



Organization IDs in Germany—Results of an Assessment of the Status Quo in 2020

RESEARCH PAPER

PAUL VIERKANT

ANTONIA SCHRADER

HEINZ PAMPEL

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Persistent identifiers (PIDs) for scientific organizations such as research institutions and research funding agencies are a further decisive piece of the puzzle to promote standardization in the scholarly publication process—especially in light of the already established digital object identifiers (DOIs) for research outputs and ORCID iDs for researchers. The application of these PIDs enables automated data flows and guarantees the persistent linking of information objects. Moreover, PIDs are fundamental components for the implementation of open science. For example, the application of PIDs for scientific organizations is of crucial importance when analyzing publications and the costs of the transition to open access at an institution.

To find out more about the status quo of the use and adoption of organization IDs in Germany, a ‘Survey on the Need for and Use of Organization IDs at Higher Education Institutions and Non-University Research Institutions in Germany’ was conducted among 548 scientific institutions in Germany in the period from July 13 to December 4, 2020, as part of the DFG-funded project ORCID DE. One hundred and eighty-three institutions participated in what was the largest survey to date on organization IDs in Germany. The survey included questions on the knowledge, adoption, and use of organization IDs at scientific institutions. Moreover, respondent institutions were asked about their needs with regard to organization IDs and their metadata (e.g., in terms of relationships and granularity). The present paper provides a comprehensive overview of the results of the survey conducted as part of the aforementioned project and contributes to the promotion and increased awareness of organization IDs.

CORRESPONDING AUTHOR:

Paul Vierkant

DataCite, DE

pvierkant@datacite.org

KEYWORDS:

organization ID;
standardization; affiliation;
publication management;
indexing; persistent identifier;
publishing; open science

TO CITE THIS ARTICLE:

Vierkant, P, Schrader, A and Pampel, H. 2022. Organization IDs in Germany—Results of an Assessment of the Status Quo in 2020. *Data Science Journal*, 21: 19, pp. 1–23. DOI: <https://doi.org/10.5334/dsj-2022-019>

1. ORCID DE PROJECT

The ‘Umfrage zum Bedarf und Nutzung von Organisations-IDs an Hochschulen und außeruniversitären Forschungseinrichtungen in Deutschland’ (Survey on the Need for and Use of Organization IDs at Higher Education Institutions and Non-University Research Institutions in Germany) was conducted within the framework of the second funding phase of ORCID DE (Bertelmann et al. 2021), a project funded by the German Research Foundation (DFG). The aim of the project is to consolidate the ORCID information infrastructure in Germany (Bertelmann et al. 2019). To this end, the expansion and consolidation of the existing network of scientific institutions that have integrated ORCID into their infrastructures is being pursued. The focus is on expanding support for the institutions that are interested in or are using ORCID and—prospectively—on an identification system for organizations that is linked to ORCID (ORCID DE 2021). The project partners, on behalf of the scientific institutions in Germany, want to participate in this development from the very beginning. Moreover, new contexts for the application of ORCID will be tested and evaluated.

The project partners of the project initiated by the German Initiative for Network Information (DINI)¹ are DataCite,² the Helmholtz Open Science Office,³ the German National Library,⁴ Bielefeld University Library,⁵ and TIB—Leibniz Information Centre for Science and Technology and University Library.⁶

The survey addressed in this contribution served to identify the state of knowledge about and the use and adoption of organization IDs at scientific institutions in Germany and to assess knowledge about existing technical, library, and organizational needs when using organization IDs.

2. METHOD, SAMPLE, AND DESIGN

2.1. METHOD

Since the aim of the work package on organizational identifiers of the ORCID DE project was to obtain an overview of the current state of organizational identifiers at scientific institutions in Germany, the method chosen had to be primarily quantitative. Given the weaknesses of a quantitative approach (e.g., lack of detail), it was proposed to complement this study with additional qualitative surveys. This was to be done by a more in-depth survey of interested institutions attending the ORCID DE workshop series. The present quantitative survey was conducted as an online survey. It was technically implemented on a Bielefeld University server using the software LimeSurvey.⁷

2.2. QUESTIONNAIRE CONCEPT

The present survey was the largest survey to date on organization IDs in Germany. Structurally, it follows on from the survey ‘ORCID in Deutschland—Ergebnisse einer Bestandsaufnahme im Jahr 2016 [ORCID in Germany—Results of an Assessment of the Status Quo in 2016]’ (Fuchs et al. 2017). Knowledge gained from previous studies on organization IDs, such as an assessment of the status quo entitled ‘Persistent Identifier für wissenschaftliche Einrichtungen [Persistent Identifiers for Scientific Institutions]’ by Rothfritz (2018), the international initiative ‘Organization Identifier Project: A Way Forward’ by Cruse, Haak, and Pentz (2016), and the analysis ‘Organisation Identifiers: Current Provider Survey’ by Bilder, Brown, and Demeranville (2016), was incorporated into the study design.

1 <https://www.dini.de>.

2 <https://datacite.org>.

3 <https://os.helmholtz.de/en/>.

4 https://www.dnb.de/EN/Home/home_node.html.

5 <https://www.ub.uni-bielefeld.de/ub/>.

6 <https://www.tib.eu/en/>.

7 <https://www.limesurvey.org>.

The online questionnaire was organized into the following four sections:

- General
- Experience with Organization IDs
- Use
- Needs with Regard to Organization IDs

Only the first 2 of the total of 22 questions in the questionnaire were mandatory.

2.2.1. General

In the first section of the questionnaire, the participants were asked to assign their institutions to an institution type. The respondents were then asked to indicate the function within the institution in which they were participating in the survey.

2.2.2. Experience with organization IDs

In the second section, participants were asked about their awareness, knowledge, and perceptions of the characteristics or application scenarios of organization IDs (multiple selection possible). First, they were asked in what context they had already heard about organization IDs (multiple selection possible).

Besides the question about hearing about organization IDs in various contexts, the participants were also asked about perceived characteristics of or application scenarios for organization IDs (multiple selection possible). In addition, they were given an opportunity in a free-text field to name further application scenarios or characteristics known to them. To capture awareness of specific organization IDs, participants were asked what organization IDs they already knew (multiple selection possible). The 17 organization ID systems offered as response options were based on the mentions of such systems in the relevant literature (Rothfritz 2018). They included commercial (e.g., Ringgold identifiers) and noncommercial (e.g., Wikidata) as well as national (e.g., the DFG's GERit ID) and international (e.g., the Research Organization Registry's ROR ID) organization ID systems. Participants were also given an opportunity in a free-text field to name further organization IDs that they knew.

2.2.3. Use

Over and above mere awareness and perception, information about the use of organization IDs is necessary to determine the status quo of this PID type at scientific institutions in Germany. Therefore, in the third section, respondents were asked to indicate the organization IDs that were already in use at their institutions (multiple selection possible). In a matrix, respondents could choose from among the following response options for the organization IDs mentioned in the previous question:

- We are already using it.
- We are not using it, but we are planning to use it.
- We are not using it, and we are not planning to use it.
- I'm not sure.

Over and above the current use of organization IDs, the response option 'We are not using it, but we are planning to use it' sheds light on the future development of organization IDs (multiple selection possible). As a cross-check, the response option 'We are not using it, and we are not planning to use it' provides information as to which organization IDs the respondents' institutions will not be using in the future.

After the question about the general use of organization IDs, the respondents were asked to indicate the areas in the institution in which organization IDs were already in use. Here, too, they could choose in a matrix from among the previously used response options:

- We are already using it.
- We are not using it, but we are planning to use it.
- We are not using it, and we are not planning to use it.
- I'm not sure.

To concretize the question as to the areas in which organization IDs were being used, respondents were then asked to indicate the software in which the organization ID in question was being or was to be implemented.

The use of organization IDs in institutional policies is a further aspect of their adoption at scientific institutions in Germany. Respondents were asked whether organization IDs were integrated in their institutional guidelines (policies) and, if yes, to what extent. To take account of the different characteristics of the various types of respondent institutions, respondents could choose between the following guidelines or regulations as response options:

- Affiliation policies
- Research data or research data management guidelines
- Doctoral degree regulations or guidelines
- Publication regulations or guidelines
- Other

To also take account of future developments with regard to the policy aspect, the response option 'The integration of the organization IDs is being considered' was offered.

2.2.4. Needs with regard to organization IDs

In the fourth and final section of the survey, the focus was on needs with regard to organization IDs from various perspectives. Respondents were first asked about their needs with regard to metadata and their curation, as the design of the metadata schema, its coverage, and the curation processes have indirect effects on the acceptance, use, and thus adoption of an organization ID.

To capture the needs with regard to the differing application scenarios among the institution types, the participants were asked to rate the need for characteristics of an organization ID for their application scenarios (multiple selection possible). Via a matrix, the respondents could rate their respective levels of need by choosing one of the following response options:

- Very high need
- High need
- Low need
- No need

Because some benefit is created by only covering different levels of an institution, granularity plays an important role in the context of organization IDs. Therefore, the participants were asked how granularly the organization structure should be represented in the metadata of an organization ID.

Relationships—that is, links between (legal and natural) persons as well as other objects (e.g., publications, projects, or events)—are an important aspect of PIDs in general and of organization IDs in particular. Such relationships are reliably possible only through the uniqueness of PIDs. They create added value, for example, in terms of the discoverability, evaluation, and monitoring of research outputs. Therefore, three questions about relationships were asked in the present survey. First, participants were asked what relationships should generally be provided in the metadata of an organization ID (multiple selection possible). Taking account of the relevance of the relationships between organizations, the second question aimed at determining what relationships between organizations should be definable (multiple selection possible). As the smallest 'unit' of organizations, researchers were the focus of the third relationships question. Analogous to the previous two questions, participants were asked what relationships between persons and organizations should be definable (multiple selection possible).

Relationships between PIDs represent a great added value of persistent identifiers. Of equal importance for the use of PIDs is that they be tagged with sufficient metadata. The structuring or depth of the organization metadata captured in the form of a metadata schema is decisive for the level of use and adoption of an organization ID as well as for the amount of curation needed. Respondents were therefore asked what metadata should be assignable to an organization ID (multiple selection possible).

