

Data sources: Dimensions

- [Database content and quality assurance](#)
- [Basic functionalities of the database](#)
- [Limitations and criticism](#)
- [Other options for using Dimensions data](#)

[Dimensions](#) is the “largest database of linked research data in the world” produced by Digital Science Ltd., with a [browser interface](#) and an [API \(application programming interface or machine-readable interface\)](#). Digital Science also provides a number of other services, such as Altmetric, Figshare, Readcube and Overleaf. Digital Science was formerly part of the Nature group, but in the 2015 Springer Nature merger it was separated from the publishing company to a completely separate company. The company is owned by the Holzbrink Publishing Group, which also holds a majority stake in Springer Nature.

Database content and quality assurance

In addition to a citation network of scientific articles, Dimensions includes the linked data of datasets, research funding, patents, clinical trials and societal publications. In the spring of 2022, the database covered and linked together metadata from about 126 million scientific publications, 11 million datasets, 6 million funding cases, 144 million patents, 0.7 million clinical trials and 0.7 million policy publications (Table 1). The database includes a growing number of scientific publications starting from 1665 (Figure 1). In addition to articles, the indexed scientific publications include preprint publications, edited books, book chapters, monographs and conference publications.

Unlike its biggest competitors, Scopus and Web of Science, the basic content and functionalities of the Dimensions database is freely accessible to everyone online without registration or fees for their personal, non-commercial use. The section of Dimension that is available for free includes scientific publications and datasets. Additionally, anyone can register for a user account for free, allowing them to use functions such as saving searches and transporting search results to [VOSviewer](#) or [CiteSpace](#) programmes for further analysis.

The database also offers its paying customers linked metadata on research grant decisions (including the Academy of Finland), patents, clinical trials, and authority and administrative reports as well as more versatile tools for analysing the search results than the free version. Upon application, Dimensions can also grant [free personal user licences for research purposes](#), including for their content and tools that are usually subject to a charge. For example, the research and development work of the [CPR citation impact indicator](#) was done with the help of one such free user licence.

From the perspective of metrics research and production, the special strength of Dimensions is its API. The [Dimensions Search Language](#) used over the user interface offers a more versatile and effective channel for searching and processing the metadata needed for metrics than the database's competitors. The constructive approach of Digital Science, the owner of the database, to the users' needs to experiment and develop their own metrics solutions is also a noteworthy factor. [JYUcite](#), a service provided by the Open Science Centre of the University of Jyväskylä measuring the citation impact of publications, is an example of a service that produces metrics while utilising the Dimensions' API and the DSL searches (Seppänen 2020). DSL searches can also be carried out easily via the service's user interface with a web browser, which means that learning or using the DSL language does not require programming skills. All data is also available via Google's BigQuery environment, which facilitates analyses that require large-scale data mining.

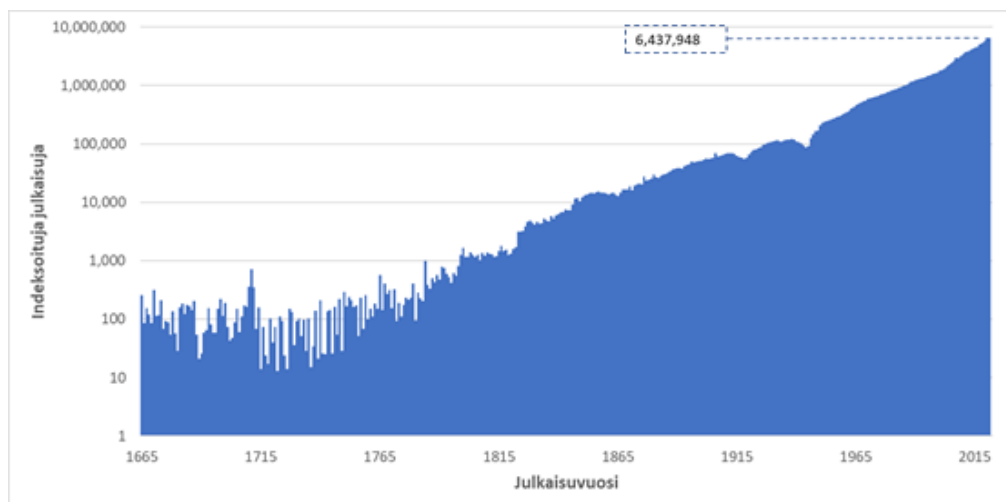


Figure 1. The number of scientific publication indexed in the Dimensions database by their publication year. Note that the scale is logarithmic.

