Project Name	**FREYA**
Project Title	**Connected Open Identifiers for Discovery, Access and Use of Research Resources**
EC Grant Agreement No	**777523**

D2.4 DOI Search Service (Common DOI Search)

Deliverable type	Other
Dissemination level	Public
Due date	31 October 2020
Authors	Martin Fenner (DataCite) Richard Hallett (DataCite)
Abstract	A common DOI search service has been developed and launched under the name “DataCite Commons”. This report describes the development, the current status and the future outlook of the service.
Status	Submitted to EC 2 November 2020

The FREYA project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 777523.

FREYA project summary

The FREYA project iteratively extends a robust environment for Persistent Identifiers (PIDs) into a core component of European and global research e-infrastructures. The resulting FREYA services will cover a wide range of resources in the research and innovation landscape and enhance the links between them so that they can be exploited in many disciplines and research processes. This will provide an essential building block of the European Open Science Cloud (EOSC). Moreover, the FREYA project will establish an open, sustainable, and trusted framework for collaborative self-governance of PIDs and services built on them.

The vision of FREYA is built on three key ideas: the **PID Graph**, **PID Forum** and **PID Commons**. The PID Graph connects and integrates PID systems to create an information map of relationships across PIDs that provides a basis for new services. The PID Forum is a stakeholder community, whose members collectively oversee the development and deployment of new PID types; it will be strongly linked to the Research Data Alliance (RDA). The sustainability of the PID infrastructure resulting from FREYA beyond the lifetime of the project itself is the concern of the PID Commons, defining the roles, responsibilities and structures for good self-governance based on consensual decision-making.

The FREYA project builds on the success of the preceding THOR project and involves twelve partner organisations from across the globe, representing PID infrastructure providers and developers, users of PIDs in a wide range of research fields, and publishers.

For more information, visit www.project-freya.eu or email info@project-freya.eu.

Executive summary

In this report we describe a Common DOI Search service that the FREYA project partners have launched in October 2020 under the name "DataCite Commons"[1]. The service allows the discovery of content registered with a DOI from FREYA partners Crossref and DataCite in a single search interface and using a common search index. DataCite Commons also supports searching for people, research organizations and funders using the ORCID and ROR services, respectively. As of October 2020, a total of 38 million persistent identifiers can be discovered using the DataCite Commons service. In addition, DataCite Commons shows the connection between works in the form of citations as well as the researchers, research organizations and funding associated with these works.

The work on DataCite Commons was driven by over 40 user stories identified early in the FREYA project and made publicly available for community feedback. These user stories were difficult to address with existing infrastructure and require flexible queries for scholarly outputs, the researchers creating them, the research organizations supporting them, and the multiple connections between all of them. To support this so called PID Graph, one of the major outputs of the FREYA project, FREYA partners built a query API using the GraphQL technology. DataCite Commons is a public web interface for this GraphQL API.

FREYA partner DataCite is committed to maintaining and improving the GraphQL API and DataCite Commons web interface, and will be adding more PIDs, more connections and more functionalities that make DataCite Commons and the PID Graph an important part of the European Open Science Cloud (EOSC) infrastructure.

[1] https://commons.datacite.org

Contents

1 Introduction

This report describes the FREYA work on launching a common DOI search service, called DataCite Commons[2], allowing the discovery of all content registered with a DOI, no matter which content type or DOI registration agency (starting with FREYA partners Crossref and DataCite), in a single place.

DataCite Commons is not only a service to search for DOIs from multiple DOI registration agencies, but also allows searching for all ORCID IDs and Research Organization Registry (ROR) IDs and their associated metadata. Additionally DataCite Commons shows the connections between DOIs in the form of citations, and connections to people (identified by ORCID IDs), research organizations and funders (both identified by ROR IDs[3]). The DataCite Commons service is built on top of the DataCite GraphQL API that the FREYA project launched in May 2020, and that is described in detail elsewhere[4].

DataCite Commons is ongoing work, and as of October 2020 only contains a subset of all DOIs, citations between DOIs, and connections to people and organizations. The service provides a statistics page that shows live data on the number of PIDs, metadata and connections included. DataCite Commons will be maintained, more content and connections added, and further developed by DataCite and partners beyond the duration of the FREYA project in November 2020. The service will replace the existing DataCite Search[5] in 2021, and is made available as a discovery and reporting platform with the European Open Science Cloud (EOSC) and beyond.

[2] https://commons.datacite.org
[3] Ferguson, C., Lambert, S., Llinares, M. B., Madden, F., Dasler, R., Fenner, M., Lavasa, A., Baars, C., Dohna, T., Koop-Jacobsen, K., & Morgan, D. (2020). Deliverable 4.4 Organizational IDs in Practice. https://doi.org/10.5281/ZENODO.3606060
[4] Fenner, M. (2020). The PID Graph in FREYA (additional project report). https://doi.org/10.5281/ZENODO.4028383
[5] https://search.datacite.org

2 Planning work

2.1 Use cases

All work on developing new services or enhancing existing services in the FREYA project is driven by the needs of the users of our services. This is also true for the work described in this document, the common DOI search that we launched as the DataCite Commons service in October 2020.

Most researchers and other users of DOI-based discovery services do not know that there are eight different DOI registration agencies for scholarly content[6] who all offer their own discovery services but may differ by content types that are mainly registered (e.g. datasets vs. publications) and/or the geographic region they serve. When the FREYA project started, there was no single place that would allow the discovery of DOIs from all DOI registration agencies in a single service. One consequence of this situation is that the discovery of some content types – e.g. research data and research software – is more difficult.

The second driver for the development of the DataCite Commons service is the FREYA work on the PID Graph and the important user stories we have identified in this context. This work is described elsewhere[1] in more detail, but briefly, that report on the PID Graph identified three important general use cases:

Reuse across versions and parts

Datasets and software (and to a lesser degree publications) are frequently versioned, and datasets can often be downloaded as subsets. Tracking the reuse (views, downloads, and citations) across versions and parts is a frequent use case and can lead to confusion in the community regarding proper versioning. An important publications user story is linking preprints and peer-reviewed publications.

Reuse of aggregated research outputs

This might be the largest category of PID Graph user stories. We want to have a summary view of the reuse (via views, downloads, and citations) of all research outputs by a particular researcher, academic institution, data repository, or funder. This summary view can help demonstrate the impact of for example a researcher or repository.

Research objects

Aggregate all scholarly resources that are linked together via a single publication, including underlying data and software used to generate the results, but also people, organizations, and funding involved in the work. Exploring the connections in and to a research object is currently very difficult, e.g. starting from a dataset included in a research object, getting a list of publications that indirectly cite this dataset by citing the publication based on the data.

2.2 Technical architecture and implementation

The work on the technical architecture of DataCite Commons service started in May 2019. One of the important architecture decisions was whether to implement the common DOI search as a federated search, or to build a central search index for all DOIs. After several discussions within the FREYA team, in particular between Crossref and DataCite, we decided to build a central search index, as this would allow for much more sophisticated search results, e.g. relevance ranking for search results. The next important question was the metadata format this common DOI index would use. As DataCite would be hosting this common DOI search (Crossref being is an unfunded FREYA partner), and also has been working extensively for

[6] Listed at https://www.doi.org/registration_agencies.html

several years on mapping metadata between different formats[7] – including Crossref XML to DataCite XML – we decided to use the DataCite Metadata format for the common DOI search index.

The API that powers the search system is built on top of GraphQL, an open source API standard designed to empower advanced querying and retrieval of specific information from a variety of connected sources. The API is exposed via one endpoint, this is then backed by an ElasticSearch index for DOI related works (including both CrossRef and DataCite) and also combines external data from ORCID, ROR, Wikidata and Unpaywall. In the future other data can be sources which can then be combined as part of the GraphQL API and exposed.

To take advantage of the FREYA PID Graph work that led to the launch of a GraphQL API to query the PID Graph, DataCite Commons was built as a web interface on top of the DataCite GraphQL API[8], using the popular Apollo Client open source library[9] and the React open source Javascript framework[10]. One of the reasons React was chosen is because of its widespread adoption that gives longevity to supporting DataCite Commons well into the future.

Additionally one of the features of React is being able to compartmentalize functionality within components, this aids rapid development of shared functionality, which we then can take advantage of throughout the interface. An example of this is the shared works display that can be seen on various pages, this is one shared component made up of multiple sub components, illustrated in the following screenshot (Figure 1) with the various components highlighted in order of red, yellow, green to signify the different layers.

