work [4] and presented as a demo at ISWC 2020 [5] include: 1) the ability to analyse journals in addition to conferences, 2) a new expert search functionality that allows users to browse, compare, and order journals and conferences according to several metrics and ranking systems, 3) a new taxonomy of high-level research areas, representing the main research fields used to classify the venues, and 4) the full integration of the dashboard with the SN Data Cloud Infrastructure.

The AIDA Dashboard was evaluated by performing a user study involving five SN editors and five researchers, which yielded excellent results in terms of usability and quality of the analytics.

In order to support the scientific community, we recently released a publicly available version of the system, that can be accessed at <https://w3id.org/aida/dashboard>. We hope that it could become a standard tool used by researchers, institutions, and funding agencies for analysing venues in Computer Science. We plan to keep updating it in the following years and also add more entities to analyse (e.g., researchers, organizations, countries, topics).

We also release the *AIDA Venue dataset*<sup>9</sup>, a machine readable version of all the analytics produced by the AIDA Dashboard on journals and conferences

---

<sup>6</sup> Focus Areas Taxonomy - <https://w3id.org/aida/fat>

<sup>7</sup> CSO - <https://cso.kmi.open.ac.uk/>

<sup>8</sup> INDUSO - <https://w3id.org/aida/downloads/induso.ttl>

<sup>9</sup> AIDA Downloads - <https://w3id.org/aida/downloads>

that can support bibliometric analysis and be used to train machine learning systems.

In summary, the novel contributions of this paper include:

- the AIDA Dashboard 2.0, a new version of the AIDA Dashboard which offers several new functionalities;
- a user study involving five SN editors and five researchers;
- a discussion of the impact and uptake of this tool within SN;
- the AIDA KG 3.0, the last version of the AIDA KG including 25M publications;
- the AIDA Venues dataset 2.0, a new resource describing 3,263 journals and 2,003 conferences in Computer Science<sup>10</sup> according to all the data produced by the AIDA Dashboard back-end.

The remainder of the paper is structured as follows. In Section 2, we describe the AIDA Dashboard back-end and the sustainability plan. Section 3 details the AIDA Dashboard GUI. Section 4 presents the evaluation study and Section 5 describes the uptake and impact of the AIDA Dashboard within SN. Section 6 presents the related work. Finally, Section 7 concludes the paper and discusses future work.

## 2 The AIDA Engine

The AIDA Dashboard is powered by a complex pipeline for data integration and analysis. Figure 1 summarises its architecture, which is composed by three main components (grey dashed boxes): i) the pipeline for the generation of AIDA KG, ii) the module for pre-computing the analytics, and iii) the AIDA Dashboard GUI. The main data about research articles and relevant metadata are stored in the SN Data Cloud Infrastructure (purple dashed box), which is based on a Google BigQuery instance. This infrastructure regularly downloads four datasets from external data sources, i.e., MAG, Dimensions, DBLP, and OpenAlex. These data are then integrated with several other knowledge bases (upper part of the figure) for generating the AIDA Knowledge Graph. We then compute several analytics that will be reported by the AIDA Dashboard GUI. The following subsections will describe more in detail the generation of AIDA KG and the analytics. The interface of the AIDA Dashboard will be instead described in Section 3.

### 2.1 The AIDA Knowledge Graph

The AIDA Knowledge Graph is automatically generated by integrating several knowledge bases, including MAG, Dimensions, CSO, DBpedia, GRID, and INDUSO. The knowledge graph describes 25M research papers and 8M patents

---

<sup>10</sup> These numbers are the results of a selection process that identifies only venues active in the last 5 years.

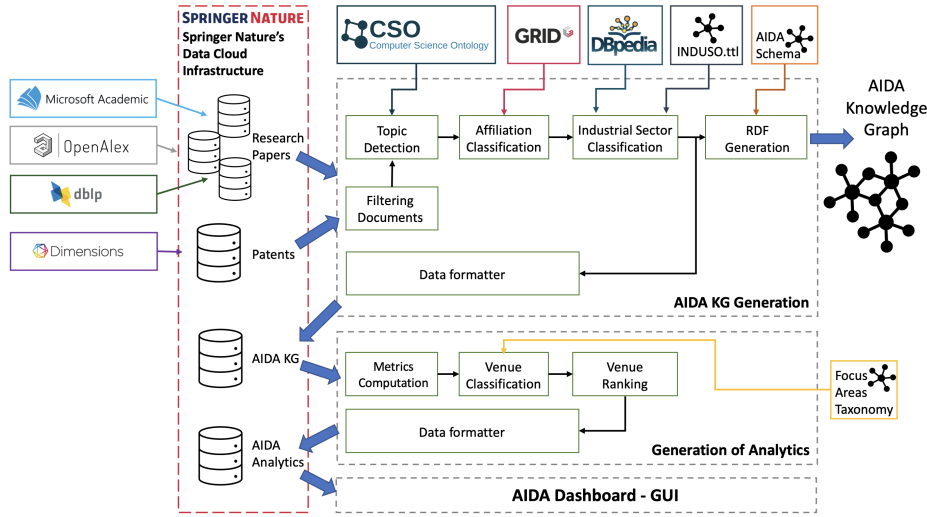


Fig. 1. The AIDA Engine Architecture

in the field of Computer Science. All these documents are classified with the Computer Science Ontology topics [25]. 6.7M papers and 5.6M patents are also classified based on the type of authors' affiliations (i.e., academia, industry, collaborative), and the industrial sectors from INDUSO. Patents are used by SN analysts for supporting analysis on research impact, but they are not employed yet by the AIDA Dashboard. Since Microsoft recently decommissioned MAG, we are now switching to a combination of OpenAlex and DBLP, as detailed by Section 2.3 (Sustainability Plan).