Coverage across different fields of science

Generally, Dimensions has more inclusive coverage than Scopus or the Web of Science. With regard to the indexing of citations, Dimensions has slightly better coverage than the Web of Science and slightly poorer coverage than Scopus (overlapping with both by 75% of total citations found), by the estimate of Martín-Martín (2021). The differences become more obvious in the fields of humanities and social sciences than in natural or medical sciences (see also Visser et al. 2021). When reviewing the findability of publications produced by the University of Jyväskylä 2007–2019, Dimensions was better than its competitors in 2020, but — like the other databases — had poor coverage in social sciences and humanities in particular. Dimensions is utilising machine learning and data processing implemented with automated algorithms on a broad basis, instead of manual curation carried out by experts.

Regarding materials written in languages other than English, Dimensions had more coverage than its competitors, but it, too, had deficiencies.

With regard to edited books, their chapters and monographs, Dimensions has much greater coverage than Scopus or Web of Science.

The [introductory section](#) of this chapter has more information on the differences in coverage between the citation databases.

Collection policy

Dimensions mainly indexes all scientific publications and data sets with a DOI identifier, automatically without the publisher’s request and without a selection process carried out by a scientific editorial board. However, the [selection criteria](#) state that the sources indexed cannot be sources that have been identified as “predatory publishers”. Search results can be filtered based on certain journal catalogues so that the results only show sources indexed in the DOAJ register or in the PubMed database, for example.

In addition to the automatic indexing of metadata, a publisher can request the indexing of their content’s full text, provided that the publisher is able to [deliver the material in accordance with Dimension’s specifications](#). About 70% of all publications indexed by Dimensions have been full text indexed.

[Grant information](#) for large funders is collected by Dimensions directly from the data published by the funders (e.g. The Academy of Finland) and for smaller operators by the machine interpretation of the Acknowledgements section of the publications, for example.

[Patent metadata](#) is collected for the database in national databases around the world as well as the European Patent Office’s database.

[Metadata on clinical trials](#) come from 12 different official medical registers.

[Metadata for policy publications](#) is collected by Dimensions from the websites of about 200 different organisations and national agencies. However, this data has limitations, as more than 80% of the indexed sources are in the UK, Switzerland (e.g. UN organisations) or the USA. ([Overton](#) offers a much better tool for finding policy publications and performing citation analysis on them).

Publication types

Table 1. The number of different publication types in the Dimensions database in the spring of 2022.

Scientific publications	125,628,854
<i>Articles</i>	101,830,235
<i>Book chapters</i>	12,205,558
<i>conference articles</i>	6,954,426
<i>Pre-prints</i>	3,180,552
<i>Monographs</i>	849,252
<i>Edited books</i>	608,424
Data sets	11,231,163
Funding data	6,141,928
Patents	143,597,975
Clinical trials	690,017
Policy publications	743,138

Database updates

The content of Dimensions is updated daily with regard to publications, publication metrics, data sets and clinical trials, and weekly with regard to patents and policy publications. Grant decisions are updated once a month. The publications are usually indexed within two to three days of the publisher saving the metadata in a widely used register, with CrossRef as the key database.

Classification of fields of science

Dimensions offers [several methods](#) for searching and filtering publications based on their field of science. In most cases, the classification is done at article level, not at publication series level: Each article is given its own classification with an algorithm based on machine-learning, which uses the article's full text or its abstract for the classification. If there is not a sufficient amount of the full text of the article available, the classification of the publication series' field of science will be used instead.

Extensive classifications covering all fields of science in Dimensions:

- *Field of Research (FoR)*: applies the [Australian and New Zealand Standard Research Classification](#) (ANZSRC) hierarchical classification system and its two middle levels with 22 upper levels including a total of 157 groups.
- *Units of Assessment (UoA)*: A division with 34 classes used in the national research assessment of the UK.
- *Sustainable Development Goals (SDGs)*: The UN's 17 goals for sustainable development.

Publications in the fields of medical and health sciences can also use the following classifications as applicable:

- *Research, Condition, and Disease Categorization (RCDC)*: The U.S. Department of Health's classification of research domains in biomedical fields.
- *Health Research Classification System (HRCS)*: The research field and health classification of funders of medical sciences in the UK.
- *Broad Research Areas (BRA)*: Four broad classes used by the Australian funder of medical science.
- *Health Research Areas (HRA)*: Dimension's own division into four broad classes to separate the continuum from basic research to the clinical care and questions of national health.
- *Common Scientific Outline (CSO)*: The international classification of cancer research.