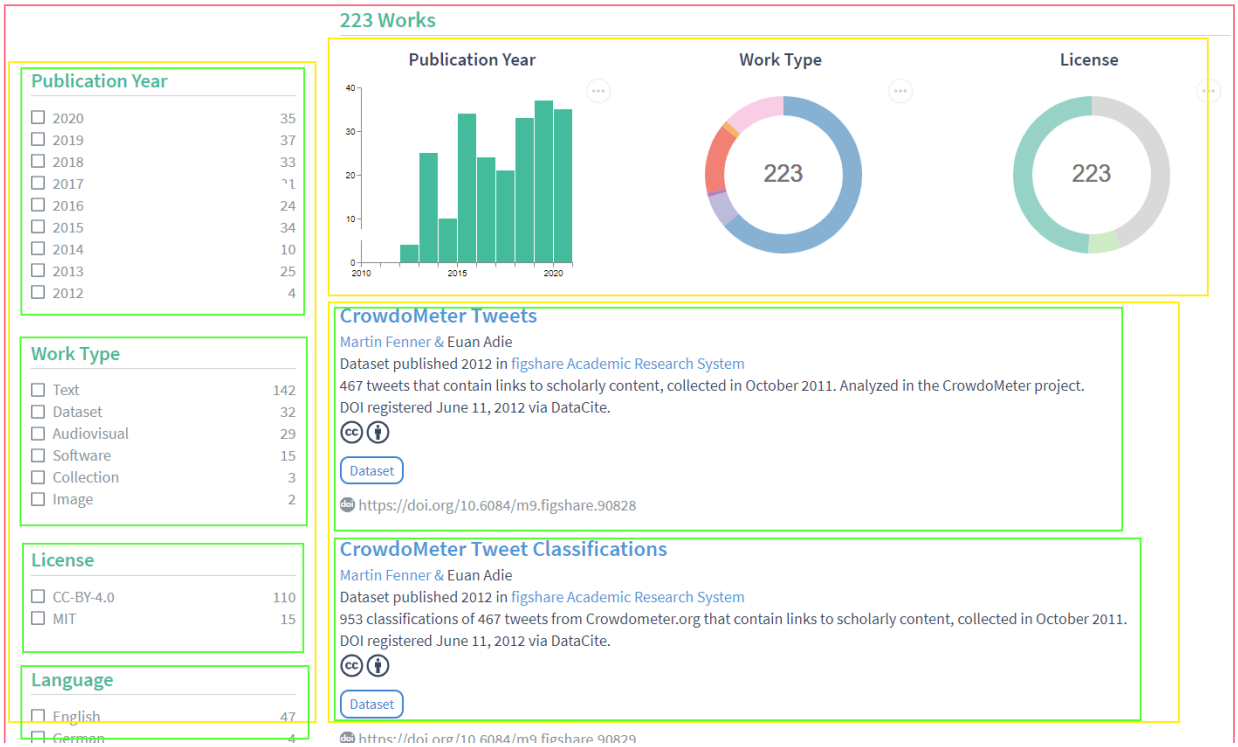


Figure 1 The DataCite Commons user interface is built using multiple components

[7] Fenner, M. (2017). Bolognese: a Ruby library for conversion of DOI Metadata. DataCite. https://doi.org/10.5438/N138-Z3MK

[8] DataCite. (2020). DataCite GraphQL API. DataCite. https://doi.org/10.25495/XKYA-0G76

[9] https://www.apollographql.com/docs/react/api/core/ApolloClient/

[10] https://reactjs.org/

Both the frontend application (akita[11]), and the backend application (lupo[12]) are released under an MIT open source license, and the source is available via GitHub.

2.3 Survey

Wrapping up the planning work was a survey that the FREYA project sent out in May 2020 to confirm that what we had planned to build resonates with the potential users of the new service, and to learn more about specific features users were interested in. We received feedback from 78 survey participants, the complete survey results can be found in Appendix 5.1, and are summarized below.

The survey confirmed earlier findings that there is a strong desire to have a single place to search for all scholarly DOIs:

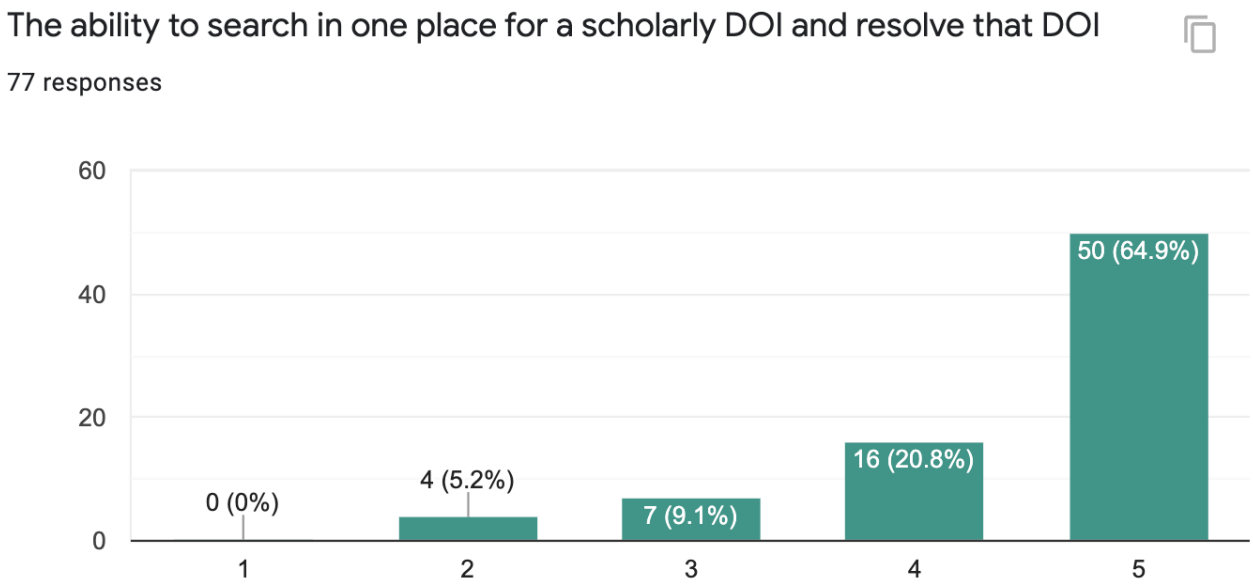


Figure 2 Survey responses. Response: 1- not important, 5- very important

It was deemed important by survey participants that they could not only search for DOIs, but also any metadata field, discover other PIDs such as ORCID or ROR IDs, and export the metadata from their search results. A visual representation (a graph) of how the PIDs in the search results are connected to each other was deemed less important.

There were 40 answers to the open feedback question on a variety of topics. We will refer to some of the answers later in the document, either in the **Results** section for functionality we have implemented, or in the **Future** section on suggestions that need further work beyond the duration of the FREYA project.

Taken together the feedback we received confirmed the work we had planned on common DOI search:

1. there is a clear need for a common DOI search irrespective of content type and registration agency;
2. there is a clear need to explore the FREYA PID Graph not only via an API, but also via an easy to use web interface.

[11] Fenner, M. (2020). Frontend for the DataCite Commons service. In DataCite (Version 1.0.4) [Computer software]. DataCite. https://doi.org/10.14454/QGK4-ZS88

[12] Fenner, M., & Garza, K. (2018). DataCite Application API. DataCite. https://doi.org/10.5438/8GB0-V673

3 Results

3.1 Common DOI search

DataCite Commons provides a single search interface for all content (**works**) registered with a DOI from either DataCite or Crossref. The search interface can be queried using one or more DOI to obtain more information about the content registered for DOIs, or the query can be a keyword query, e.g. **climate**. The keyword can be searched in all fields, or in specific fields, e.g. **titles.title:climate**.

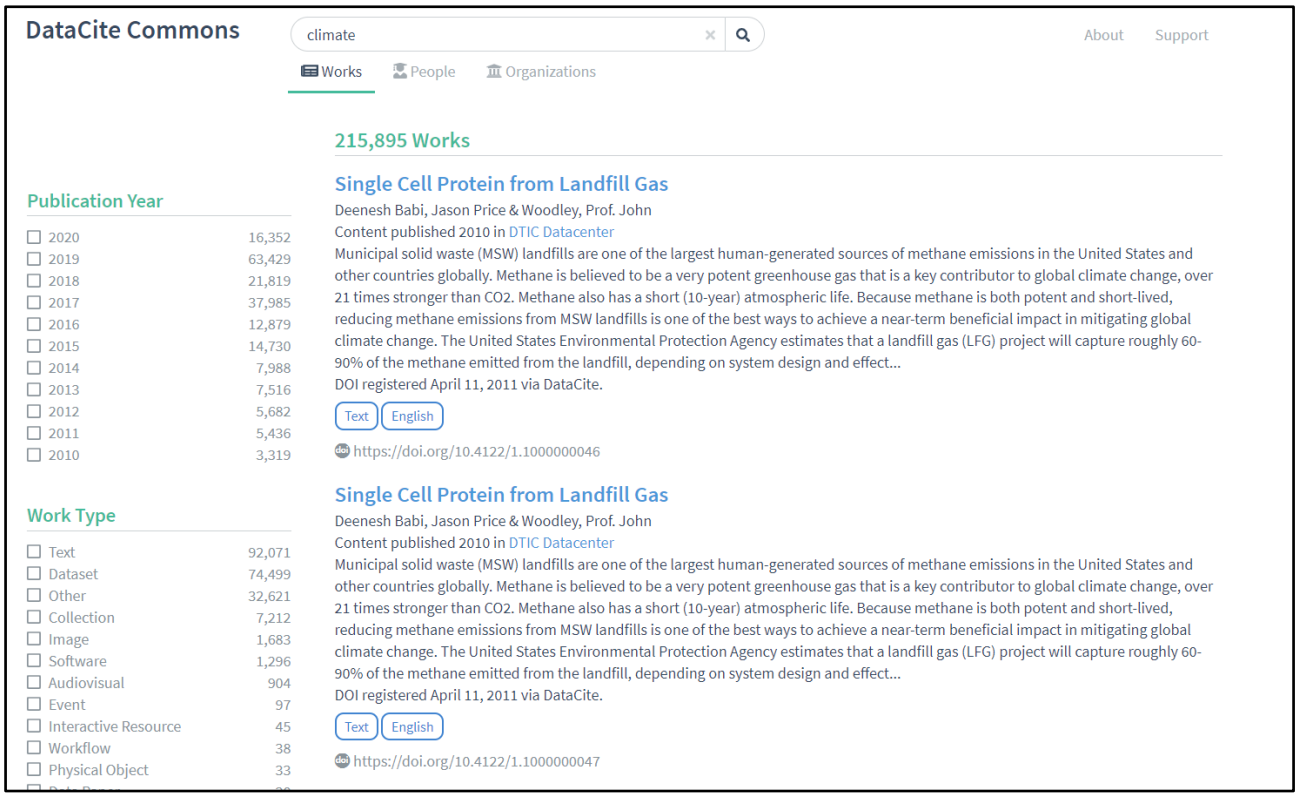


Figure 3 Search for the keyword "climate" in DataCite Commons. URL https://commons.datacite.org/doi.org?query=climate

Fields can also be queried by number or date range, e.g. **publicationYear:[2016 TO 2020]**, and multiple keywords can be combined, e.g. **titles.title:climate AND publicationYear:[2016 TO 2020]**. Because all DOIs are in the same search index, sophisticated queries similar to the examples above become possible, which would be impossible in a federated search. DataCite Commons allows the faceting (filtering) of search results, for example by publication year or work type, but also field of science, license and of course DOI registration agency. The search options for DataCite Commons are documented at the DataCite Support site[13].