The integration pipeline first selects all research papers and patents respectively from MAG and Dimensions. It applies filters to select the documents within the Computer Science field. It then uses the CSO Classifier [20] to annotate all documents with their relevant research topics. Next, it leverages the GRID IDs associated with the original data to determine whether the documents are authored by either academic or industrial institutions, or through a collaborative effort. For all papers authored by industrial affiliations, it uses DBpedia to classify their industrial sectors according to INDUSO. The reader can refer to [2] for additional details about the AIDA KG generation.

The data model of the resulting knowledge graph builds on the AIDA Schema<sup>11</sup> (`aida:`), Schema.org (`schema:`), FOAF, OWL, CSO schema<sup>12</sup> (`cso:`), Microsoft Academic KG schema<sup>13</sup> (`mag:`), GRID schema<sup>14</sup> (`grid:`) and others. Specifically, in the context of the AIDA Dash-

<sup>11</sup> The AIDA Schema - <https://w3id.org/aida/ontology>

<sup>12</sup> The CSO Schema - <https://w3id.org/cso/schema/cso>

<sup>13</sup> The MAKG Schema - <https://makg.org/ontology.owl>

<sup>14</sup> The GRID ontology - <http://www.grid.ac/ontology/>

board, we leverage five main entities defined in the AIDA Schema (`aida:paper`, `aida:author`, `aida:affiliation`, `aida:industrialSector`, and `aida:DBpediaCategory`) as well as seven additional entities reused from external ontologies (`mag:paper`, `mag:author`, `grid:affiliation`, `mag:Journal`, `mag:ConferenceSeries`, `mag:ConferenceInstance`, and `cso:Topic`). These entities are interconnected through 23 semantic relations, 10 defined in the AIDA schema and 13 reused from external ontologies (e.g., `cito:cites`, `datacite:doi`, `dc:title`, `prism:publicationDate`, `schema:memberOf`). In particular, the following 4 relations (out of the 10 defined in the schema) characterise articles according to the relevant information from DBpedia, GRID, INDUSO, and CSO:

- `aida:hasDBpediaCategory`, indicating the industrial sectors (DBpediaCategory) obtained from several DBpedia fields, such as `About:Property` and `About:Industry`;
- `aida:hasGridType` showing the type of an affiliation according to the GRID classification (e.g., education, company, government, non-profit);
- `aida:hasIndustrialSector`, indicating the INDUSO industrials sector of an affiliation;
- `aida:hasTopic`, indicating the CSO topics identified in a paper.

A more comprehensive description of the AIDA Schema is available at <http://w3id.org/aida/#aidaschema>. The AIDA knowledge graph is serialised in RDF and can be downloaded from <https://w3id.org/aida/downloads>. It can also be queried via a SPARQL endpoint at <https://w3id.org/aida/sparql>.

## 2.2 Generation of Analytics

The second component of the AIDA Engine takes in input the AIDA KG and produces the analytics related to journals and conferences. We pre-computed all analytics to improve scalability and response time. The analytics are computed in three phases: i) we retrieve all information about venues and produce a very comprehensive set of metrics about them and related entities (e.g., topics, authors, organisations); ii) we classify venues according to their main research fields, using the Focus Areas Taxonomy; iii) we produce the venue rankings in different fields according to both our metrics and a set of external ratings, such as SJR and CORE.

**Metrics Computation.** We first get from AIDA KG the journals and conferences which counted at least 50 publications in the last 5 years. The current version includes 3,263 journals and 2,003 conferences. We then compute a set of performance metrics based on citations, such as h-index, h5-index, and impact factor (on the previous 2 years). Next, we use the `schema:creator`, `aida:hasAffiliation`, `grid:countryName`, and `aida:hasTopic` relationships to select the top 100 authors, organisations, countries, and main topics in terms of publications and citations. These are computed both as totals (e.g., all years, last 5 years, last 10 years) and as distributions over time. For authors and organisations, we also compute their h-index and h5-index considering all their

publications in the knowledge graph. In addition to the main topics, we also identify the top 100 *fingerprint topics*, which are ranked according to the difference between the fraction of relevant publications in the venue and the average distribution in the whole Computer Science. This metric usually identifies the topics that are most significant for the underlying venue, also according to the user assessment [4]. Finally, we compute the number of publications and citations received from the research papers written by academia, industry, and collaborations, and the distribution of the industrial sectors at the venue.

**Venue Classification.** It is crucial to categorise journals and conferences with their research fields, in order to contextualise performance metrics and comparisons with other venues. Indeed, comparing the h-index or impact factor of a journal in the area of *Neural Networks* with one in *Formal Logic* would neither be fair nor informative. To this end, we created the *Focus Areas Taxonomy*<sup>15</sup> containing 124 broad areas organised into 4 levels. In contrast, the taxonomy used by Google Scholar Metrics for characterising venues includes only 26 categories relevant to Computer Science.