Basic functionalities of the database

Search options

The default basic search, which is the only option available in the free version, is very simple and similar to the most modern online search interfaces: write something in the search field and click Enter. By default, the search is targeted at all data available: about 70% of the articles recognised by Dimensions have been full text indexed, which means that the article content will also be searched, including using concepts determined by artificial intelligence, so the search simply does not look for keywords.

If desired, the search can be limited to the title or abstract instead of all content, or just the DOI.

The paid version of Dimensions also offers:

- A larger selection of filters for filtering search content.
- An advanced search function that supports the use of logical operators (AND, OR, NOT), parenthesis, wildcard characters and proximity search.
- AI search based on similarities in text: you can paste even a long piece of text into the search field, such as the entire introduction of an article, and the AI will determine certain concepts from the text and recover records that contain similar concepts.
- [Dimensions Search Language](#), which can be used to carry out complex and effective user interface queries in the database without any programming expertise, directly via the browser's user interface.

Filtering search results using factors such as the year of publication, the name of the researcher, publication series, research area, funder, research organisation, country, or the type of publication is simple by using the menu next to the results. Filters can also be used directly for starting a search, without entering a search term.

Search results can be arranged by the search terms' relevance, date of publication, [RCR](#) index, [FCR](#) index, number of citations or their [Altmetric](#) index. The RCR and FCR indexes in particular, which strive for field and time-normalisation, are often a quick way to find the publications with the best information on the subject from a large amount of search results.

The publication authors, organisations, funders, publication series, list of sources and citing publications can all be found in the publication's information as direct links.

The search options are [described in detail on Dimension's own website](#).

Dimensions' analysis tools

See the chapter [Dimensions Analytics](#) of this guide.

Researcher profile in Dimensions

Dimensions aims to differentiate researcher information into individuals using algorithms and based on several sources. This means that a researcher does not need to — and they cannot — determine their profile in the Dimensions database by themselves. The profile presents a view of the researcher created by the algorithm: their affiliations, publications, datasets, funding, patents, research areas, the key terms of their research and their partner network.

Mechanical processing allows for a massive network to be created, but this comes at the price of occasional deficiencies and inaccuracies. In particular, those researchers who have not registered for an ORCID account for themselves and who have a fairly common surname, meaning that another researcher with the same name may work at the same university, may become mixed with each other in Dimensions profiles.

Organisation profiles in Dimensions

The Global Research Identifier Database (GRID) system developed and maintained by Digital Science for permanent identification, differentiation and linking of research organisations works in the background of the Dimensions database.

However, a new, community maintained Research Organization Registry (ROR) has been developed on the foundation built by GRID, with the support of Digital Science. This registry works independently of any commercial operators and it is now replacing GRID as the publicly shown ID. GRID and ROR can differentiate between more than 100,000 different organisations.

Limitations and criticism

At the same time, the significant strength of Dimensions is also its weakness: data processing based on machine learning and automated algorithms enables greater scalability and faster updating of data than human content production and quality assurance. In addition, full text concepts based classification and search organizes and groups research publications directly, without relying on a serial level classification. On the flip side, there are unexpected errors in the content, and the opacity and repeatability of the methods are lost in the algorithms' black box.

For example, publications written by researcher "null" with the title "null" have been found in the database. These publications have been cited very frequently, because sources that were not found elsewhere have been interpreted as citations to this unnamed pseudo-publication. Additionally, thousands of citations were found to the brief definition of the word "small" in a certain dictionary, because the publisher of the dictionary had, for some reason, registered a DOI for the word, and there is an existing journal named "Small", and some citations to that were linked to the dictionary entry.

However, in relation to the massive size of the database, the number of such errors is fairly low. In a regular information search, these kinds of mistakes are also obvious to the user and are thus easy to ignore. When processing large quantities of search results for the purposes of publication metrics, however, a sanity check should be performed in order to weed out any mistakes and ensure the integrity of the source data before using it. Do the search results include publications that have an unreasonably high number of citations, either as is or in relation to the publication's age, for example? Do the search results have publications where the title, the name of the publication series or the author's name are just a couple of characters long?

The machine-generated funder information derived from the Acknowledgements sections fairly often has mistakes.