To facilitate or enable the uniform external presentation of the institution and the entire research output of persons affiliated with the institution, names are increasingly specified

and laid down in affiliation policies.⁸ Respondents were therefore asked what name form of the organization should be the preferred and primarily displayed name form that is presented together with the organization ID.

To remain up to date, metadata that have been created must be maintained (i.e., curated). The participants were therefore asked what actor should have decision-making authority over the curation (e.g., updating) of the metadata of an organization ID. Specifically addressing the task of metadata curation, they were then asked what actor within their own institution would assume responsibility for the maintenance of its metadata record (multiple selection possible).

Over and above the needs with regard to metadata and their curation, topics such as the governance, operation, and business model of a service for organization IDs are of importance for its acceptance and adoption. The aim of the first question on the topic of governance was to determine what aspects of the governance of the service were important to the organizations participating in the survey. The participating institutions were further asked what aspects of the financially sustainable operation of a service for organization IDs were important to their organizations (multiple selection possible). The third and final question on the topic of governance measured whether a fee-based membership of a service for organization IDs would be an option for the respondent organizations.

At the end of the section and, at the same time, the survey, the focus moved away from the current needs with regard to organization IDs and toward the future need for PIDs for other resources. In the final question, the participants were asked for what resources or products of science further PIDs should be developed (multiple selection possible). Via a matrix, the respondents could rate their respective needs by choosing one of the following response options:

- Very high need
- High need
- Low need
- No need

3. SAMPLE, RESPONSE, AND ANALYSIS

The invitation to participate in the ‘Survey on the Need for and Use of Organization IDs at Higher Education Institutions and Non-University Research Institutions in Germany’ was sent to a total of 548 scientific institutions in Germany. The survey took place in the period from July 13 to December 4, 2020. In the introduction to the survey, the topic of organization IDs was introduced, and, among other things, the concept of (institutional) affiliation was explained.

Besides all the higher education institutions listed by the German Rectors’ Conference (HRK),⁹ the institutes of the Fraunhofer-Gesellschaft,¹⁰ the Helmholtz Association,¹¹ the Leibniz Association,¹² and the Max Planck Society¹³ were especially taken into account when selecting the institutions to be invited. Of the 548 research institutions contacted, 183 participated in the survey. Of these, 163 fully completed and 20 partially completed the questionnaire. This corresponds to a rate of fully complete participation of 29.7%.

After the initial analysis of the responses, it was decided to include the partially completed questionnaires in the analysis and to take the answered questions into account (pairwise case exclusion). The minimum requirement was defined as ‘at least one question answered.’ As a result, 183 datasets were included in the analysis described in what follows. This corresponds to a share of 33.4% of the institutions invited to participate in the survey.

8 The affiliation policy of the Technische Universität Braunschweig is one example of an affiliation policy that stipulates the use of organization IDs. Last checked on 19 May 2021, <https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Einrichtungen/UB/PDF/publikationsrichtlinie.pdf>.

9 <https://www.hochschulkompass.de/en/hochschulen/search-for-a-higher-education-institution.html>.

10 <https://www.fraunhofer.de/en.html>.

11 <https://www.helmholtz.de/en/>.

12 <https://www.leibniz-gemeinschaft.de/en/>.

13 <https://www.mpg.de/en>.

4. RESULTS

The results of the survey are presented in what follows according to the four sections of the questionnaire.

4.1. GENERAL

With regard to the participation of the different institution types, higher education institutions (universities and universities of applied sciences) accounted for more than half of the participants. The nonuniversity research institutions (Helmholtz centers, Leibniz institutes, Max Planck institutes, and the federal government's departmental research institutions) together accounted for a share of almost one third. The question was mandatory and was answered by 183 institutions (see [Figure 1](#)).

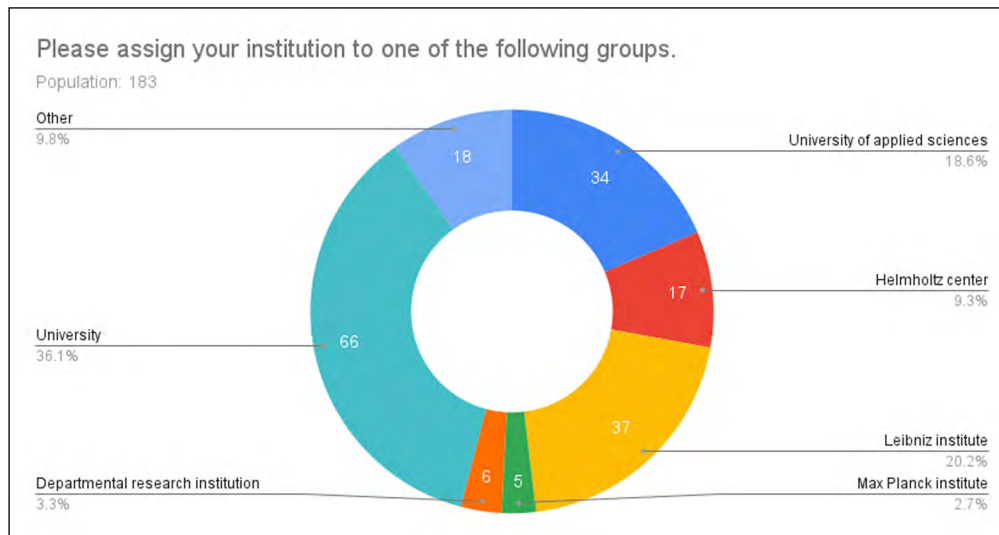


Figure 1 Distribution and number of responses to the question, 'Please assign your institution to one of the following groups.'

Recipients of the survey invitation were mainly either persons in senior positions at the libraries of the respective institutions or contact persons for electronic publishing or open access/open science. In response to the question, 'In what function within your institution are you participating in this survey?', a large majority (81.4%; $n = 149$) answered 'as a library staff member.' By contrast, research and teaching staff (3.8%; $n = 7$), administrative staff (3.3%; $n = 6$), data center staff (1.6%; $n = 3$), and university press staff (1.1%; $n = 2$) participated only to a small extent in the survey. Further functions were reported under 'Other' (8.7%; $n = 16$), for example, 'data manager,' 'RIS [research information system] officer,' and 'information management.' The question was mandatory and was answered by 183 institutions (see [Figure 2](#)).

4.2. EXPERIENCE WITH ORGANIZATION IDS

Answers to the question, 'In what context have you already heard about organization IDs?', show that, for the most part, respondents found out about organization IDs both in conversation with colleagues (59.3%; $n = 108$) and at events (56%; $n = 102$). Respondents also mentioned specialist literature (42.3%; $n = 77$) and mailing lists (41.8%; $n = 76$) as further important sources for finding out about organization IDs. Only 9.3% ($n = 17$) of the respondents found out about organization IDs for the first time in the questionnaire. Multiple selection was possible. The question was answered by 183 institutions (see [Figure 3](#)).

In response to the question, 'What characteristics or application scenarios do you perceive in the case of organization IDs?', the most important application scenarios perceived by respondent institutions from among the large number of possible application areas were the linking of persons to organizations (e.g., affiliation details), at 81.4% ($n = 149$), and the unique designation of organizations (disambiguation), at 80.3% ($n = 147$). Moreover, at 75.4% ($n = 138$), the linking of research outputs (publications, research data, software) to organizations was considered the third most important characteristic or application scenario. Other possible applications, such as the reporting/measurement of research outputs (bibliometrics; $n = 114$),

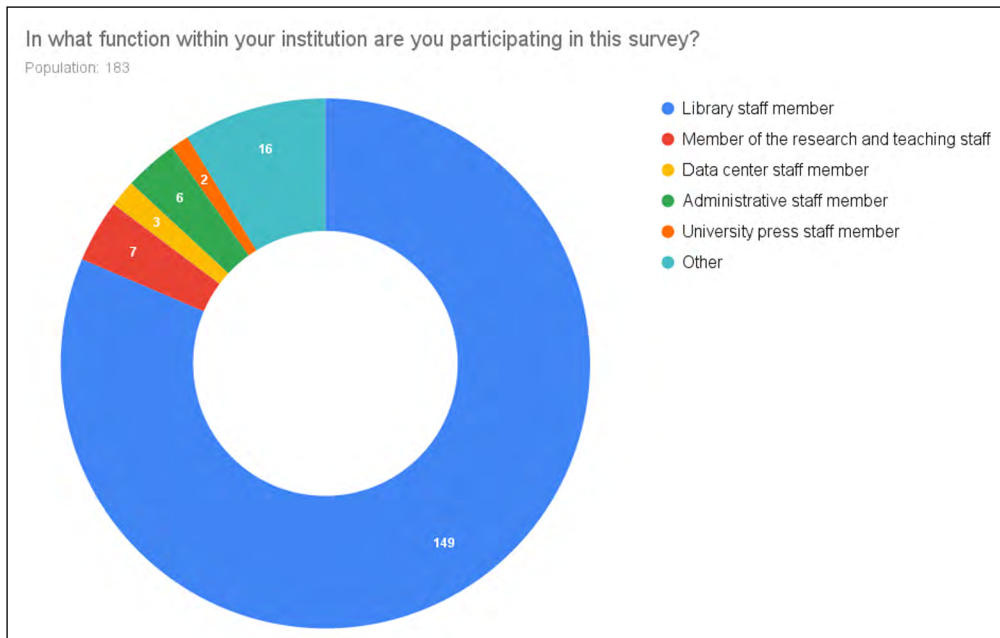


Figure 2 Distribution and number of responses to the question, 'In what function within your institution are you participating in this survey?'