Not only does DataCite Commons provide a unified search interface for DOIs from multiple DOI registration agencies, but the results are also harmonized as they come from a single search index using the DataCite Metadata schema[14]. Besides the required and recommended metadata such as title, description, publisher,

[13] https://support.datacite.org/docs/datacite-commons

[14] DataCite Metadata Working Group. (2019). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data v4.3. https://doi.org/10.14454/7XQ3-ZF69

publication year and license, DataCite Commons also shows detailed information about creators, contributors and funding, and linking to people (via ORCID ID), or organizations (ROR ID or Crossref Funder ID) if this information is available (links in blue).

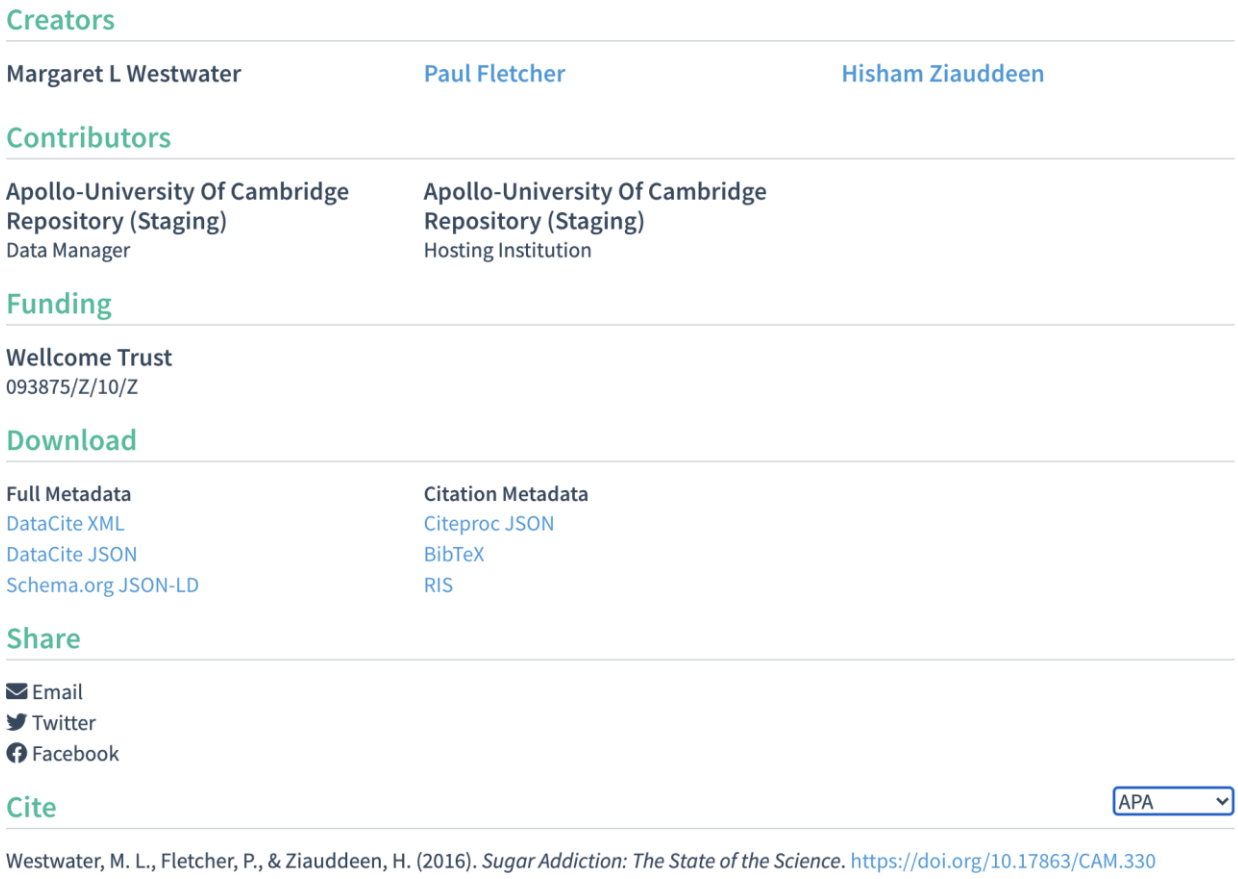


Figure 4 Creators, contributors and funding information for a single DOI in DataCite Commons

In addition, metadata in a variety of metadata formats, and a formatted citation using the thousands of citation styles available via the open source citation style language[15] are provided for each DOI.

3.2 Search for people or organizations

Similar to the queries for works, the DataCite Commons search interface allows queries for people or organizations, again using a persistent identifier or keyword. These queries go directly to the respective APIs and thus differ slightly in the advanced functionality, e.g ORCID does not support faceting (filtering) of search results, and metadata beyond the ORCID ID, name and affiliation only when looking up a single record.

[15] https://citationstyles.org/

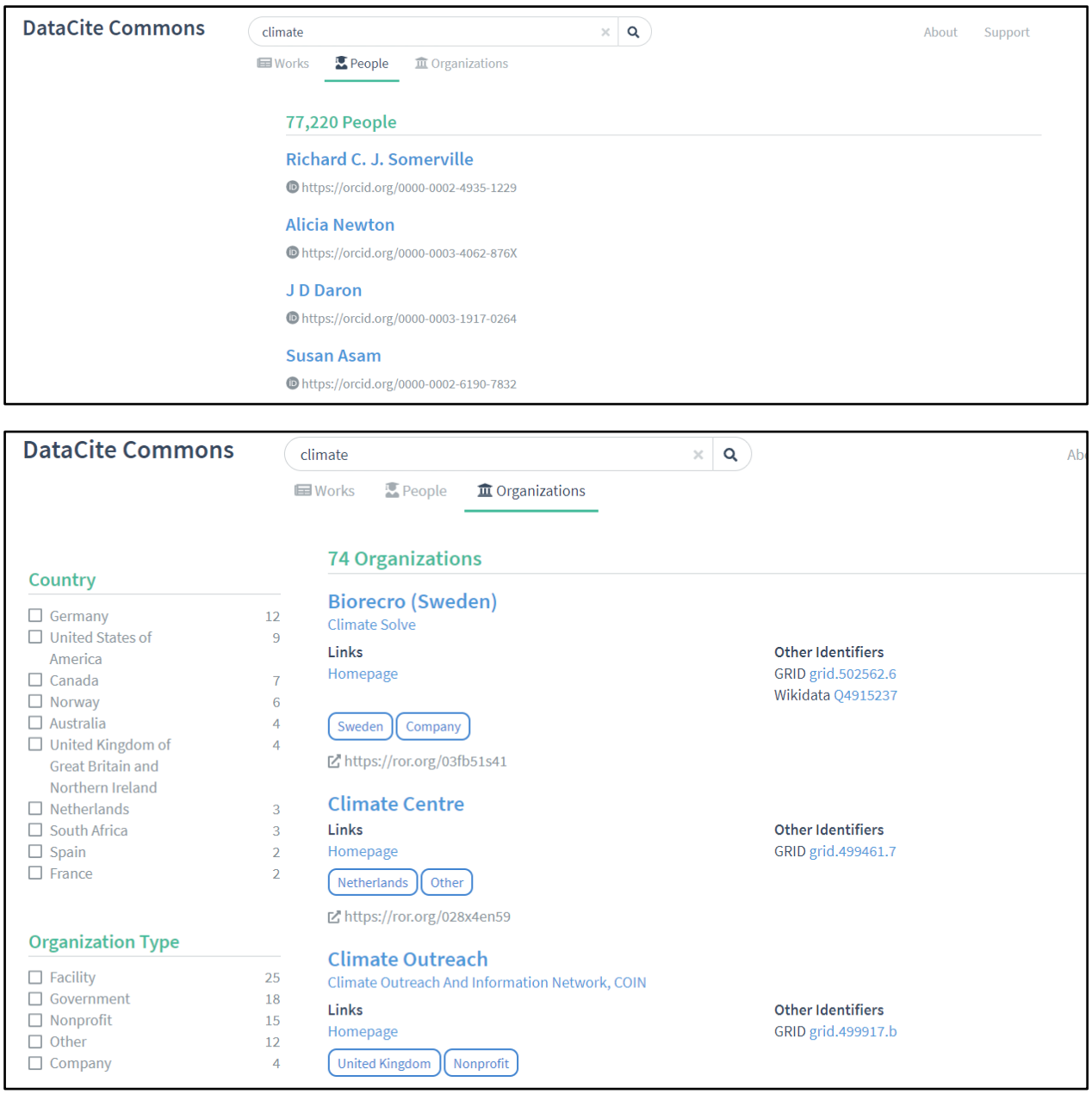


Figure 5 Search for people or organizations via a common search interface.