The Focus Areas Taxonomy has been created following both bottom-up and top-down strategies. We first selected 200 research areas from CSO that appeared as sub-string in a venue name and were also in the top 10 fingerprint topics for at least 10 venues. We then included all the super-topics of the first set (which are 392), resulting in a total of 592 candidate topics. We then associated them with various metrics linked with their prevalence in the 5.2K venues in our system. These included the frequency of appearance in journal or conference titles and the number of journals and conferences in which it appears among the top 10 topics. Finally, we arranged all the topics within a taxonomy following the same structure of CSO. We then asked three senior researchers in Computer Science to revise the taxonomy by 1) selecting the most significant topics on the basis of the metrics and their expertise, and 2) rearranging their position in the taxonomy if needed. We plan to keep updating this knowledge base according to feedback from the editors and the community.

In order to classify venues with the fields of the Focus Areas Taxonomy, we first check if the venue mentions a focus area in the name and in that case we directly assign to this area. For instance, the *International Conference on Robotics and Automation (ICRA)* is automatically assigned to “Robotics”. Otherwise, we identify the focus area with the highest coverage in the distribution of fingerprint topics and, among its descendants, we select the most specific area with at least 20% of publications in the venue.

**Venue Ranking.** The last step consists of pre-computing the rankings of the venues across specific fields. For each focus area, we generate a list of all relevant venues along with a set of metrics, such as h5-index, average h5-index of the relevant organizations, and average number of papers in the last 5 years. We also include well-known external journal and conference ratings made available by various associations. For journals, we use the SCIMAGO Journal Rank (SJR),

---

<sup>15</sup> Focus Areas Taxonomy is browsable here: <https://w3id.org/aida/fat>

and the SCIMAGO Quartile<sup>16</sup>. For conferences, we use the ranks provided by liveSHINE<sup>17</sup>, CORE<sup>18</sup>, and GII-GRIN-SCIE (GGS)<sup>19</sup>.

### 2.3 Sustainability Plan

We plan to keep maintaining and updating the dashboard in the following years. For this reason, we set up an automatic pipeline that will update the data every 6 months. In addition, we will keep developing the main functionalities, following the feedback of the community and the editors (see Section 4).

At the end of 2021 Microsoft decommissioned the MAG project<sup>20</sup>. We thus decided to introduce two additional datasets within our integration pipeline: OpenAlex<sup>21</sup> and DBLP<sup>22</sup>, as shown in Fig. 1. We included OpenAlex because it shares the same schema with MAG and it has a low cost of integration. However, since OpenAlex does not disambiguate conferences yet, we leveraged the conference representation of DBLP, by mapping papers across the two datasets. To achieve this, we designed a two-stage pipeline. We firstly mapped papers with the same DOI. Then, for the conferences that do not assign DOIs to articles (e.g., AAAI, NeurIPS), we mapped the papers across the two datasets by computing the string similarity of their titles. We worked in close collaboration with the SN Data Science team and we now have stable version with over 95% conference papers matched between DBLP and OpenAlex. Future versions of AIDA KG and the generated analytics will be based on these newly integrated datasets.

## 3 The AIDA Dashboard

The AIDA Dashboard is a web application that allows users to analyse and compare journals and conferences. In May 2022, we released a public version, independent from Springer Nature internal workflow, that is available at <https://w3id.org/aida/dashboard>.

In the starting page, the users can search either for a venue or a research topic. If the user selects a conference or a journal, the system will retrieve the pre-computed analytics (as JSON file) and display them in the *venue panel*. If the user selects a research field (e.g., *Artificial Intelligence*) the application will return the *advanced search panel* that ranks all venues in the field according to a variety of metrics and allows the user to navigate and compare them. In the following, we describe these two interfaces in details.

<sup>16</sup> SCIMAGO - <https://www.scimagojr.com>

<sup>17</sup> liveSHINE - <http://web.archive.org/web/20180728060959/http://liveshine.icomp.ufam.edu.br/>

<sup>18</sup> CORE - <https://www.core.edu.au/team>

<sup>19</sup> GGS - <https://scie.lcc.uma.es:8443/conferenceRating.jsf>

<sup>20</sup> Next Steps for Microsoft Academic – Expanding into New Horizons - <https://www.microsoft.com/en-us/research/project/academic/articles/microsoft-academic-to-expand-horizons-with-community-driven-approach/>

<sup>21</sup> OpenAlex - <https://openalex.org>

<sup>22</sup> DBLP - <https://dblp.org>

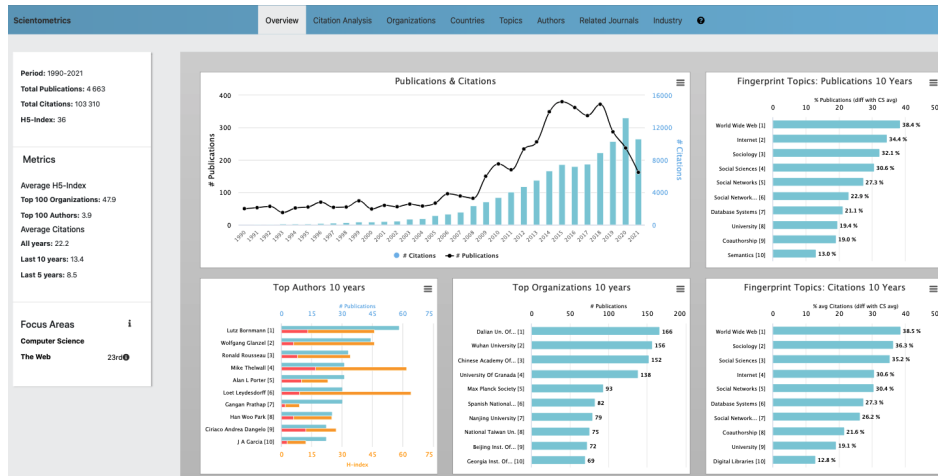


Fig. 2. Overview page of Scientometrics (journal).