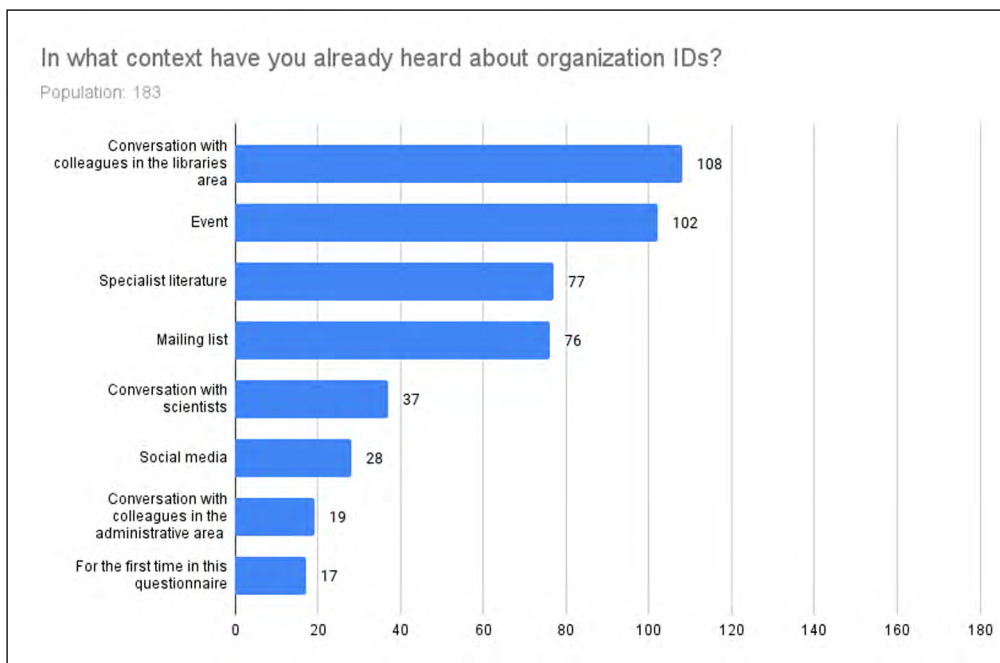


Figure 3 Distribution and number of answers to the question, 'In what context have you already heard about organization IDs?'

the possibility of implementing organization IDs in existing systems (e.g., research information systems; $n = 110$), and the simplification of metadata management ($n = 104$), were also seen as potential application scenarios. Only one other possible application—'project funding'—was mentioned in the free-text field. As no other application areas were mentioned over and above the application scenarios offered in the question, it can be assumed that the characteristics listed in the question and the above-mentioned characteristics represent the current state of potential application scenarios for organization IDs at scientific institutions in Germany. Multiple selection was possible. The question was answered by 183 institutions (see [Figure 4](#)).

With regard to the question as to what organization IDs were known to the participants, the Integrated Authority File (Gemeinsame Normdatei, GND), at 77% ($n = 141$), proved to be the best-known organization ID. The level of awareness of the remaining 16 organization IDs was very heterogeneous. At 57.4% ($n = 105$), the Ringgold identifier was the second most well-known organization ID system among the participants, while the ULAN ID (Union List of Artist Names), at 0% ($n = 0$), was the least well known. The 'Web of Science ID' and the 'EU Transparency Register, EU Participants Portal' were mentioned as further organization IDs in the free-text field. As no other organization ID systems were mentioned over and above the response options offered in the question, it can be assumed that the identifiers listed in the

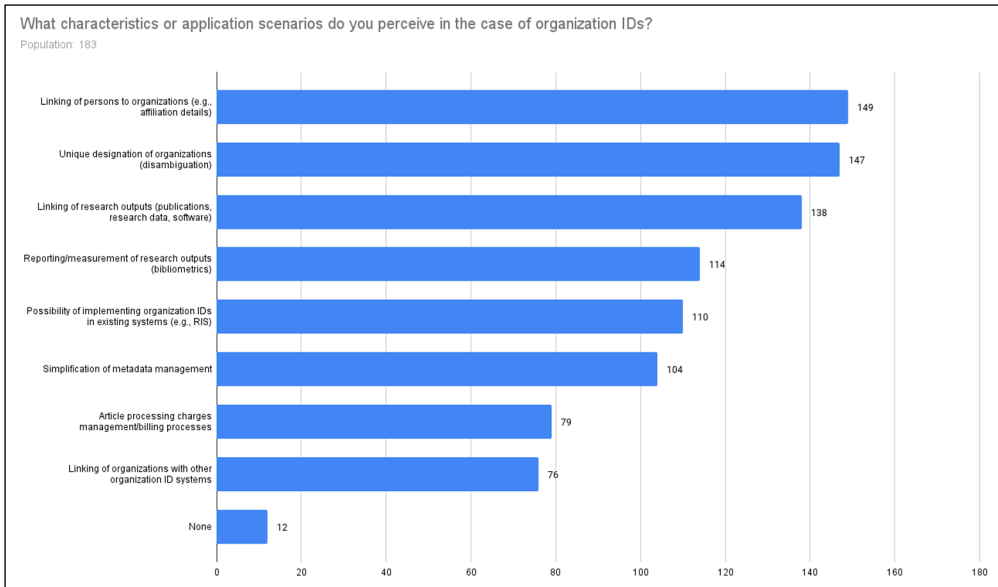


Figure 4 Distribution and number of responses to the question, 'What characteristics or application scenarios do you perceive in the case of organization IDs?'

question represent the current state of knowledge of organization IDs at scientific institutions in Germany. Multiple selection was possible. The question was answered by 183 institutions (see [Figure 5](#)).

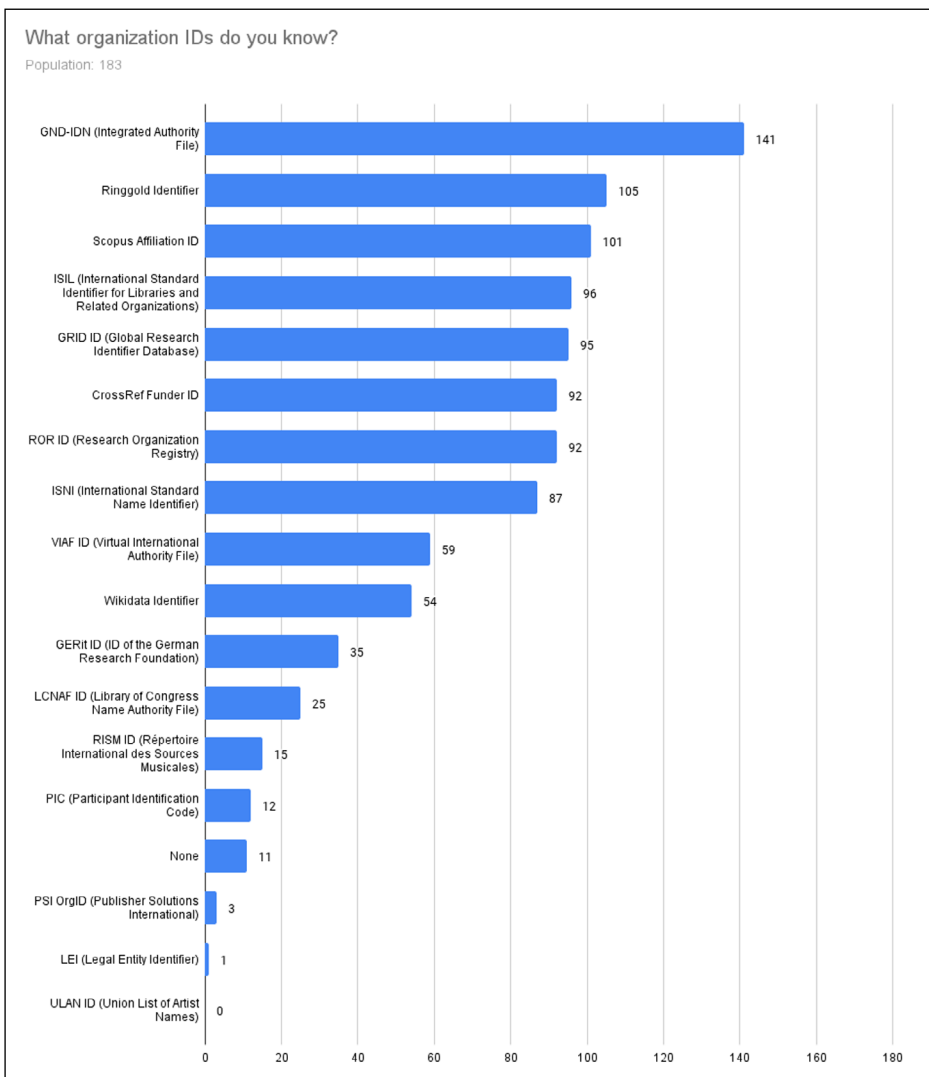


Figure 5 Distribution and number of responses to the question, 'What organization IDs do you know?'

4.3. USE

The Integrated Authority File (GND), as a national PID for organizations, was (at 77%) the best-known organization ID among the respondent institutions in Germany (see [Figure 5](#); n = 141).

With regard to the question, ‘What organization IDs are you already using?’, it turned out that the GND was also the most widely used organization ID: 85 participants (N = 134) reported that they were already using it (see Figure 6). Fifty-nine institutions (N = 118) reported that they were using the International Standard Identifier for Libraries and Related Organizations (ISIL), which was thus the second most used organization ID. Similarly well-known organization IDs, such as the Crossref Funder ID (n = 16, N = 105), the GRID ID (n = 35, N = 116), the ISNI (n = 26, N = 107), the Ringgold ID (n = 40, N = 113), the ROR ID (n = 22, N = 110), and the Scopus Affiliation ID (n = 35, N = 111), were used less often. Multiple selection was possible; 149 institutions answered all response options offered in the question. However, because of the question structure (not a mandatory question, matrix structure, and lack of a ‘None’ option), the populations (N) of the individual response options varied (see Figure 6). These differences were taken into account in the analysis.