Taken together, DataCite Commons provides a simple and standardized search interface for works, people and organizations.

3.3 Citations, views and downloads for research data and software

One important category of PID Graph use cases demonstrates the reuse of research data and research software via citations, views and downloads, and DataCite Commons is addressing these use cases, based on work in the Research Data Alliance (RDA) Scholix initiative[16] and the Make Data Count (MDC) project[17]. One example is shown in Figure 6.

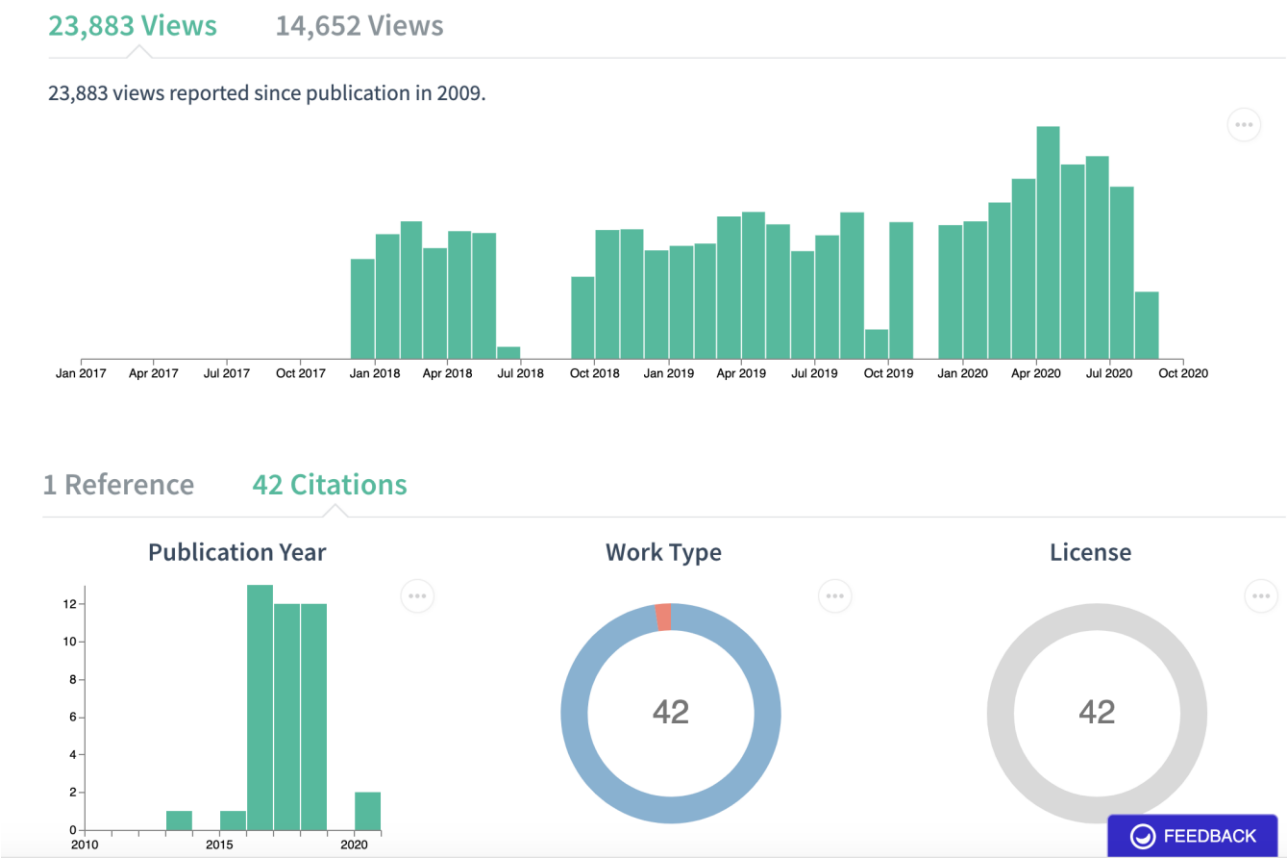


Figure 6 Citations, views, and downloads for a dataset in the Dryad data repository [18], shown in DataCite Commons[19].

[16] Burton, A., Aryani, A., Koers, H., Manghi, P., La Bruzzo, S., Stocker, M., Diepenbroek, M., Schindler, U., & Fenner, M. (2017). The Scholix Framework for Interoperability in Data-Literature Information Exchange. D-Lib Magazine, 23(1/2). https://doi.org/10.1045/JANUARY2017-BURTON

[17] Fenner, M., Lowenberg, D., Matt, J., Needham, P., Vieglas, D., Abrams, S., Cruse, P., & Chodaki, J. (2018). Code of practice for research data usage. Metrics release 1. Zenodo. https://doi.org/10.5281/ZENODO.3340590

[18] Zanne, A. E., Lopez-Gonzalez, G., Coomes, D. A., Ilic, J., Jansen, S., Lewis, S. L., Miller, R. B., Swenson, N. G., Wiemann, M. C., & Chave, J. (2009). Data from: Towards a worldwide wood economics spectrum (Version 5) [Data set]. Dryad. https://doi.org/10.5061/DRYAD.234

[19] https://commons.datacite.org/doi.org/10.5061/dryad.234

3.4 Aggregated research outputs and reuse

Aggregation of research outputs by researcher, research organization or funder, and combining the citations, views and downloads for these research outputs are important PID Graph use cases, and have been implemented in DataCite Commons. One important category is the aggregation of research outputs by researcher, showing not only all research outputs (if they could be linked via DOI and ORCID ID), including publications, datasets, software and others, in a single place, but also generating aggregated statistics, such as percentage of research outputs with an open license, or combined citations, views and downloads of all research outputs per person.

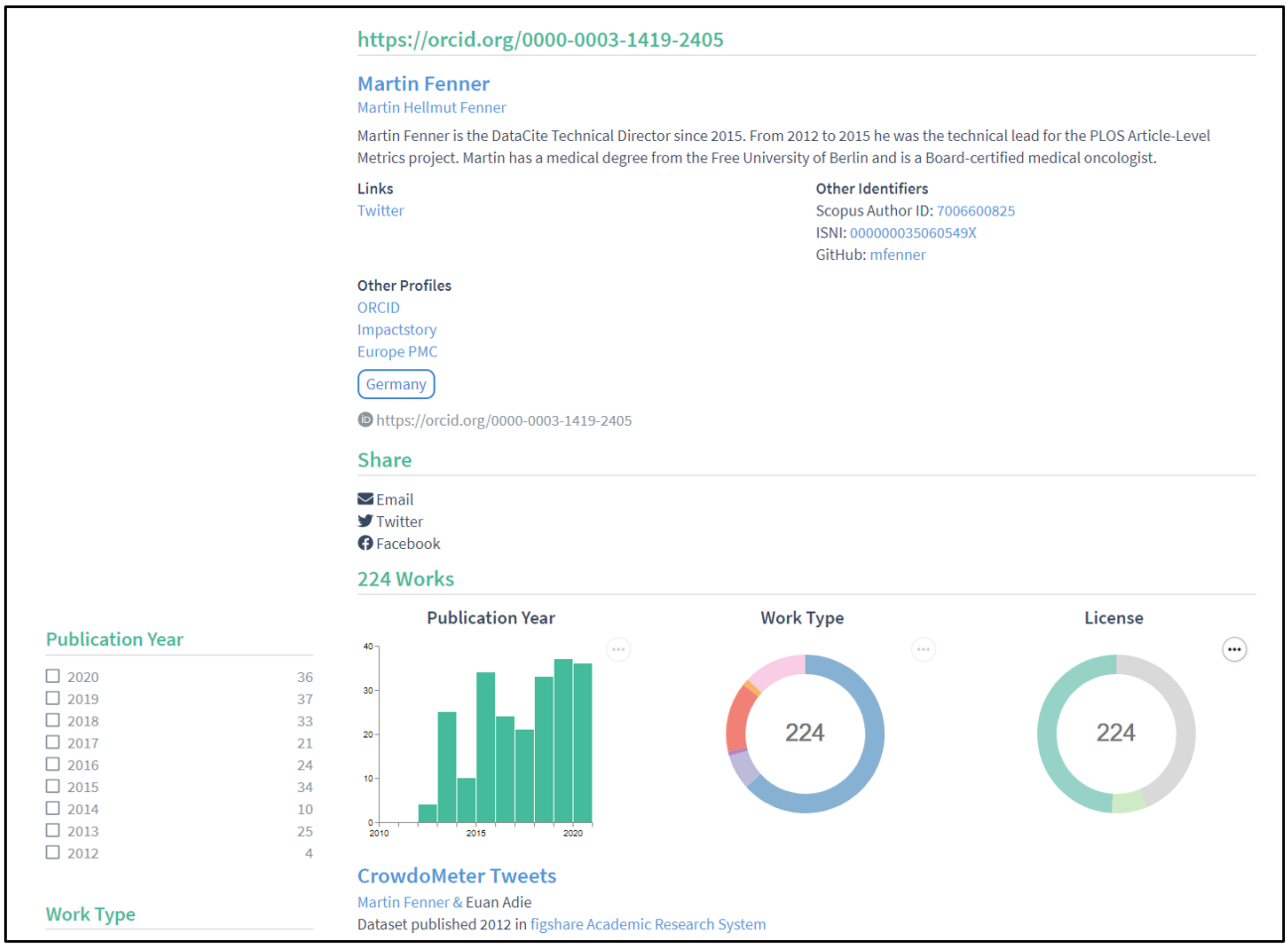


Figure 7 Aggregated research outputs by researcher, including facets for filtering and visualizations for the aggregations.