### 3.1 Venue Panel

The venue panel is structured in eight tabs: i) *Overview*, ii) *Citation Analysis*, iii) *Organizations*, iv) *Countries*, v) *Authors*, vi) *Topics*, vii) *Related Conferences/Journals*, and viii) *Industry*.

The users first lands on the **Overview** tab, shown in Fig. 2, which displays the most important metrics and a selection of charts, including: a) publications and citations across time, b) the top fingerprint topics in terms of publications and citations in the last 10 years, and c) the top 10 authors and organisations in the last 10 years.

The **Citation Analysis** tab reports how the venue is performing in terms of citation-based metrics, such as impact factor and average citations over time. Notably, it also shows the evolution of the venue’s rank and percentile within its focus areas. This is a very intuitive measure of the prominence of the conference in the field over time, which is not available in alternative tools.

The **Organizations**, **Authors**, **Countries**, and **Topics** tabs allows users to rank and inspect these entities in terms of publications, citations, and average citations. All metrics can be displayed either as totals (all years, last 5 years, or last 10 years) or as time-based distributions.

The **Related Conferences/Journals** tab allows users to compare the venue of interest with other venues in the same fields according to their number of publications, citations, and average citations across time. This diachronic view is very useful for identifying emerging conferences or journals that may not be dominant yet, but exhibit a strong positive trend.

The **Industry** tab presents the distribution of publications and citations of academic institutions, industrial organisations, and collaborative efforts. This tab displays also the distribution of publications and citations of the different



industrial sectors that published at the venue. For instance, the *NeurIPS* conference attracts publications prevalently from companies in the sectors: “*Technology*”, “*Computing and IT*”, “*Marketing*”, and “*Electronics*”.

### 3.2 Advanced Search

Figure 3 displays the Advanced Search panel, which allows users to browse and compare venues according to their fields. The user can browse the different fields using the selection menus and switch between journals and conferences with the button in the upper right. For instance, a user checking all the conferences in the field “*The Web*” can decide to focus further the analysis and only show the subset of venues within the sub-area “*Semantic Web*”. Clicking on a specific venue will bring the user to the relevant venue panel.

ID	Conference Name	Avg Citations (Last 5 Years)	Avg Publications (Last 5 Years)	H5-Index	Avg Orgs H5-Index	Avg Authors H5-Index	LiveSHINE Rank	CORE Rank	GGS Rank
	<a href="#">Neural Information Processing Systems (NeurIPS)</a>	30.8	1405	188	91	25.2	A++	A++	A++
	<a href="#">International Conference on Learning Representations (ICLR)</a>	49.3	637	175	88.9	29.1	-	A++	A++
	<a href="#">Meeting of the Association for</a>	23.6	970	146	84.9	22.4	A++	A++	A++

**Fig. 3.** The Advanced Search Panel displaying conferences in Artificial Intelligence ranked by h5-index.

Journals and conferences can be ranked according several metrics, including: a) average citations received in the last five years, b) average articles published in the last 5 years, c) h5-index, d) the average h5-index of the relevant organisations, and e) the average h5-index of the relevant authors. The last two metrics are not typically offered by alternative systems, but are very useful to identify emergent conferences that are attracting strong research groups but may not have yet received a good number of citations. Venues can be also ranked according to the set of external ratings discussed in Section 2.2.

## 4 User Study

We performed a user study on the AIDA Dashboard to assess the quality and usefulness of the analytics as well as the usability of the user interface. To this

end, we organised individual sessions with 5 SN editors and 5 researchers in Computer Science. In each session, we first presented the AIDA Dashboard 2.0 for about 20 minutes. We then assigned to the users the task of analysing two venues and a focus area of their expertise in order to assess the quality of the resulting analytics. After the hands-on session the users filled a five-parts survey about their experience. The first part covered the users background and expertise. The second part was a standard System Usability Scale<sup>23</sup> (SUS) [7] questionnaire to gauge the usability of the AIDA dashboard. The third section asked the users to rate the quality of the analytics for the two venues and the focus area on a Likert scale in the [1-5] range. The fourth part included four open questions about strengths and weaknesses of the dashboard asked to all users and two further questions that were asked only to the editors. Finally, the fifth part asked to list at least three of the most useful functionalities.

The data produced during the user study are available online<sup>24</sup>.

#### 4.1 User Background

The five researchers in the user study are all senior researchers, with an average of 13.4 years of experience, and come from different institutions: i) University of Cagliari (IT), ii) Institute for Applied Informatics (DE), iii) FIZ Karlsruhe – Leibniz (DE), iv) University of Paris 13 (FR), and v) National Council of Research (IT). The five editors are at various career stages (1, 5, 13, 21, and 25 years of experience) and come from different departments within SN.

The areas of expertise of the 10 users include Artificial Intelligence, Natural Language Processing, Semantic Web, Robotics, Machine Learning, Multimedia Systems, and Theoretical Computer Science.

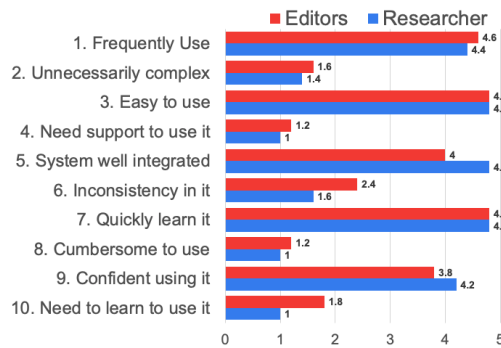


Fig. 4. The SUS Questionnaire results.