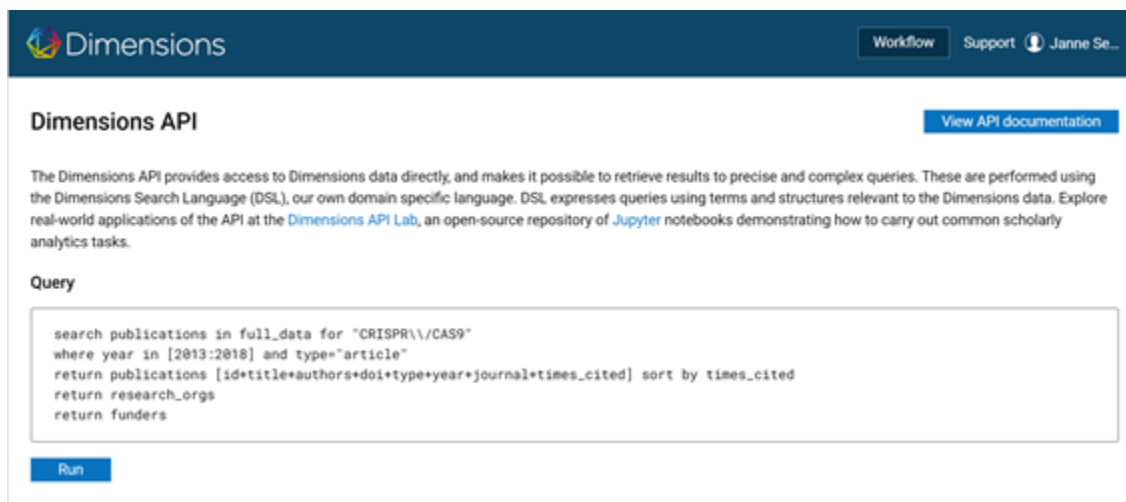
As in other databases, the coverage of humanities and social sciences in the Dimensions database, especially when it comes to research not written in English, is lacking.

Other options for using Dimensions data

The special strength of Dimensions is its powerful API: [Dimensions Search Language](#) interface query. This is available to the subscribers of [Dimensions Analytics](#) tool.

The DSL interface query, for example, allows users to retrieve a list of all the articles presented in the reference lists of articles citing the subject article — with just one query ("find publications that cite this target article and return their reference lists"). Or you can retrieve the information on up to 400 articles with just one query ("find all publications with DOI presented in this list"). Additionally, you can use the same query to search for not only publications, but also funding decisions, research organisation information, patents and author information related to them — or vice versa.

Users can learn to use queries even without prior programming skills using a feature available in the browser interface, and the DSL documentation is easy to understand and offers practical examples. Utilising the search results provided in JSON format on a larger scale requires some data processing, but tools are available for this even in Excel (Excel Power Query).



Dimensions API [View API documentation](#)

The Dimensions API provides access to Dimensions data directly, and makes it possible to retrieve results to precise and complex queries. These are performed using the Dimensions Search Language (DSL), our own domain specific language. DSL expresses queries using terms and structures relevant to the Dimensions data. Explore real-world applications of the API at the [Dimensions API Lab](#), an open-source repository of [Jupyter](#) notebooks demonstrating how to carry out common scholarly analytics tasks.

Query

```
search publications in full_data for "CRISPR\\CAS9"
where year in [2013:2018] and type="article"
return publications [id*title*authors*doi*year*journal*times_cited] sort by times_cited
return research_orgs
return funders
```

[Run](#)

Figure 2. An example of a DSL interface query in Dimensions' browser interface.

Sources

Digital Science (2022) *Dimensions AI*. Available: <https://www.dimensions.ai/> [Accessed 13.6.2022]

Digital Science (2022) *Dimensions APIs*. Available: <https://www.dimensions.ai/dimensions-apis/> [Accessed 13.6.2022]

Digital Science (2022) *Dimensions käyttöliittymä*. Available: <https://app.dimensions.ai/discover/publication> [Accessed 13.6.2022]

Digital Science (2022) *Dimensions scientometric research*. Available: <https://www.dimensions.ai/scientometric-research/> [Accessed 13.6.2022]

Digital Science (2022) *Dimensions Search Language. Version 2.1.0*. Available: <https://docs.dimensions.ai/dsl/> [Accessed 13.6.2022]

Martín-Martín, A., Thelwall, M., Orduna-Malea, E. and López-Cózar, E.D. (2021) Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, 126, pp. 871–906 (2021). Available: <https://doi.org/10.1007/s11192-020-03690-4>

Visser, M., van Eck, N.J. and Waltman, L. (2021) Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1): pp. 20–41. Available: https://doi.org/10.1162/qss_a_00112

Seppänen, J.T. (2020) *Source code and example data for article: Co-Citation Percentile Rank and JYUcite: a new network-standardized output-level citation influence metric*. JYX dataset. <https://doi.org/10.17011/jyx/dataset/71858>