A similar approach can be taken for research organizations and funders. And the aggregation can be filtered using the existing search facets, or a query string. We can for example search for all European Commission-funded research outputs that were funded as part of the FREYA grant**,** using the following query string: https://commons.datacite.org/ror.org/00k4n6c32?query=fundingReferences.awardNumber%3A777523

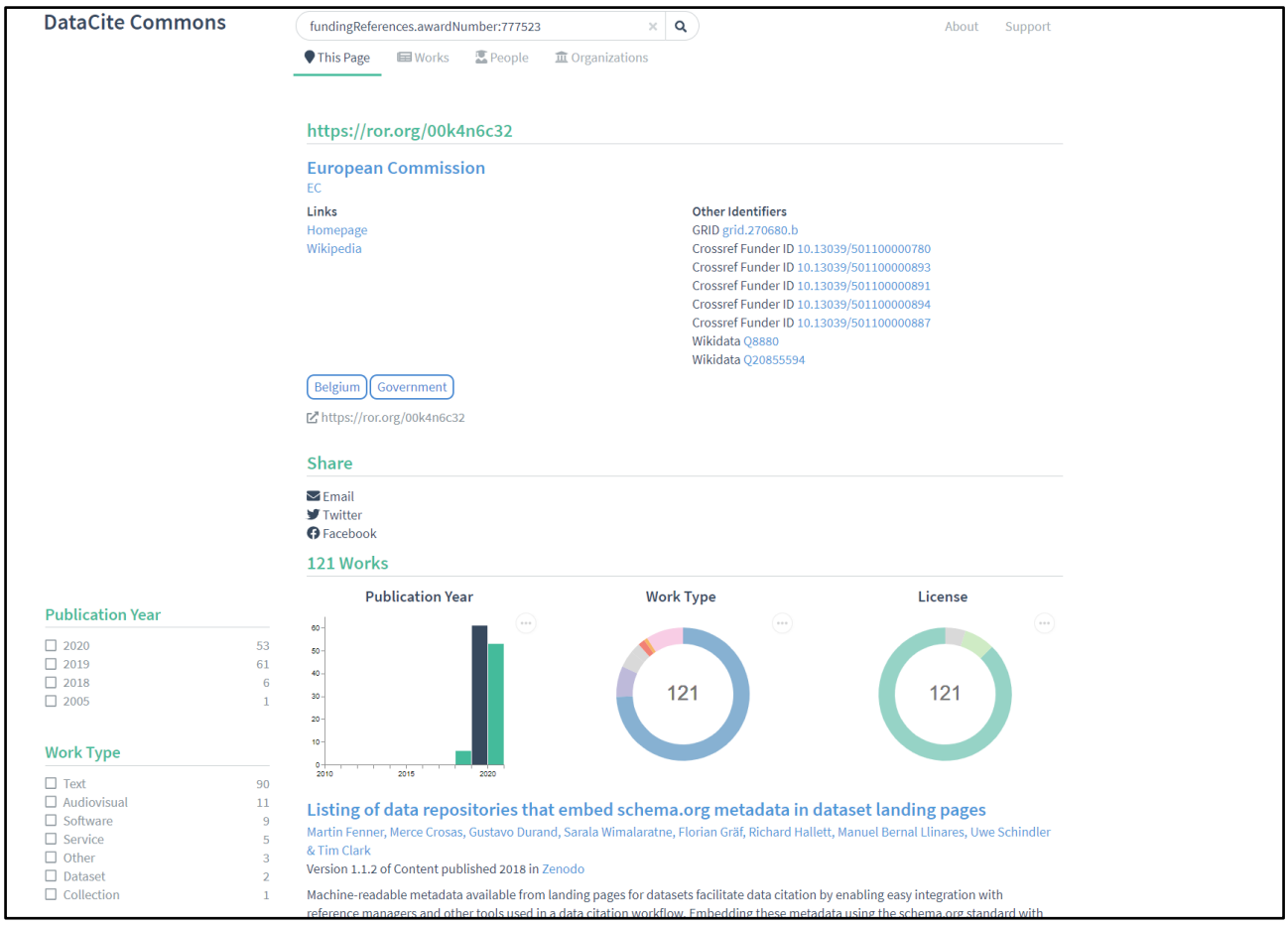


Figure 8 Research outputs funded via the FREYA grant

These examples demonstrate that DataCite Commons is addressing important PID Graph use cases that were identified at the beginning of the FREYA project. One important category of use cases, **Reuse across versions and parts**, has not been addressed in DataCite Commons yet, and will be worked on in 2021.

3.5 Statistics

DataCite Commons has a publicly accessible statistics page[20] that gives an overview of the information available via the service, focussing on data sources, the content categories works, people, and organizations, and the connections between these categories.

3.5.1 Data sources

The Data Sources section of the statistics page shows the PID providers who contribute content to DataCite Commons, the content category, and the number of PIDs contributed. The section also lists other data sources contributing information to DataCite Commons.

Figure 9 DataCite Commons Data Sources on the Statistics Page on 18 October 2020

At the time of writing this report, DataCite Commons included 38 million PIDs and associated metadata. The biggest gap is missing Crossref metadata, but the import process is ongoing and is much more complex than including ORCID or ROR PIDs, as it requires metadata conversion into the DataCite Metadata Schema and the import into the DataCite Search index. We expect DataCite Commons to reach 50 million PIDs before the end of 2020, and 100 million PIDs in 2022.

3.5.2 Content categories

The biggest content category is works, which includes the major research output categories publications, datasets, and software, and uses DOIs from DataCite and Crossref as PIDs.

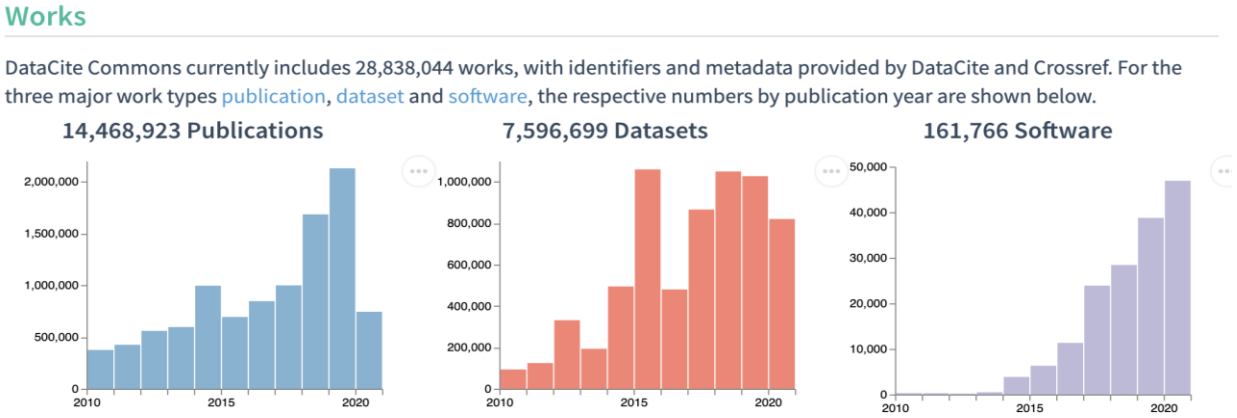


Figure 10 Works published by year on the DataCite Commons Statistics Page on 18 October 2020

[20] https://commons.datacite.org/statistics

Both by the total count and the number of new content published by year, publications are the largest category in DataCite Commons. This will become even more pronounced once all DOIs from Crossref (totaling 117 million) have been included in DataCite Commons. Because the initial import of Crossref DOIs was not random but favored more recently published content, conclusions about trends over time should be drawn cautiously. Publications can also be registered with DataCite – in particular, grey literature, preprints, reports, etc. – and currently make up 40.12% of all publications in DataCite Commons.

The number of datasets published by year is increasing every year, but there are significant fluctuations, mainly caused by big data repositories starting to register DOIs for all their content, as happened in 2015. While Crossref also allows the registration of DOIs for datasets, 99.69% of all datasets in DataCite Commons have been registered with DataCite. Datasets frequently have multiple versions, and the granularity (e.g. one big dataset vs. multiple subsets registered individually) may differ by community, as there are no widely adopted best practices regarding granularity across communities. These are important co-variables that need to be taken into account when comparing numbers of datasets over time.

The number of software releases published per year is much smaller, but shows a much faster growth rate compared to publications and datasets. Crossref does not register DOIs for software, and 85.18% of all DataCite DOIs for software have been registered by a single repository, Zenodo[21]. Frequent versioning is common for software, the current count of 161,766 DOIs represents about 40,000 software packages, with on average four versions per package.