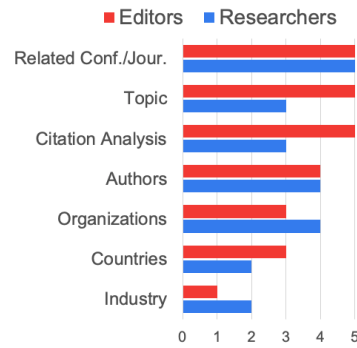


Fig. 5. Number of votes received by each Section/Functionality.

<sup>23</sup> System Usability Scale (SUS) - <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

<sup>24</sup> AIDA Evaluations - <https://w3id.org/aida/downloads#evaluation>

## 4.2 SUS questionnaire

The SUS questionnaire provided excellent results obtaining a score of 88.5/100 considering all users. This corresponds to the 97% percentile rank in terms of usability (A+ grade) according to the SUS guidelines<sup>25</sup>. In general, editors were more severe than researchers, mostly because they consider the dashboard an important working tool and they were very motivated in suggesting further improvements. Indeed, editors scored an average 84.5 SUS score (96% percentile rank), while researchers yielded 92.5 (98%). This version of the dashboard (2.0) showed a better usability than the previous one, which achieved a SUS score of 87.5 in a user study involving 10 researchers [4].

Figure 4 reports the average score given by researchers (red bars) and editors (blue bars) to specific questions in the SUS questionnaire. Odd questions are positive (a higher score is better) while even ones are negative (a lower score is better). Overall, all the users found the system very easy to use (high values in question 3), they could easily learn the system (question 7), and they do not need support to use the system (question 4). The editors found some inconsistency in the integration of the functionalities (question 5). Finally, all users would like to frequently use the dashboard (question 1).

## 4.3 Quality Assessment

We asked the users to evaluate the quality of the analytics produced by the AIDA Dashboard for the two venues and the focus area according to a Likert scale. On average, editors scored 3.8 for venues and 4 for focus areas, whereas researchers 4.2 for both venues and focus areas. The range of fields and venues analysed by the users included Artificial Intelligence (AAAI, ICML, EANN, NC&L, Machine Learning), Natural Language Processing (EMNLP, ACL, EACL), Multimedia Systems (ACM Multimedia, Multimedia Tools & Applications), Robotics (ICRA, IROS), The Web (The Web Conference), Information Retrieval (SIGIR), Digital Library (TPDL), Semantic Web (ISWC), and Theoretical Computer Science (Information & Computation, iConference).

## 4.4 Open Questions

We summarise here the main feedback emerged from questions Q1-Q4 (all users) and questions Q5-Q6 (only editors).

**Q1. What are the main strengths of AIDA Dashboard?** Users were positively impressed by the easy and intuitive interface and the large amount of analytics. Other positive feedback regarded the granularity of the topic classification and the fact that the system addressed a real need in the community, i.e. analysing and comparing venues.

**Q2. What are the main weaknesses of AIDA Dashboard?** Users listed a range of issues that we plan to address in the future. One researcher suggested

<sup>25</sup> Interpreting a SUS score - <https://measuringu.com/interpret-sus-score/>

that the major limitation is that the coverage is constrained to the Computer Science domain. Another one reported some disambiguation issues, in particular regarding authors with similar names. One more suggested that certain functionalities were hard to locate because the second level tabs were not particularly discernible. One editor mentioned the need of analysing venues in time ranges smaller than 5 years. Another one criticised the current interface for navigating the taxonomy based on selection menu. Finally, one editor did not find smooth the integration of journals and conferences and asked to be able to compare both of them in the same panel.

**Q3. Can you think of any additional features to be included in AIDA Dashboard?** Researchers mentioned: 1) adding more type of scholarly entities to analyse (e.g., organisations, researchers), 2) the ability to compare specific charts from different venues, 3) some additional metrics (e.g., number of papers that contributed to the citation count), 4) various minor GUI improvements, and 5) the ability to rank topics alphabetically. Editors mentioned: 1) the ability to directly compare conferences to journals; 2) a better integration with the CSO taxonomy; 3) adding information about the publishers of the venues, and 4) considering also books series.

**Q4. How comprehensive/accurate do you consider the list of focus areas associated with the venues in AIDA Dashboard?** All the researchers found the list of focus areas accurate and comprehensive. However, two of them suggest that they were sometimes too broad and would have liked the ability to browse venues also according to arbitrary research topics. Four editors found the list very accurate and comprehensive, while one of them identified some missing areas in their field of expertise and suggest edits for the Machine Learning branch (already implemented in the current version).

**Q5. In which way the AIDA Dashboard support your work?** Two editors reported that the system was very useful for supporting junior or new editors in analysing specific research fields. Two found it very helpful in identifying notable trends in venues topics and performing country-centric analysis. One found it very useful in identifying and comparing venues. Some editors also highlighted how the dashboard supports the detection of conferences and workshops that could produce special issues about specific emerging topics.