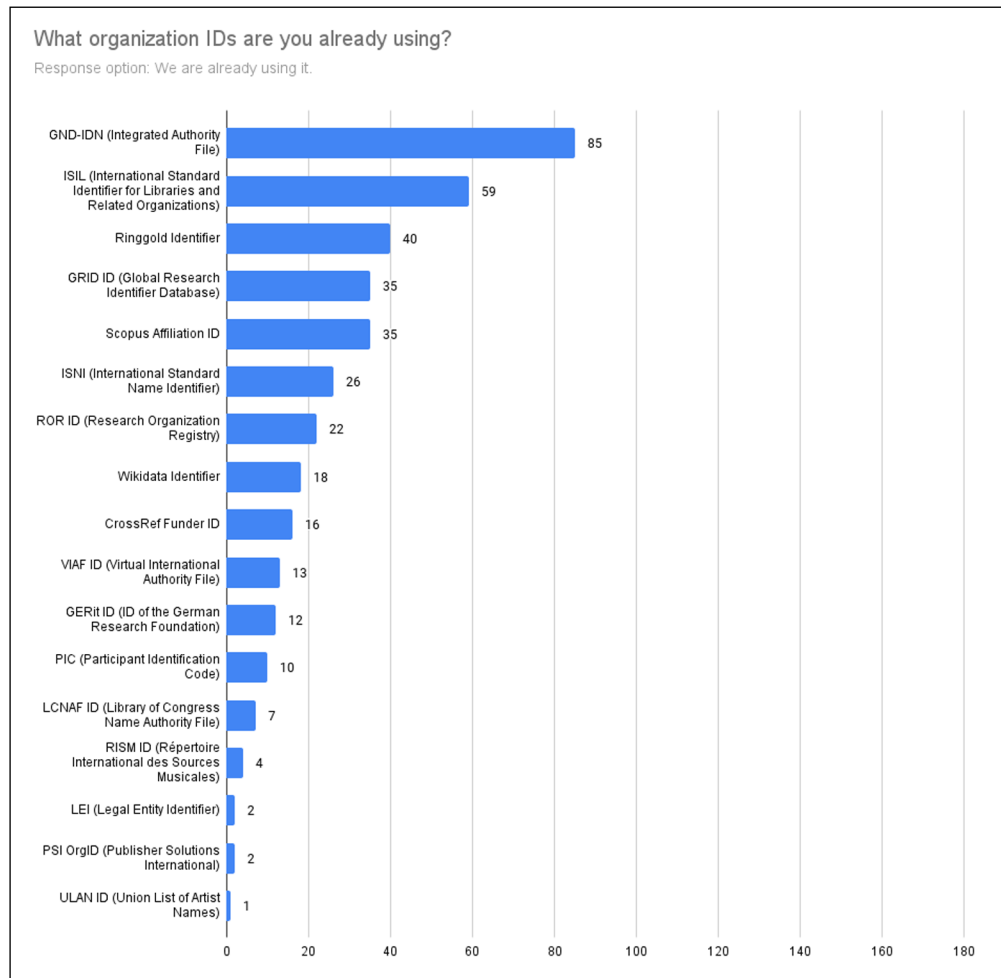


Figure 6 Distribution and number of the responses to the question, ‘What organization IDs are you already using?’ (Response option: We are already using it.).

In the case of the response option, ‘We are not using it, but we are planning to use it,’ it appears that scientific institutions in Germany will in future be relying on the ROR ID as their organization ID of choice (see Figure 7). Thirty-eight of the respondent institutions (n = 38, N = 110) reported that they were planning to use the ROR ID. This is almost as many as all the other four organization IDs in the top five put together: GRID ID (n = 12, N = 116), Crossref Funder ID (n = 9, N = 105), GERit ID (n = 9, N = 102), and Wikidata Identifier (n = 9, N = 104).

In the case of the response option, ‘We are not using it, and we are not planning to use it,’ the field was relatively homogeneous. In the case of most response options, between 40 and 60 institutions reported that they did not want to use the respective organization IDs. The IGND (n = 16) and the ROR ID (n = 23) stand out as the two organization IDs that the fewest respondent institutions were not planning to use (see Figure 8).

The response option ‘I’m not sure’ was also homogeneously distributed across all the organization IDs offered. Between 25 and 36 institutions expressed their uncertainty regarding the various response options for this question. However, it is unclear whether this uncertainty related to ignorance of the use of the respective organization IDs at their own institutions or to

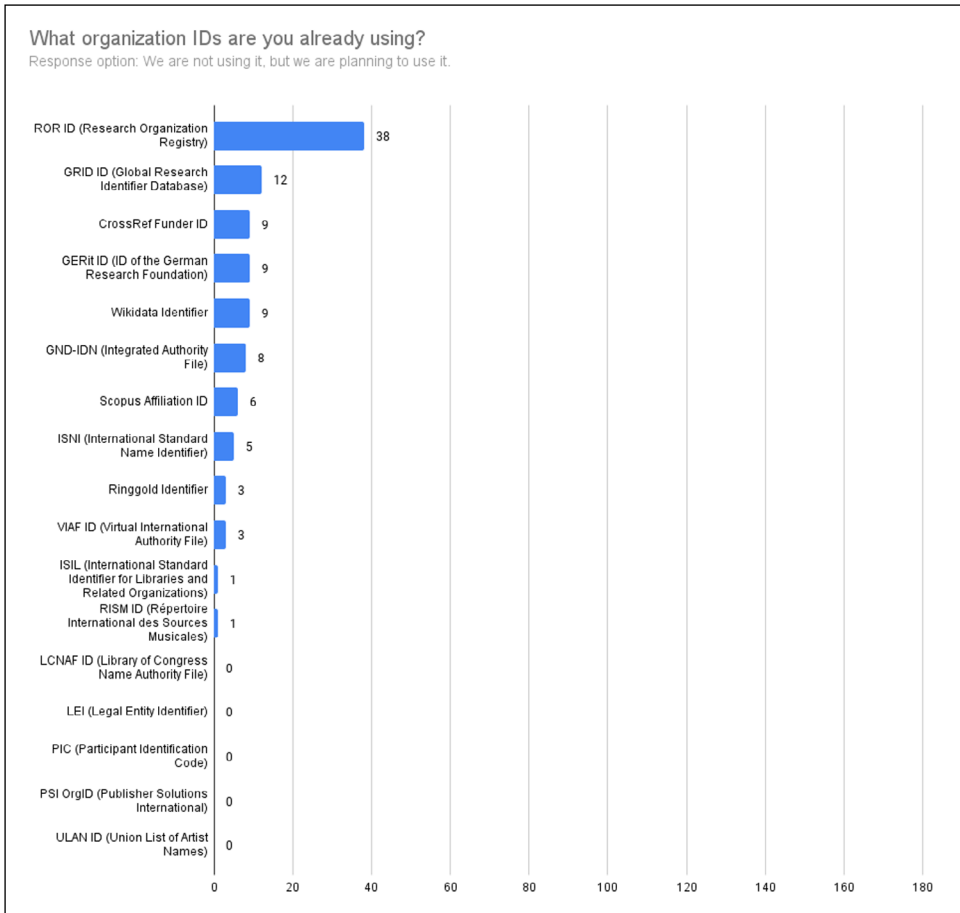


Figure 7 Distribution and number of responses to the question, ‘What organization IDs are you already using?’ (Response option: We are not using it, but we are planning to use it.).

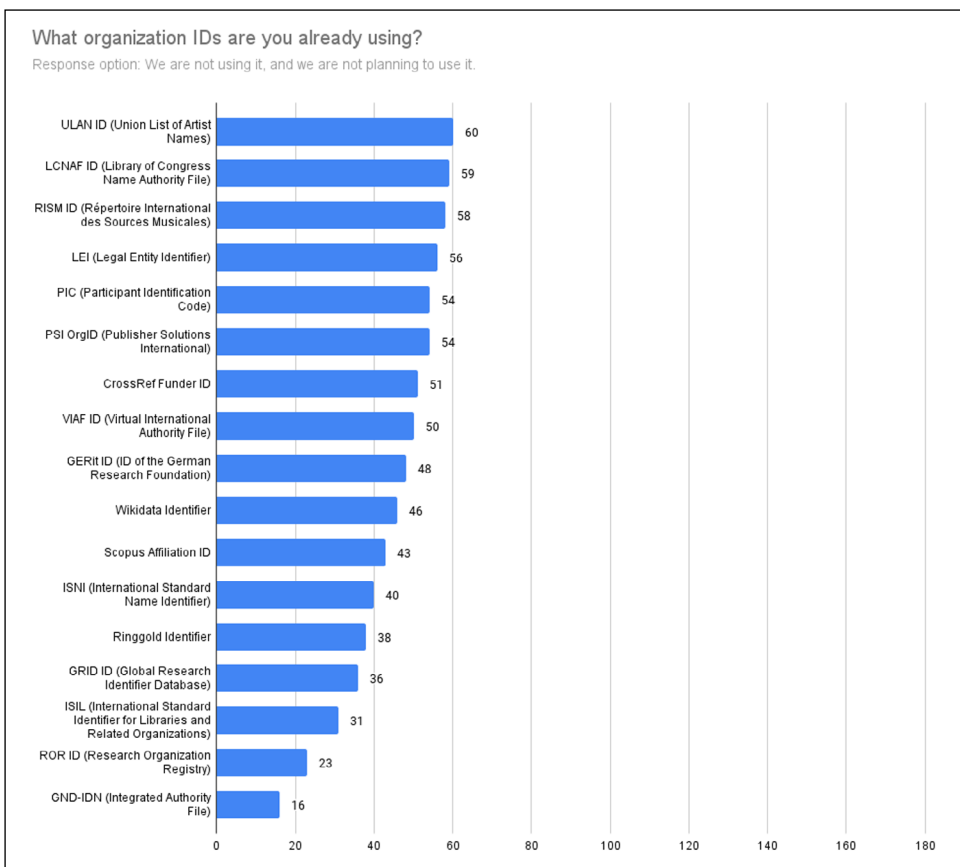


Figure 8 Distribution and number of responses to the question, ‘What organization IDs are you already using?’ (Response option: We are not using it, and we are not planning to use it.).

ignorance of the respective organization IDs. The vagueness of the question formulation should be noted (see Figure 9).