The number of people who register for an ORCID ID is growing steadily and will reach the important milestone of 10 million ORCID IDs before the end of 2020.

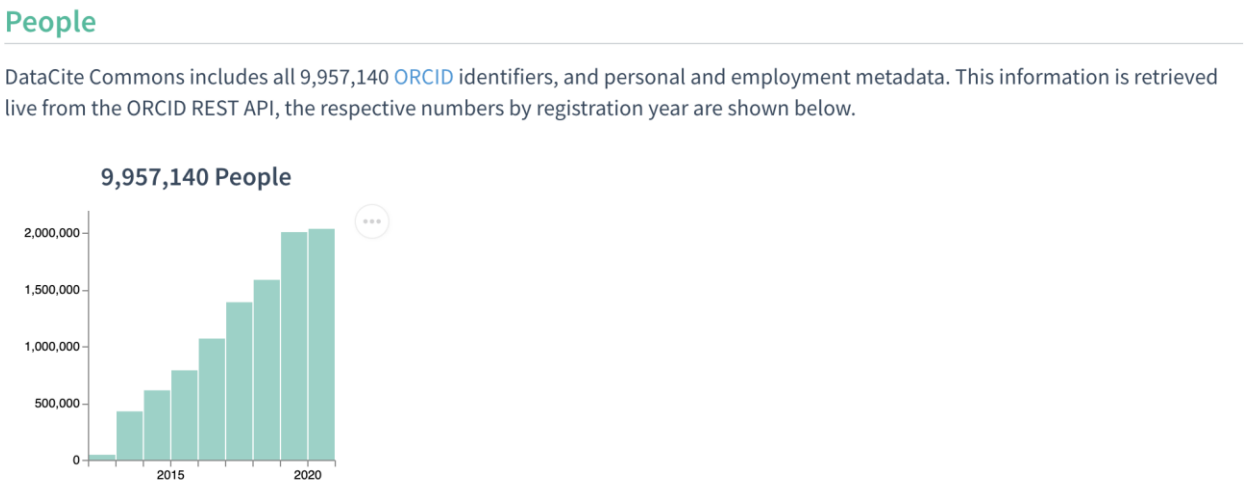


Figure 11 People registered by year on the DataCite Commons Statistics Page on 18 October 2020

The Research Organization Registry ROR currently re-uses the metadata from existing GRID[22] identifiers, with 80,248 identifiers on 9 November 2017 right before the launch of ROR, and 98,598 identifiers on 6 October 2020, with releases every 2-4 months.

[21] https://zenodo.org
[22] https://grid.ac/

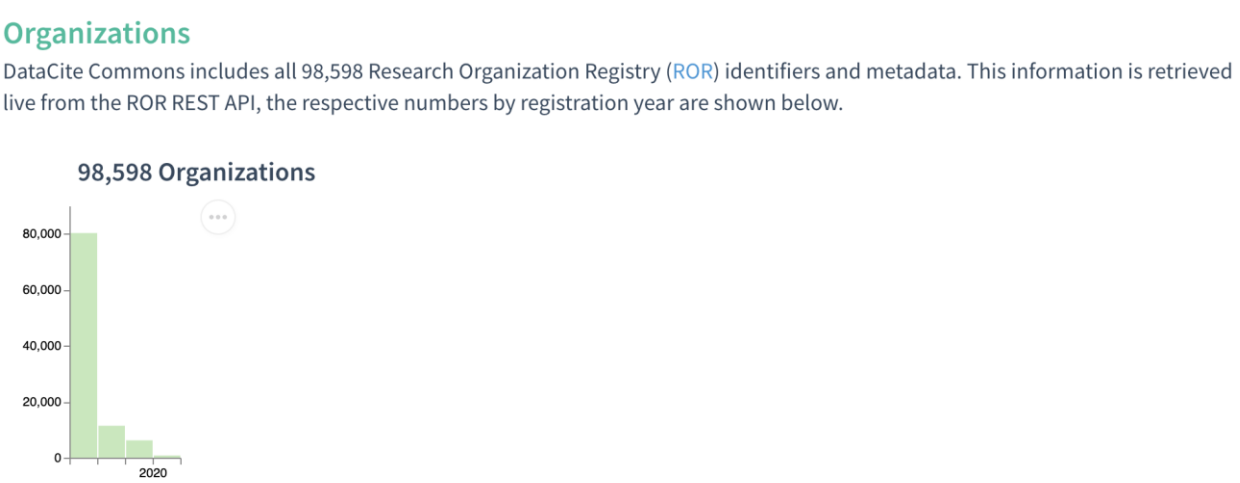


Figure 12 Research organizations registered by year on the DataCite Commons Statistics Page on 18 October 2020

The growth of the number of ROR identifiers in the past few years has been slowing down, and a similar pattern can be seen with GRID identifiers before the launch of ROR. This may indicate that identifying research institutions for the ROR affiliation identifier use case has been mostly addressed.

3.5.3 Connections

The DataCite Commons statistics page reports the following PID connections:

1. Number of works that have been cited, separate for publications, datasets, and software
2. Number of works that have been linked to a person
3. Number of works that have been linked to an organization (research organization or funder)

We see that 42.05% of the publications in DataCite Commons have been cited at least once, whereas only about 1% of datasets and software have ever been cited.

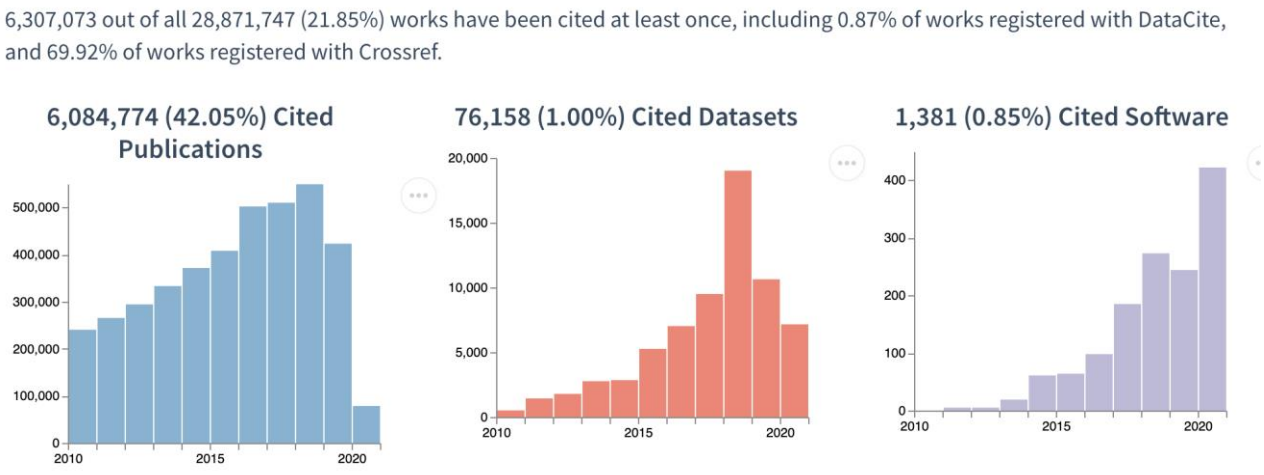


Figure 13 Citations by publication year of the cited work on the DataCite Commons Statistics Page on 18 October 2020

While we expect publications to be cited more frequently, these numbers also heavily depend on finding all citations for datasets and software and properly reporting them in DataCite Commons. Following the Scholix metadata schema[23] and additional work by DataCite[24], we only show relations that use the relation types *cites*, *references* and *is supplemented by* (and the corresponding inverse relationships), and we add to make changes to the Crossref/DataCite Event Data service to properly show those citations in the DataCite GraphQL API. This work is still ongoing, and we will properly show all citations collected by the Event Data service in the coming months.

On the statistics page we report that as of 18 October 2020, 3,654,901 out of all 28,871,747 (12.66%) works have been connected to at least one ORCID record, including 5.21% of works registered with DataCite, and 29.75% of works registered with Crossref. Broken down by work type

1. 19.33% of all publications are connected with at least one person via an ORCID ID,
2. 7.44% of all datasets are connected with at least one person via an ORCID ID, and
3. 14.11% of all software is connected with at least one person via an ORCID ID.

Consistent with other findings, we see higher adoption of ORCID for authors for publications than for creators of datasets or software. There is a wide variety of these percentages across publishers and repositories, and DataCite Commons and the underlying GraphQL API allow us to dig deeper into these numbers. The data repository PANGAEA for example has as of 18 October 2020 32.13% of its 389,536 datasets connected with at least one person via an ORCID ID.

On the statistics page we report that as of 18 October 2020, 14,277,095 out of all 28,871,747 (49.45%) works are connected with at least one organization via ROR ID or Crossref Funder ID, including 65.60% of works registered with DataCite, and 12.43% of works registered with Crossref. Broken down by work type, we see that

1. 30.95% of all publications are connected with at least one organization via a ROR ID or Crossref Funder ID.
2. 62.13% of all datasets are connected with at least one organization via a ROR ID or Crossref Funder ID, and
3. 97.13% of all software is connected with at least one organization via a ROR ID or Crossref Funder ID.

The difference between DataCite and Crossref, and indirectly between publications vs. datasets and software, is mainly explained by two factors:

1. DataCite supports ROR as an affiliation identifier in DOI registrations since August 2019, Crossref has not yet launched support for ROR IDs in DOI registrations.
2. DataCite supports ROR as an identifier for DataCite members, and content hosted by repositories managed by these members can be linked to this ROR ID.