**Q6. What competitive advantages would you say the AIDA Dashboard provides with respect to Scopus/Google Scholar (if any)?** One editor pointed out that the AIDA Dashboard provides better visualisations as well as more granular analytics compared to Scopus and Google Scholar. One considered the auto-suggested search more helpful and simpler than the one in Scopus search. Finally, an editor found the AIDA Dashboard more powerful in analysing conferences and journals, preferring instead Google Scholar for analysing individual researchers or articles.

#### 4.5 Best Functionalities

We asked the ten users to list at least three of the most useful sections of the AIDA Dashboard. Figure 5 reports the user preferences. The *Related Confer-*

*ences/Journals* tab was the most appreciated section for both editors and researchers. This highlights how comparing venues is a critical task that was not well supported by previous solutions. Interestingly, researchers preferred the analytics about topics and citation analysis, while editors the analysis on authors and organisations.

## 5 Uptake and Impact

The partnership between The Open University and SN has produced a wide range of intelligent tools and services for automatically classifying articles [23] and proceeding books [21], recommending publications [26], evolving domain ontologies [18], and predicting the emergence of research topics [22]. The very first prototype of the AIDA Dashboard was introduced at SN in 2020 and has since been used by their editors and analysts to assess venues in Computer Science. The aim was to inform editorial and marketing decisions regarding the 162 journals and the about 800 proceedings books produced each year. In particular, editors need to monitor the performance across time of journals and conferences within specific fields and take action for improving the coverage and quality of SN catalogue. They also need to scan the horizon to detect the emergence of new scientific communities and relevant venues. It is typically ideal to establish a solid presence in new fields as early as possible by starting relevant journals and publishing the proceedings of new conferences. The editors need also to gain an understanding of the key persons within specific communities that may be invited to editorial boards or to organise special issues. Finally, it is important to assess what industrial sectors are interested in specific community in order to support targeted marketing campaign and specific editorial products. These are very complex and time-consuming analyses that were performed by senior editors on the basis of their personal knowledge and standard metrics offered by commercial datasets, resulting into bottlenecks, delays, and high costs. A task such as comparing all the journals and conferences within a specific research areas (e.g., Cloud Computing), considering also the potential for growth, used to take days of work by editors and assistant editors.

The adoption of AIDA has drastically improved the situation and brought three major benefits. First, it halved the time needed for analysing venues and prepare relevant analytics. Second, it reduced the complexity of the task, allowing less experienced junior editors and editorial assistants to also perform these analysis, improving the distribution of the workload and freeing up the time of the senior editors and analysts. Overall, this resulted in an estimated 60% cost reduction. Finally, the dashboard had a positive effect on both the velocity and the quality of the decision making process, that can be now continuously supported by advanced large-scale analytics.

In preparation to the public release of the AIDA Dashboard, we also made it available to selected members of the research community. During this pilot, the AIDA Dashboard informed the decisions of members of the organiser committees of conferences (SEMANTICS, EKAW) and workshops (DL4KG, Text2KG, Sci-

K). For instance, one of the chairs of SEMANTICS 2023 used the dashboard for gaining a deeper understanding of the conference trends in terms of topics, countries, and organisations, with the aim of supporting strategic decisions for the next edition.

## 6 Related Work

Within the scholarly domain, many knowledge graphs offer a good representation of research papers and their metadata, such as authors, affiliations, topics, and so on. Among them, we find AMiner [30], Microsoft Academic Graph (MAG) [28], PID Graph [8], the Research Graph [6], ScholarlyData [17], the OpenAIRE Research Graph [15], SciGraph<sup>26</sup> [11], OpenCitations [19], the Open Research Knowledge Graph (ORKG) [12], Nanopublications [9,29], and the AIDA Knowledge Graph [2].

Several bibliometric tools and search engines can be used to query information about journals and conferences. For instance, Microsoft Academic Search, now dismissed, offered several metrics relevant to scientific venues, including their citations, topics, related venues, authors, and institutions. However, it did not let users compare venues or examine how research topics evolved across time. AMiner<sup>27</sup> and Semantic Scholar [1] support users in exploring journals and conferences, but they report only the most prominent authors and papers. Scholia<sup>28</sup> [16] consists of a Web service that builds scholarly profiles for topics, people, organisations, and venues on top of the information available in Wikidata<sup>29</sup>. If a journal or conference are selected, Scholia reports some relevant information, such as the main articles ranked by their citations, the main topics, related authors, and organisations. One drawback of this tool is that the topics are associated to venues as a whole and cannot be used to evaluate their temporal evolution. The Scopus<sup>30</sup> web application is a widely used online platform that offers several analytics regarding both researchers and scientific papers. However, it does not aggregate information on conference series. Lens.org [13] is another web application that integrates data from MAG, Crossref<sup>31</sup>, CORE [14], and PubMed<sup>32</sup>. It enables the analysis of several entities (e.g., authors, institutions, countries, journals, conferences, topics), but it is built on top of MAG and therefore shares the same limitations of Microsoft Academic Search. RelPath [10] leverages the citation network to identify experts in a certain domain that can act as reviewers of a target paper. The rationale behind the approach is that if a given paper shares similar scientific elements with some of its references, then the authors of such references can be considered experts. The approach may be

<sup>26</sup> SciGraph datasets - <https://sn-scigraph.figshare.com>

<sup>27</sup> AMiner - <https://www.aminer.org/>

<sup>28</sup> Scholia - <https://scholia.toolforge.org>

<sup>29</sup> Wikidata - <https://www.wikidata.org>

<sup>30</sup> Scopus.com - <https://www.scopus.com/>

<sup>31</sup> Crossref - <https://www.crossref.org/>

<sup>32</sup> PubMed - <https://pubmed.ncbi.nlm.nih.gov/>

extended at journal and conference level to suggest who can act as a programme committee member or co-editor of a journal special issues. SciKGraph is another approach that leverages semantic technologies and natural language processing techniques to identify research fields from research papers [27]. Given a dataset of papers, it finds their main concepts and creates a knowledge graph based on their co-occurrence in papers. Concepts are then clustered to show how a scientific area is organised. Likewise, it is straightforward to apply the same approach to journals and conferences for identifying similar papers through their topical characterisation.