Responses to the question, ‘What areas of your organization are already using organization IDs?’ show that the use of organization IDs was differentially distributed across the various

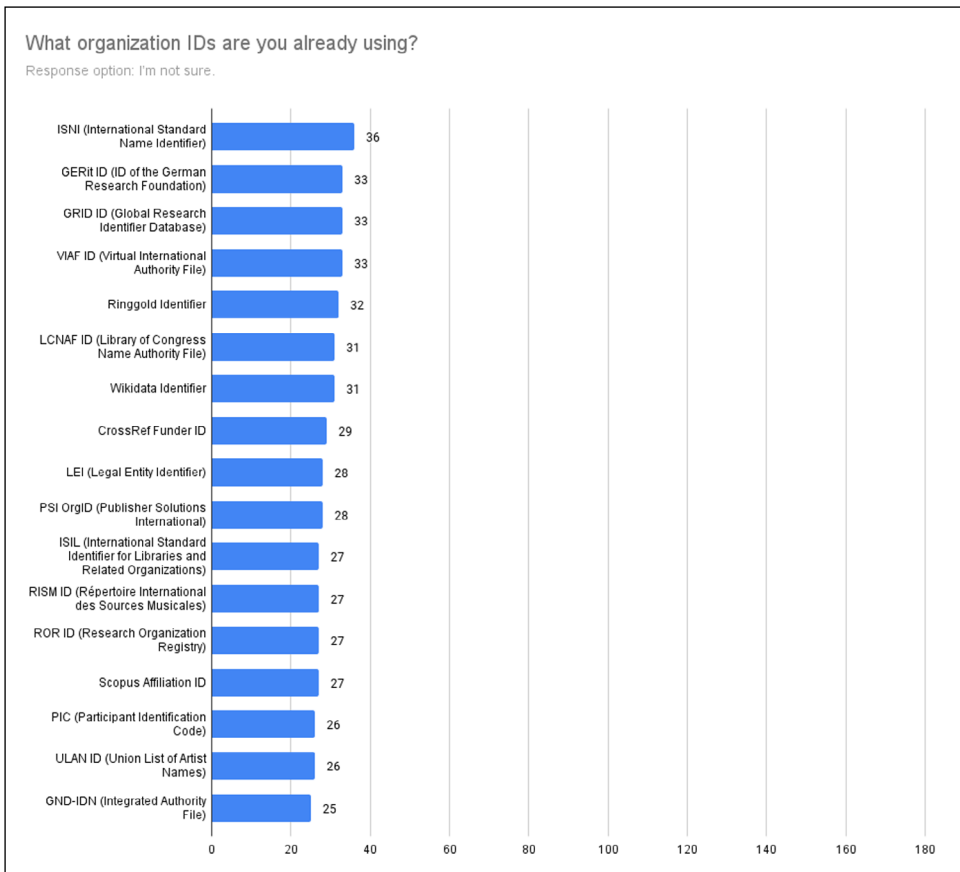


Figure 9 Distribution and number of responses to the question, 'What organization IDs are you already using?' (Response option: I'm not sure.)

areas of an organization, but the overall level of use was low. Thus, only 38 of the institutions' respondents reported that organization IDs were used in their open access repositories (N = 122). The second most frequent use of organization IDs (reported by 30 respondents) was in university bibliographies (N = 101); the third most frequent use of organization IDs (reported by 26 respondents) was in publication databases (N = 93). Compared with information infrastructures, administration departments tended to use organization IDs less for administrative software (12 respondents; N = 85) or identity management systems (5 respondents; N = 78). Overall, organization IDs were already in use in 164 information infrastructures or administration systems in Germany. The use of organization IDs was planned in 292 such areas and was not planned in 209 such areas. Multiple selection was possible; 141 institutions answered all response options offered in the question (see Figure 10).

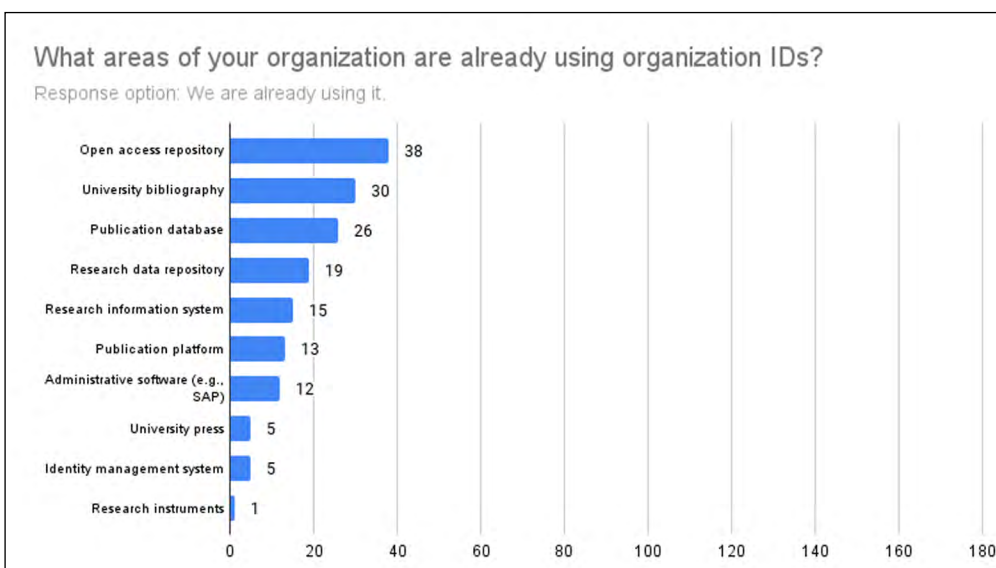


Figure 10 Distribution and number of responses to the question, 'What areas of your organization are already using organization IDs?'

The distribution of organization IDs in the various software solutions corresponded to the heterogeneous distribution of the organization IDs across the various areas. As a classical

repository application, OPUS, with 40 responses, was the software most frequently mentioned in response to the question, ‘In what software is an organization ID implemented or is an organization ID to be implemented?’ followed by the editorial management system Open Journal Systems (OJS), with 37 responses. In addition, the diversity in the various systems was reflected in the high number of ‘Other’ responses, which included not only a range of in-house developments but also systems such as FACTScience and Open Monograph Press. Multiple selection was possible. The question was answered by 177 institutions (see Figure 11).

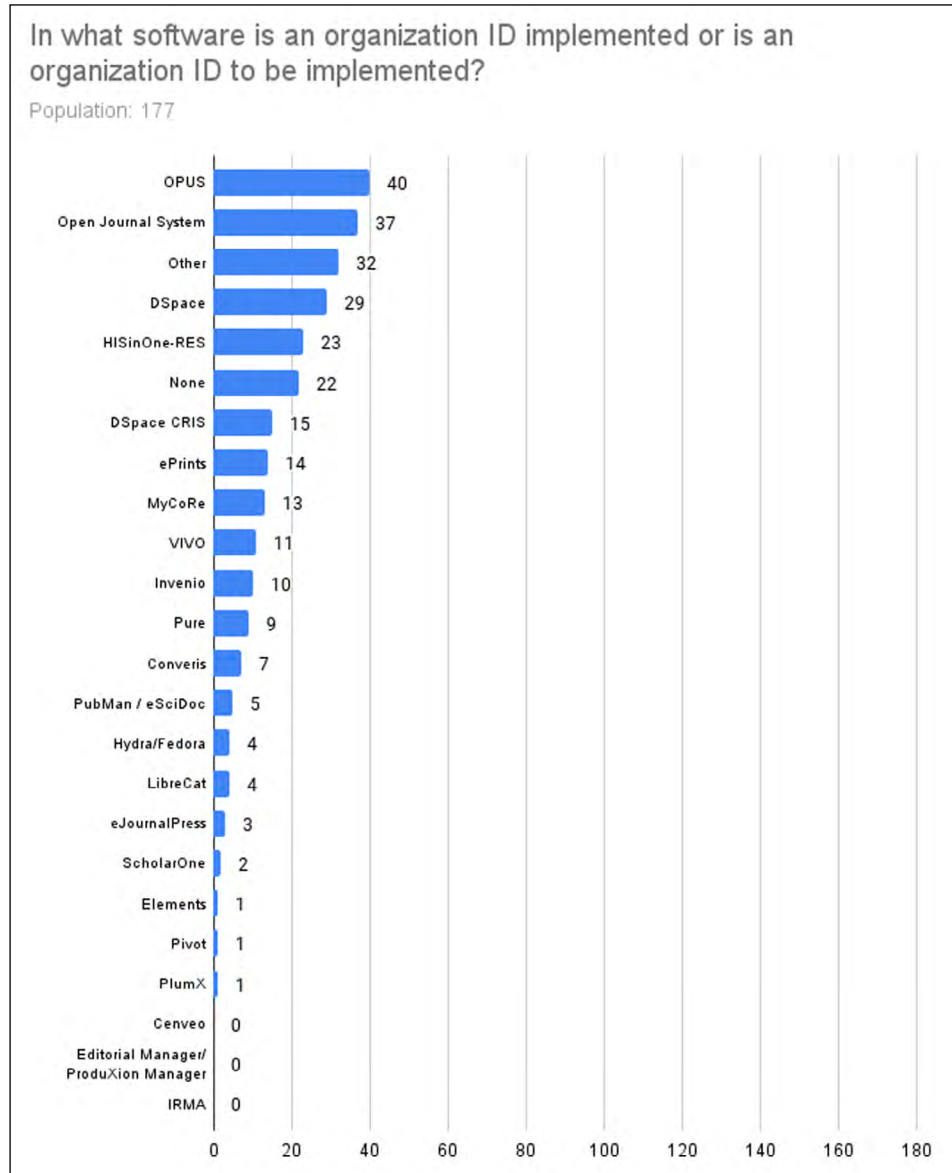


Figure 11 Distribution and number of responses to the question, ‘In what software is an organization ID implemented or is an organization ID to be implemented?’