The Crossref DOIs linked to ROR are at this time all linked via funding information and the Crossref Funder ID, which in turn is linked to a ROR ID.

[23] https://github.com/scholix/schema

[24] Garza, K. (2020). Datacite Citation Display: Unlocking Data Citations. https://doi.org/10.5438/1843-K679

4 Conclusions and outlook

4.1 Conclusions

In this report we describe the DataCite Commons service that was launched in October 2020 as part of the FREYA project. DataCite Commons successfully addresses the two main goals of the DataCite Commons service, (a) a common DOI search for all DOIs irrespective of content type and DOI registration agency (starting with FREYA partners Crossref and DataCite), and (b) a web interface for exploring the PID Graph of connected scholarly resources that was built in the FREYA project.

DataCite Commons is a very ambitious project and it is not surprising that there is still work left to do at the end of the FREYA project. But the service has a solid foundation in terms of addressing important use cases, the technical architecture, and the integration into the European Open Science Cloud. The sustainability and further development of the service are provided by DataCite, a not-for-profit membership organization that does not depend on ongoing grant funding to maintain the DataCite Commons service.

4.2 Adoption

DataCite Commons has only recently been released (October 2020). DataCite will continue the adoption outreach activities started in the FREYA project, and will scale the DataCite Commons infrastructure as the traffic to the service scales.

4.3 EOSC coordination

The DataCite GraphQL API and DataCite Commons provide important PID infrastructure for EOSC, in particular regarding new PID types such as organizations or instruments, and regarding open science graphs for discovery services and reporting. FREYA and DataCite have been coordinating with the OpenAIRE project on their respective open science graph projects via the Research Data Alliance (RDA) Open Science Graphs for FAIR Data Interest Group[25]. DataCite Commons will be registered in the EOSC Marketplace.

4.4 Sustainability of the service

DataCite Commons is maintained and will be further developed beyond the duration of the FREYA project by DataCite. Initially the focus will be on adding more PIDs and their metadata and more connections, bug fixes and performance improvements, and feedback collection, adding to the ideass we already received (see Annex).

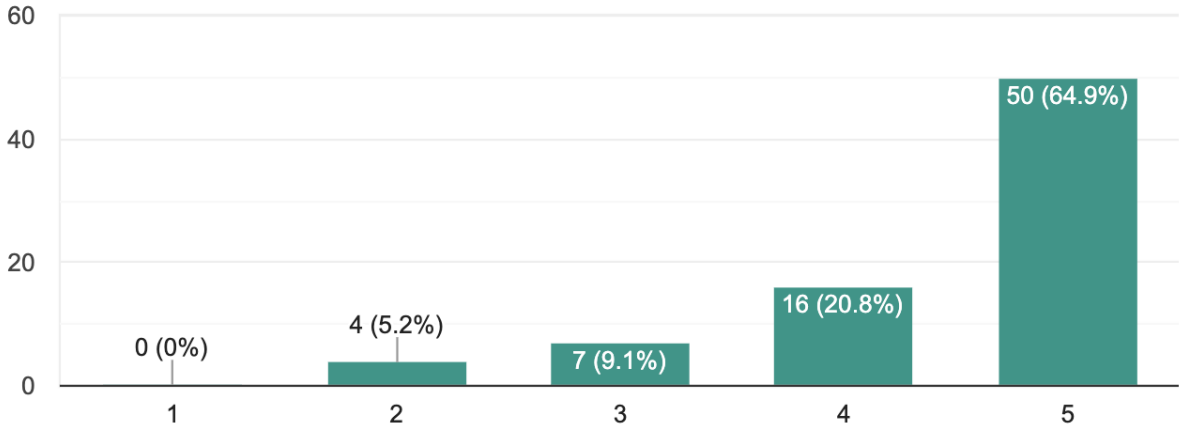
[25] https://www.rd-alliance.org/groups/open-science-graphs-fair-data-ig

Annex: FREYA Questionnaire May 2020

In May 2020 the FREYA project sent out a questionnaire to learn more about the desired functionality for DataCite Commons. We received feedback from 78 survey participants regarding five specific and one open question. The strongest support (1- not important, 5- very important) was given to the following three questions:

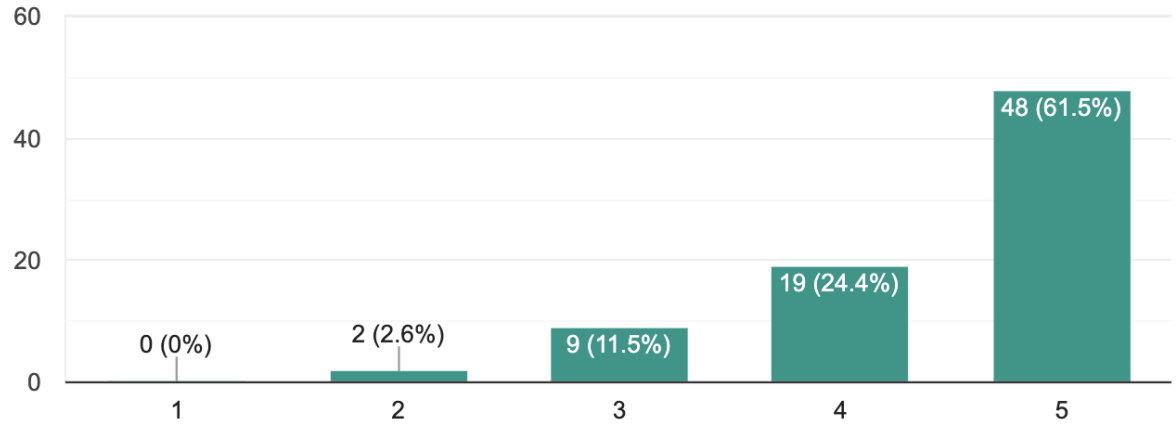
The ability to search in one place for a scholarly DOI and resolve that DOI

77 responses



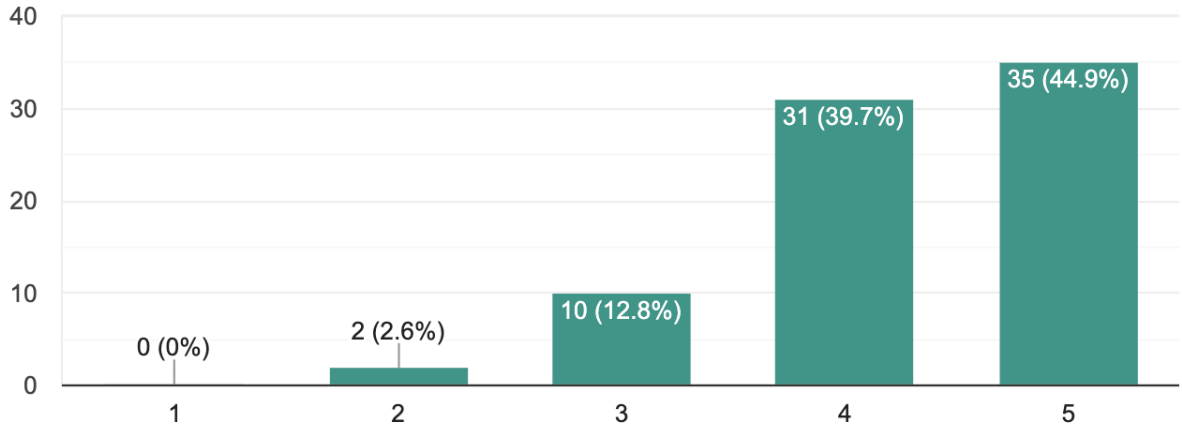
The ability to search across all metadata fields associated with a DOI, e.g. creator, description, date etc.

78 responses



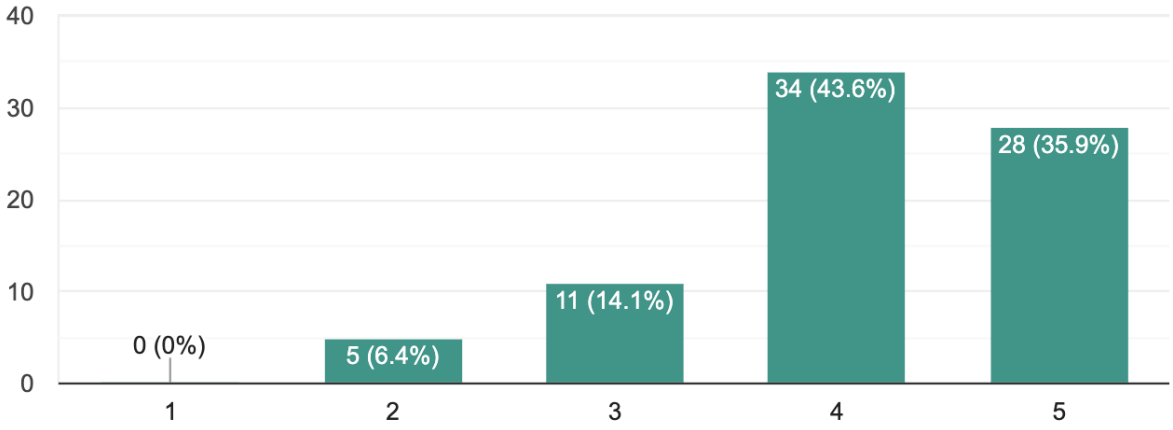
The ability to view all metadata associated with a DOI in the search interface