All the mentioned systems allow only a coarse-grained analysis of the involved actors (e.g., conferences, journals, authors, organisations, countries, topics). Furthermore, they do not take into account how much a venue attracts commercial organizations or specific industrial sectors. Therefore, the original idea when building the AIDA Dashboard was to integrate different knowledge graphs with the goal of enhancing the set of available analytics and performing more fine-grained analyses.

## 7 Conclusions

We have illustrated the second version of the AIDA Dashboard, a system developed within SN to support the analysis and comparison of journals and conferences according to several metrics. The AIDA Dashboard is built on top of the Academia/Industry Dynamics Knowledge Graph, a large knowledge graph containing over 1.5B triples obtained by merging data of 25M papers conferences from Microsoft Academic Graph, Dimensions, DBpedia and GRID. This version greatly improves the first prototype [4] by offering i) journals in addition to conferences, ii) an advanced search functionality to browse, compare, and rank journals and conferences, iii) the Focus Areas Taxonomy, a new taxonomy of research areas that we have produced to classify research venues, and iv) the integration of the dashboard with the SN Data Cloud Infrastructure. We have carried out a user evaluation involving 10 users of which 5 SN editors and 5 researchers, obtaining excellent results.

The AIDA Dashboard is now freely available online and the underneath data can be downloaded as well.

In future work, we plan to further enhance the AIDA Dashboard according to the feedback from editors and researchers. In particular, we are working in collaboration with the SN Data Science team on improving the intuitiveness of the interface and widening the coverage by expanding to other research fields, starting with Engineering. We also plan to include new types of entities for the user to inspect and compare, such as countries, researchers, organisations, and scientific communities.

*Supplemental Material Availability:* The AIDA KG 3.0, the AIDA Schema, the AIDA Venues dataset 2.0, INDUSO, the Focus Areas Taxonomy, and the evaluation data are available at <https://w3id.org/aida>.

## References

1. Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., Dunkelberger, J., Elgohary, A., et al.: Construction of the literature graph in semantic scholar. In: Proc. of the 2018 Conference of the North American Chapter of the Assoc. for Computational Linguistics: Human Language Technologies. pp. 84–91. Association for Computational Linguistics (2018)
2. Angioni, S., Salatino, A., Osborne, F., Recupero, D.R., Motta, E.: AIDA: A knowledge graph about research dynamics in academia and industry. *Quantitative Science Studies* **2**(4), 1356–1398 (02 2022). [https://doi.org/10.1162/qss\\_a\\_00162](https://doi.org/10.1162/qss_a_00162), [https://doi.org/10.1162/qss\\_a\\_00162](https://doi.org/10.1162/qss_a_00162)
3. Angioni, S., Salatino, A., Osborne, F., Reforgiato Recupero, D., Motta, E.: Integrating knowledge graphs for analysing academia and industry dynamics. In: ADBIS, TPD and EDA 2020 Common Workshops and Doctoral Consortium. Springer International Publishing, Cham (2020)
4. Angioni, S., Salatino, A.A., Osborne, F., Recupero, D.R., Motta, E.: The AIDA dashboard: A web application for assessing and comparing scientific conferences. *IEEE Access* **10**, 39471–39486 (2022). <https://doi.org/10.1109/ACCESS.2022.3166256>, <https://doi.org/10.1109/ACCESS.2022.3166256>
5. Angioni, S., Salatino, A.A., Osborne, F., Recupero, D.R., Motta, E.: The AIDA dashboard: Analysing conferences with semantic technologies. In: Taylor, K.L., Gonçalves, R.S., Lécué, F., Yan, J. (eds.) Proceedings of the ISWC 2020 Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 19th International Semantic Web Conference (ISWC 2020), Globally online, November 1-6, 2020 (UTC). CEUR Workshop Proceedings, vol. 2721, pp. 271–276. CEUR-WS.org (2020), <http://ceur-ws.org/Vol-2721/paper570.pdf>
6. Aryani, A., Poblet, M., Unsworth, K., Wang, J., Evans, B., Devaraju, A., Hausstein, B., Klas, C.P., Zapilko, B., Kaplun, S.: A research graph dataset for connecting research data repositories using rd-switchboard. *Scientific Data* **5**(1), 180099 (2018). <https://doi.org/10.1038/sdata.2018.99>, <https://doi.org/10.1038/sdata.2018.99>
7. Brooke, J.: Sus: A 'quick and dirty' usability scale. *Usability evaluation in industry* **189** (1996)
8. Fenner, M., Aryani, A.: Introducing the pid graph (2019). <https://doi.org/10.5438/jwvf-8a66>
9. Groth, P., Gibson, A., Velterop, J.: The anatomy of a nanopublication. *Information Services & Use* **30**(1-2), 51–56 (2010)
10. Guilarte, O.F., Barbosa, S.D.J., Pesco, S.: RelPath: an interactive tool to visualize branches of studies and quantify the expertise of authors by citation paths. *Scientometrics* **126**(6), 4871–4897 (2021). <https://doi.org/10.1007/s11192-021-03959-2>, <https://doi.org/10.1007/s11192-021-03959-2>
11. Iana, A., Jung, S., Naeser, P., Birukou, A., Hertling, S., Paulheim, H.: Building a conference recommender system based on scigraph and wikicfp. In: International Conference on Semantic Systems. pp. 117–123. Springer, Cham (2019)
12. Jaradeh, M.Y., Oelen, A., Farfar, K.E., Prinz, M., D'Souza, J., Kismihók, G., Stocker, M., Auer, S.: Open research knowledge graph: Next generation infrastructure for semantic scholarly knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture. pp. 243–246 (2019)