The overall picture that emerges from responses to the question, ‘Are organization IDs integrated in your institutional guidelines (policies) [and] if yes to what extent?’ is that only very few scientific institutions in Germany had integrated organization IDs into their institutional guidelines. The respondents reported that organization IDs were currently anchored in just 16 policies (10 affiliation policies, 3 research data or research data management guidelines, and 3 publication regulations or guidelines). Overall, there were plans to integrate organization IDs into 46 guidelines of the scientific institutions in Germany that participated in the survey (15 research data or research management guidelines, 13 publication regulations or guidelines, 3 doctoral degree regulations or guidelines, and 3 ‘Other’). Multiple selection was possible; 139 institutions answered all response options offered in the question. However, because of the question structure (not a mandatory question, matrix structure, and a lack of a ‘None’ option), the populations of the individual response options varied (see Figure 12). These differences were taken into account in the analysis.

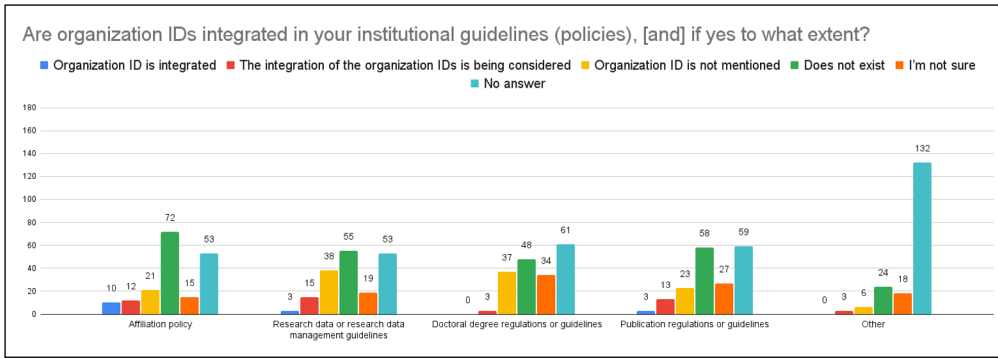


Figure 12 Distribution and number of responses to the question, ‘Are organization IDs integrated in your institutional guidelines (policies) [and] if yes to what extent?’

4.4. NEEDS

The distribution of responses to the question, ‘What need is there for characteristics of an organization ID for your application scenario?’, paints a picture of diverse needs for characteristics among the respondents. The need for 5 of the 14 response options offered was reported by over half of the respondents to be very high or high. For 132 respondents, ‘global uniqueness and persistence’ was the most important characteristic for their institutions’ application scenarios. ‘Machine readability of the metadata (e.g., via API)’ was relevant for 123 respondent institutions. The response ‘Metadata can be downloaded and reused (e.g., CSV, JSON, RDF)’ was chosen by 107 institutions. Almost as many (106) institutions chose the response option ‘Metadata are (re)usable under a free license.’ The fifth characteristic needed by the majority of respondents was interoperability (n = 101). Contrasting with these five favored characteristics, the three least reported needs were granularity (n = 60), versioning (n = 59), and curation by the operator of the organization ID directory (n = 58). Multiple selection was possible; 148 institutions answered all response options offered in the question. However, because of the question structure (not a mandatory question, matrix structure, and the lack of a ‘None’ option), the populations of the individual response options varied (see Figure 13). These differences were taken into account in the analysis.

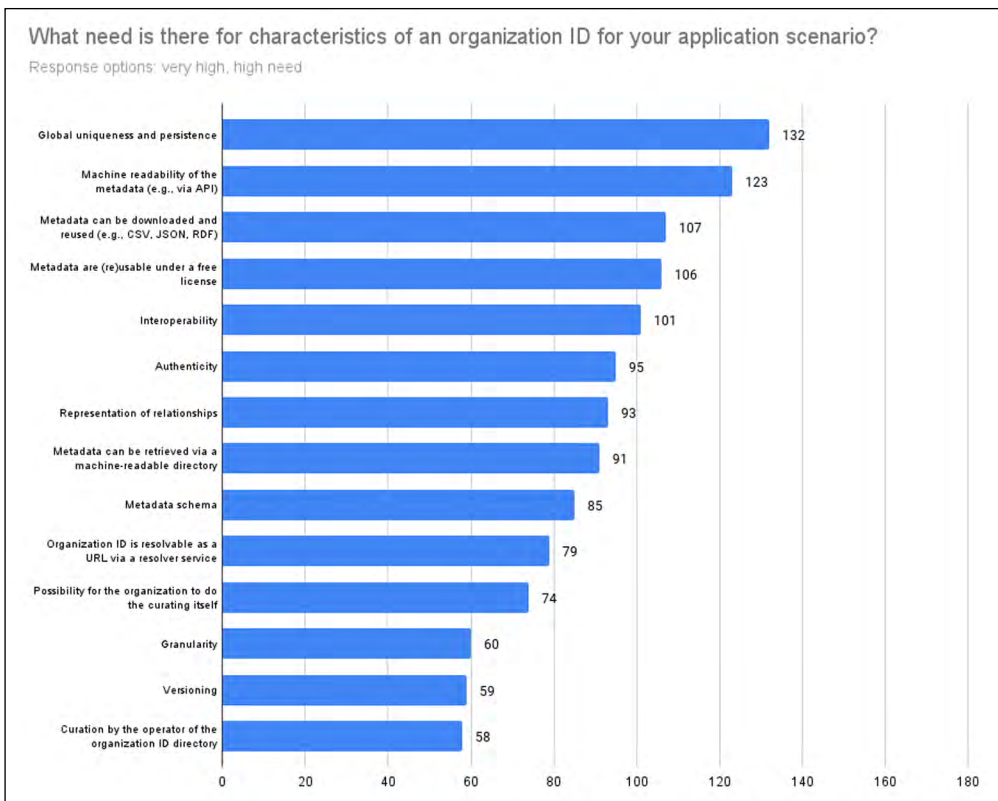


Figure 13 Distribution and number of responses to the question, ‘What need is there for characteristics of an organization ID for your application scenario?’

Responding to the question of how granularly the organization structure should be represented in the metadata, there was a clear preference (n = 72) for the option ‘Down to the third organizational level (e.g., institute).’ The majority of the respondents that chose this option were higher education institutions and Leibniz institutes. Thirteen universities of applied sciences, 31

universities, and 16 Leibniz institutes that participated in the survey considered coverage down to the third organizational level to be most important level of granularity in the case of an organization ID. The question was answered by 156 institutions (see Figure 14).

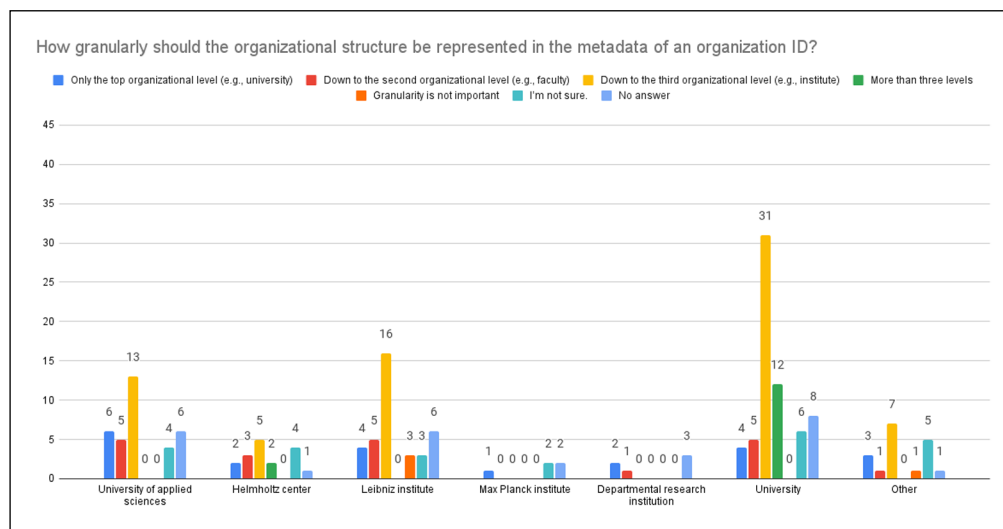


Figure 14 Distribution and number of responses to the question, 'How granularly should the organizational structure be represented in the metadata of an organization ID?'

In response to the question as to what relationships should generally be provided in the metadata of an organization ID, the majority of respondents (n = 130) reported that organizational relationships (e.g., university-faculty-institute) were most important. Relationships with publications (n = 103) and relationships with persons (n = 102) were also reported by many respondents. The response option 'Relationships with events' was chosen by the fewest respondents (n = 37). Furthermore, the answers given by respondents in the free-text field under 'Other' show that 'predecessor-successor relationships between organizations' was the only relationship that could not be directly selected from among the response options offered (however, it could also have been indirectly reported via the response option 'Temporary relationships'). It can therefore be assumed that the relationships listed in the question and the above-mentioned relationships represent the currently desired relationships that should be provided in the metadata of organization IDs. Multiple selection was possible. The question was answered by 167 institutions (see Figure 15).

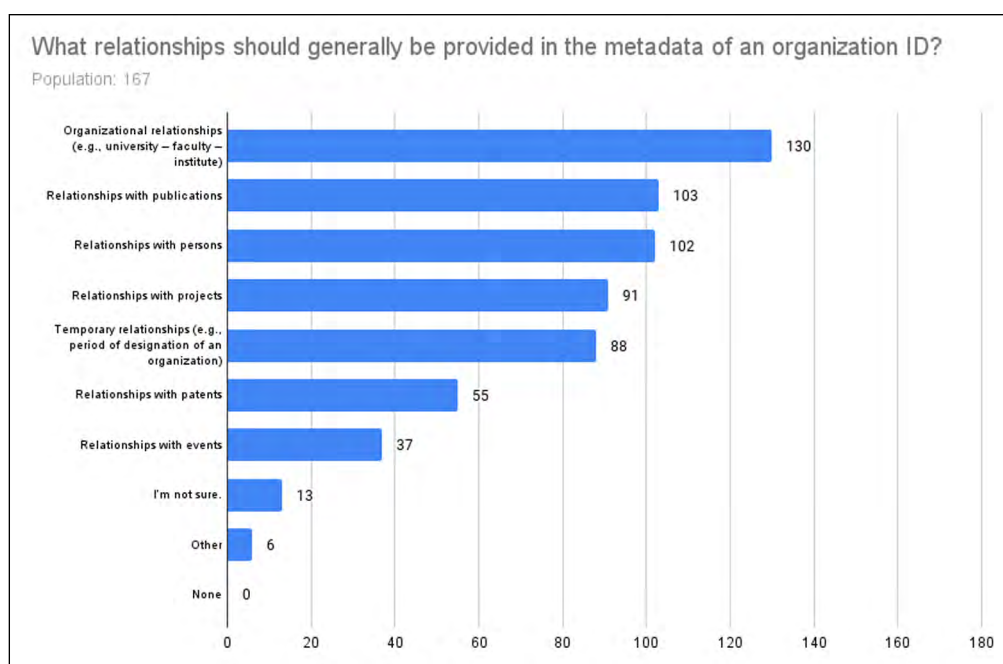


Figure 15 Distribution and number of responses to the question, 'What relationships should generally be provided in the metadata of an organization ID?'