78 responses



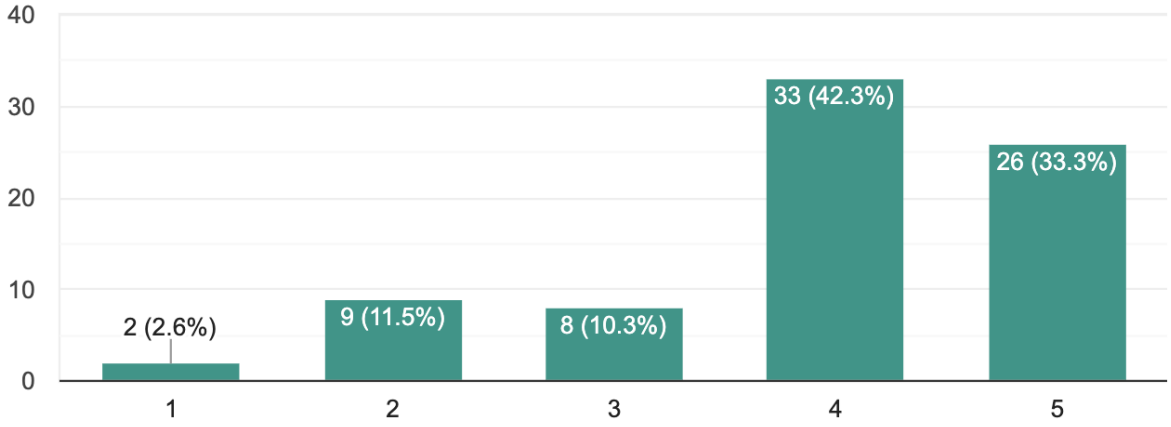
The ability to resolve all associated PIDs from the search results e.g. ORCID IDs, ROR IDs

78 responses



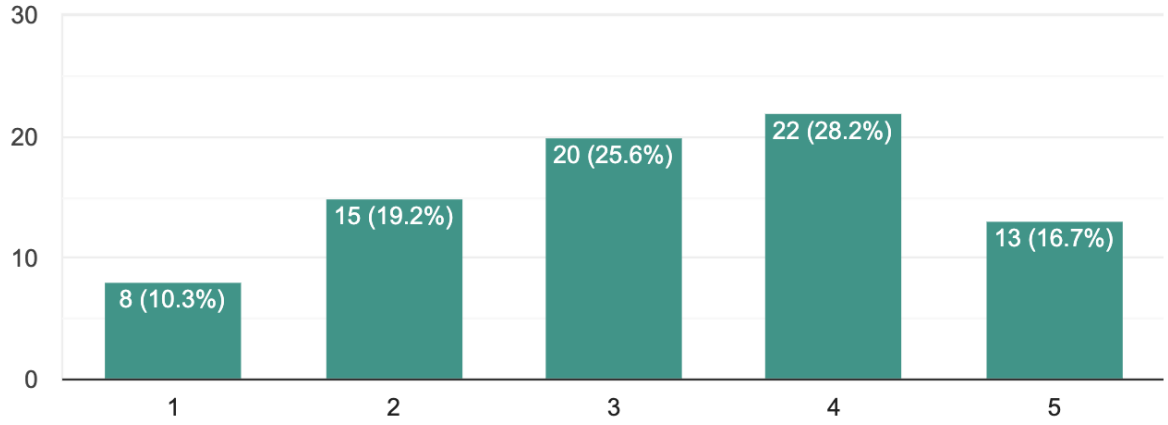
The ability to search for other PIDs e.g. ORCID IDs, ROR IDs within this interface

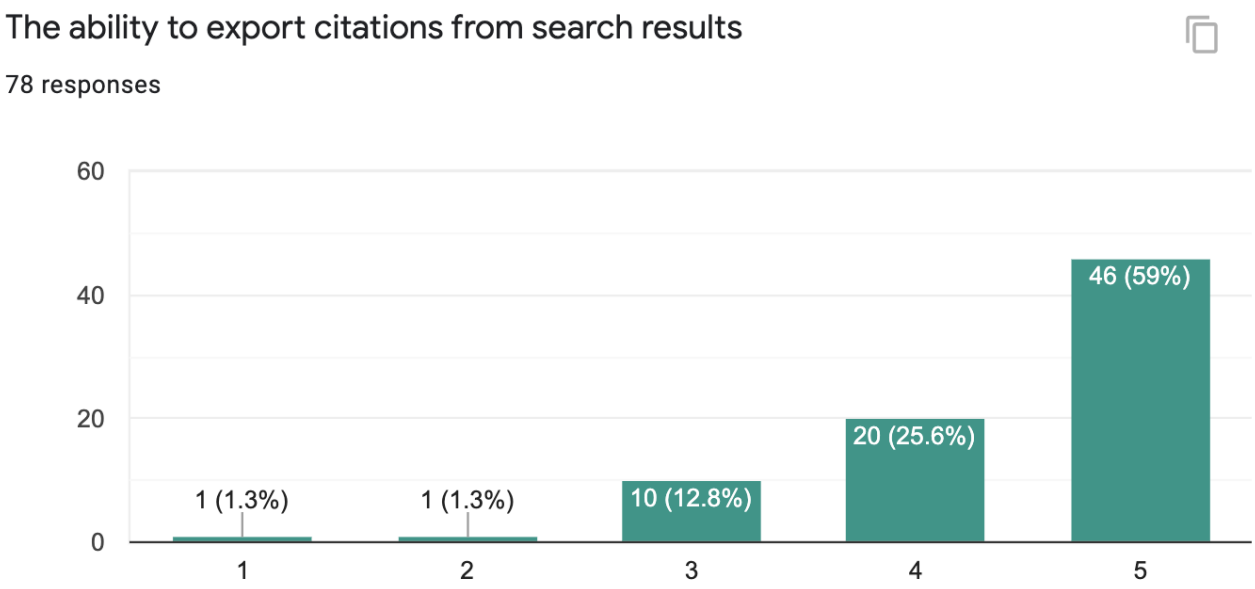
78 responses



The ability to view the relationship between PIDs graphically, such as in the image

78 responses





The following answers were given to the open question:

1. While I don't think a visual graph is important, a list of (linked) related identifiers is. I also have problems with non-CrossRef/Datacite providers such as those used in China
2. support journal citation verification
3. Highlight licenses, re-use rights, correctly identify publishers
4. visualization of search results metrics
5. good analytics
6. indication of where the PID was found (e.g. references, text, appendix, figure, etc)
7. Ability for repositories or data discovery software to connect to via Common DOI Search's API.
8. Offer way for me to assert facts against DOIs so that others can see what I know.
9. Should direct you to the data repository to get the "raw" data
10. I would assume most people will not know a specific DOI so being able to search by other fields that links to resources (data, articles, etc) through the DOI would be important.
11. An API. For example when asked if I want to "view" relationships between PIDs - that's a nice to have feature, but do I want to query them? That's a must have feature . The API & its endpoints is as important as the resolver and the human facing UI
12. Discover citations by DOI prefix and within ranges of publication dates of the citing articles..
13. A very usable search interface that returns relevant results reliably. Its UI needs to be a huge improvement over what's currently available at search.datacite.org or scholexplorer.openaire.eu.
14. It should be as simple as possible; it should supplement, and not replace existing platforms.
15. Citations/backlinks
16. Nothing
17. Difficult to answer as I am not someone who would explicitly use it.
18. Automated DOI search - notification for search results from DOIs maintained by a given archive.
19. nothing else
20. I think the clue is not SEARCH but rather counting. So database indexing. Because this will never be a true Search tool - it will rather be a database tool that other tools can re-use
21. search via map
22. It would be nice if there was a way to import the results into my own DOI manager to update my DOIs.
23. I like the idea of all PIDs being included, but while ORCID insists on validation of ORCIDs before including in your content this creates issues - either everyone does this or nobody does it to allow fairness and trust. Depending on who does what first, some services can pull PIDs for their workflows from this service
24. NA

25. The ability to nominate new IDs for inclusion in the search (e.g. arXiv ID)
26. An open API
27. Ability to search across metadata that might be different from each provider.
28. Easy-to-find info about what is in and out of scope. Will all orgs that mint DOIs all contribute all the same kinds of metadata?
29. If so, showing which other versions of the DOI referenced content there are. E.g. Zenodo provides different versions. I would like to know if my DOI refers to the latest version of the content, in which (hopefully) more errors are corrected than in older versions.
30. The ability to click through from your Orcid account to see "your" graph and new 1st and 2nd associations. Researchers constantly ask who's using their work and for what purpose.
31. Open Access check will be wonderful
32. Search for all types of PIDs (for institutions, people, research resources, funders, projects etc). Ability to create sets of PIDs linked to one project or grant. Ability to group versioned PIDs. Offer concordances with oher (proprietary) IDs like Scopus, MAG, LENS ARXIV, Redalyc etc
33. Alerting on changes to the graph associated with a node
34. Resolution to open access sources, and an API to allow use by library search engines and link resolvers.
35. APIs
36. An interface like this might also help with understanding the trends of how DOIs are curated, for instance, are certain related id types used over others.
37. Ability to sort, ability to export in a range of formats
38. You've really named the biggest things for me.
39. Should be capable of extension to cover all kinds of Handles and not just DOI.
40. import to Reference management software, possibly integration of open access status and/or integration von unpaywall to jump to free version of the DOI document (if a publication)