13. Jefferson, O.A., Koellhofer, D., Warren, B., Jefferson, R.: The lens metarecord and lensid: An open identifier system for aggregated metadata and versioning of knowledge artefacts (Nov 2019). <https://doi.org/10.31229/osf.io/t56yh>, [osf.io/preprints/lissa/t56yh](https://osf.io/preprints/lissa/t56yh)
14. Knoth, P., Zdrahal, Z.: Core: three access levels to underpin open access. *D-Lib Magazine* **18**(11/12), 1–13 (2012)
15. Manghi, P., Atzori, C., Bardi, A., Baglioni, M., Schirrwagen, J., Dimitropoulos, H., La Bruzzo, S., et al.: Openaire research graph dump (Nov 2020). <https://doi.org/10.5281/zenodo.4279381>
16. Nielsen, F.Å., Mietchen, D., Willighagen, E.: Scholia, scientometrics and wikidata. In: Blomqvist, E., Hose, K., Paulheim, H., Lawrynowicz, A., Ciravegna, F., Hartig, O. (eds.) *The Semantic Web: ESWC 2017 Satellite Events*. pp. 237–259. Springer International Publishing, Cham (2017)
17. Nuzzolese, A.G., Gentile, A.L., Presutti, V., Gangemi, A.: Conference linked data: The scholarlydata project. In: *Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 9982 LNCS, pp. 150–158 (2016). [https://doi.org/10.1007/978-3-319-46547-0\\_16](https://doi.org/10.1007/978-3-319-46547-0_16)
18. Osborne, F., Motta, E.: Pragmatic ontology evolution: reconciling user requirements and application performance. In: *International Semantic Web Conference*. pp. 495–512. Springer (2018)
19. Peroni, S., Shotton, D.: OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies* **1**(1), 428–444 (02 2020). [https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023), [https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023)
20. Salatino, A., Osborne, F., Motta, E.: Cso classifier 3.0: a scalable unsupervised method for classifying documents in terms of research topics. *International Journal on Digital Libraries* **23**(1), 91–110 (2022)
21. Salatino, A.A., Osborne, F., Birukou, A., Motta, E.: Improving editorial workflow and metadata quality at springer nature. In: Ghidini, C., Hartig, O., Maleshkova, M., Svátek, V., Cruz, I., Hogan, A., Song, J., Lefrançois, M., Gandon, F. (eds.) *The Semantic Web – ISWC 2019*. pp. 507–525. Springer Int. Publishing, Cham (2019)
22. Salatino, A.A., Osborne, F., Motta, E.: Augur: Forecasting the emergence of new research topics. In: *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. pp. 303–312. JCDL '18, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3197026.3197052>
23. Salatino, A.A., Osborne, F., Thanapalasingam, T., Motta, E.: The cso classifier: Ontology-driven detection of research topics in scholarly articles. In: *Digital Libraries for Open Knowledge*. pp. 296–311. Springer International Publishing, Cham (2019)
24. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Birukou, A., Osborne, F., Motta, E.: The computer science ontology: A comprehensive automatically-generated taxonomy of research areas. *Data Intelligence* pp. 1–38 (2020). [https://doi.org/10.1162/dint\\_a.00055](https://doi.org/10.1162/dint_a.00055)
25. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: The computer science ontology: a large-scale taxonomy of research areas. In: *International Semantic Web Conference*. pp. 187–205. Springer (2018)
26. Thanapalasingam, T., Osborne, F., Birukou, A., Motta, E.: Ontology-based recommendation of editorial products. In: Vrandečić, D., Bontcheva, K., Suárez-Figueroa,

- M.C., Presutti, V., Celino, I., Sabou, M., Kaffee, L.A., Simperl, E. (eds.) *The Semantic Web – ISWC 2018*. pp. 341–358. Springer International Publishing, Cham (2018)
27. Tosi, M.D.L., dos Reis, J.C.: Scikgraph: A knowledge graph approach to structure a scientific field. *Journal of Informetrics* **15**(1), 101109 (2021). <https://doi.org/10.1016/j.joi.2020.101109>, <https://www.sciencedirect.com/science/article/pii/S175115772030626X>
  28. Wang, K., Shen, Z., Huang, C., Wu, C.H., Dong, Y., Kanakia, A.: Microsoft academic graph: When experts are not enough. *Quantitative Science Studies* **1**(1), 396–413 (2020)
  29. Wijkstra, M., Lek, T., Kuhn, T., Welbers, K., Steijaert, M.: Living literature reviews. arXiv preprint arXiv:2111.00824 (2021)
  30. Zhang, Y., Zhang, F., Yao, P., Tang, J.: Name disambiguation in aminer: Clustering, maintenance, and human in the loop. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 1002–1011 (2018)