In response to the question, 'What relationships between organizations should be definable?', the relationship 'formerName/laterName' was the one that was most desired among the participants (n = 112). The purpose of such a link is to relate datasets of organizations that were

renamed in the past to each other. Chosen by 104 respondent institutions, the relationship 'isMemberOf/hasMember' was the second most desired relationship. Ninety-seven participants reported the need to be able to link subunits with an organization via 'isUnitOf/hasUnit.' Multiple selection was possible. The question was answered by 167 institutions (see Figure 16).

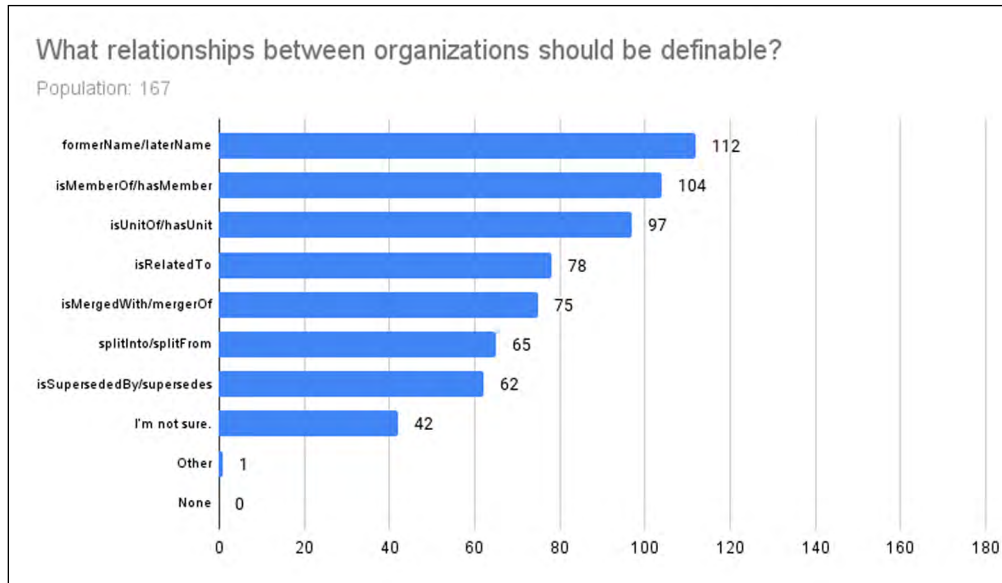


Figure 16 Distribution and number of responses to the question, 'What relationships between organizations should be definable?'

In response to the question as to what relationships between persons and organizations should be definable, 103 respondents chose the option 'isAffiliatedWith/hasEmployee,' which was thus the relationship most preferred by the participating institutions. The second most required relationship between persons and organizations was 'IsMemberOf/hasMember,' which was chosen by 98 respondents. Multiple selection was possible. The question was answered by 167 institutions (see Figure 17).

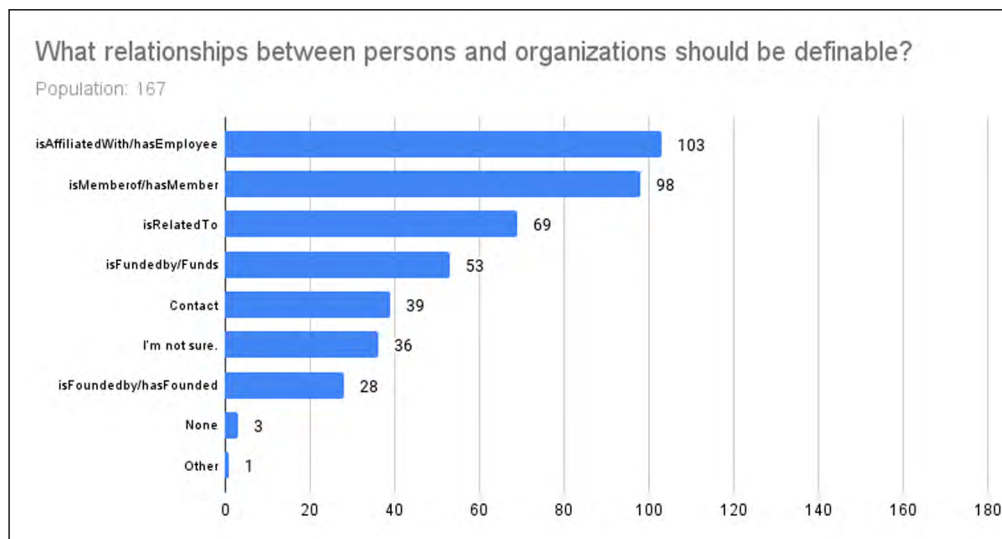


Figure 17 Distribution and number of responses to the question, 'What relationships between persons and organizations should be definable?'

Responding to the question, 'What metadata should be assignable to an organization ID?', the indication of the 'Preferred name' within a metadata schema for organization IDs was most important to most of the participants (n = 145). The response options 'Name abbreviations/acronyms' and 'Name variants' were selected by a similar number of participants (n = 143 and n = 140, respectively). Almost three quarters of the respondents (n = 122) reported that the metadata should include the 'Name in other languages' and the 'URL of the organization.' The indication of the country code (n = 115) and the period of designation of the organization (n = 100) were the only other organization-related metadata desired by the majority of respondents. Multiple selection was possible. The question was answered by 167 institutions (see Figure 18).

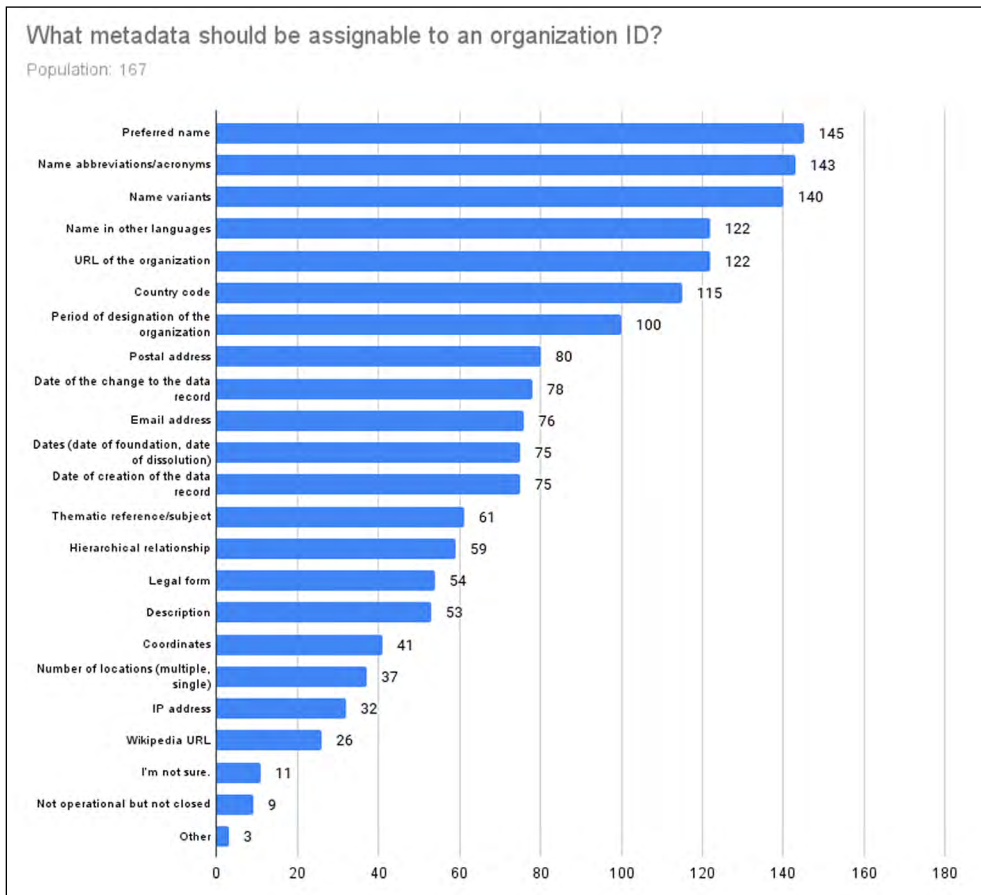


Figure 18 Distribution and number of responses to the question, 'What metadata should be assignable to an organization ID?'

In response to the question, 'What name form of an organization should be the preferred and primarily displayed name form that is presented together with the organization ID?,' the name form that was by far the most preferred by the participants was 'Always the name form of the organization by which it wishes to be identified' (n = 74). Thirty-seven respondent institutions preferred 'Always the national-language name form of the organization,' and 22 respondents opted for 'Always the English-language name form because it is an international database.' The question was answered by 153 institutions (see [Figure 19](#)).

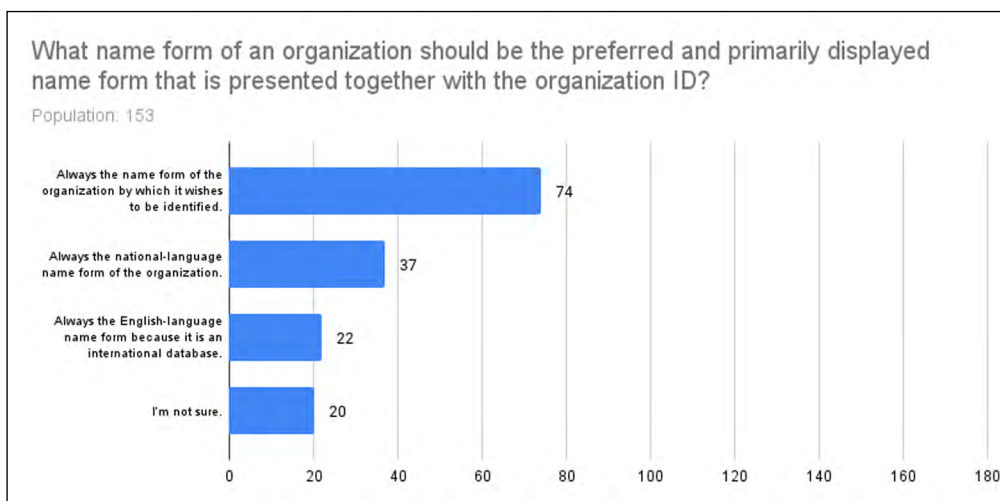


Figure 19 Distribution and number of responses to the question, 'What name form of an organization should be the preferred and primarily displayed name form that is presented together with the organization ID?'

In response to the question as to what actor should have decision-making authority over the curation (e.g., updating) of the metadata of an organization ID, the majority of respondents chose the option 'Organizations should be able to curate the metadata of their records themselves' (n = 105). This clearly reflects the aspiration for sovereignty over the metadata of their own institutions intimated in the responses to the question about name form. The response option 'The metadata should be maintained by the service for organization IDs' was the second

most often selected response (n = 23). The responses given in the free-text field under ‘Other’ favored, among other things, a hybrid form comprising curation by the organization itself and by the service for organization IDs. In addition, curation ‘as an internationally coordinated task of the national libraries’ was suggested. The question was answered by 144 institutions (see Figure 20).

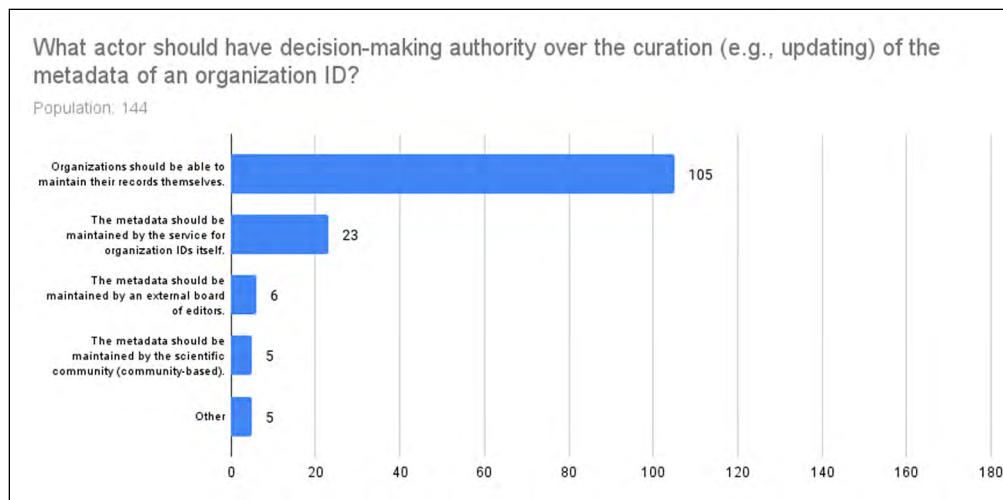


Figure 20 Distribution and number of responses to the question, ‘What actor should have decision-making authority over the curation (e.g., updating) of the metadata of an organization ID?’

In response to the question, ‘Supposing your organization could maintain the metadata of your organization ID itself, what actor at your organization would maintain your organization’s record?’, the only response option selected by the majority of respondents designated the library as the responsible actor (n = 107). Here, it should be noted that when asked about their work areas, the main group of persons that participated in the survey selected the response option ‘Library staff member’ (n = 149; see Figure 2). Accordingly, some of the library staff did not in turn consider that responsibility for the curation of the metadata should lie with the library but rather with the administration department (n = 65), the data center (n = 12), or the scientific departments (n = 11), whereas others were not sure (n = 38) or selected the ‘Other’ option (n = 6). Multiple selection was possible. The question was answered by 167 institutions (see Figure 21).

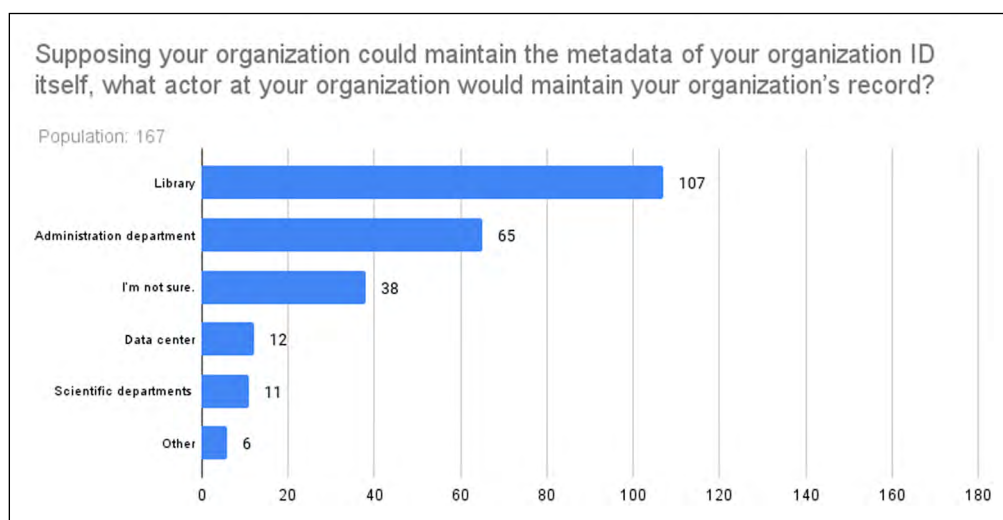


Figure 21 Distribution and number of responses to the question, ‘Supposing your organization could maintain the metadata of your organization ID itself, what actor at your organization would maintain your organization’s record?’

In response to the question, ‘What aspects of the governance of the service for organization IDs are important to your organization?’, 126 respondent institutions reported that the service should be operated in accordance with open and transparent statutes.¹⁴ The second most important aspect (n = 112) was that the service should not be operated on a for-profit basis. Also of importance for the majority of respondents was that the software of the service for

14 One example of a PID service that is already operated according to open and transparent statutes is [ORCID Inc.](https://info.orcid.org/what-is-orcid/) for author identification. See ‘Our Founding Principles,’ <https://info.orcid.org/what-is-orcid/>.

organization IDs should be open source and available under a free license (n = 100). The only aspect among the response options offered that was considered less important by respondents (n = 47) was that the service should be operated by actors from the scientific community themselves. Multiple selection was possible. The question was answered by 167 institutions (see Figure 22).

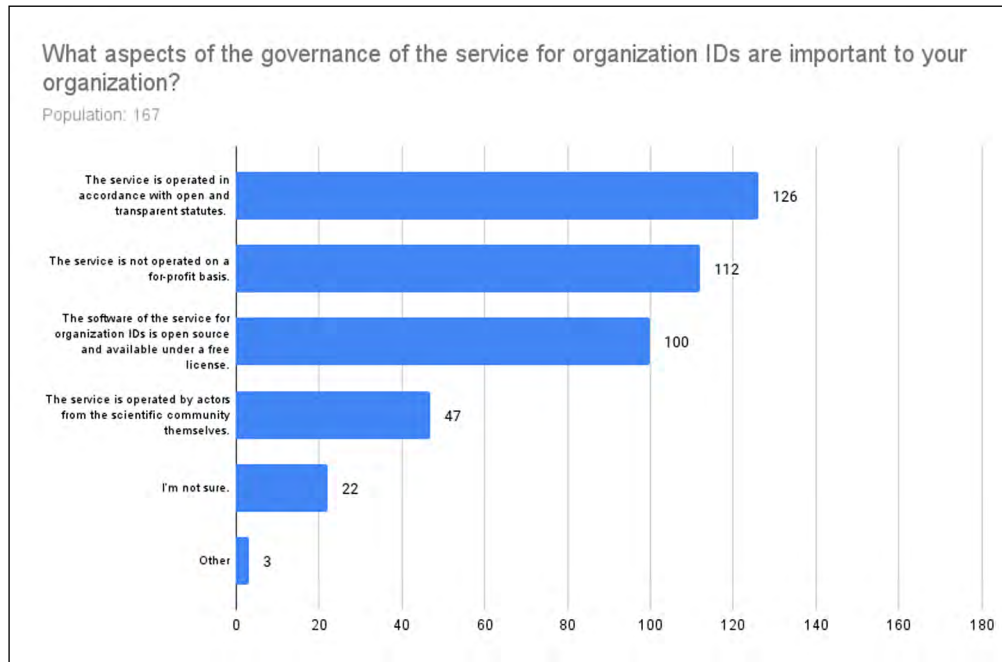


Figure 22 Distribution and number of responses to the question, 'What aspects of the governance of the service for organization IDs are important to your organization?'

In response to the question as to what aspects of the financially sustainable operation of a service for organization IDs were important to their organizations, 99 respondents reported that the service should be accessible free of charge. The second most important aspect was that the revenue of the service should be based on services and not on the metadata (n = 48). A similar number of respondents (n = 42) were of the opinion that the service should be financed by membership fees. That this question gave rise to some uncertainty can be seen from the fact that 43 respondent institutions reported that they were not sure. The response option that the service should be crowdfunding-based was an important financially sustainable aspect of the operation of a service for organization IDs for only four institutions, and thus for a minority. Multiple selection was possible. The question was answered by 167 institutions (see Figure 23).

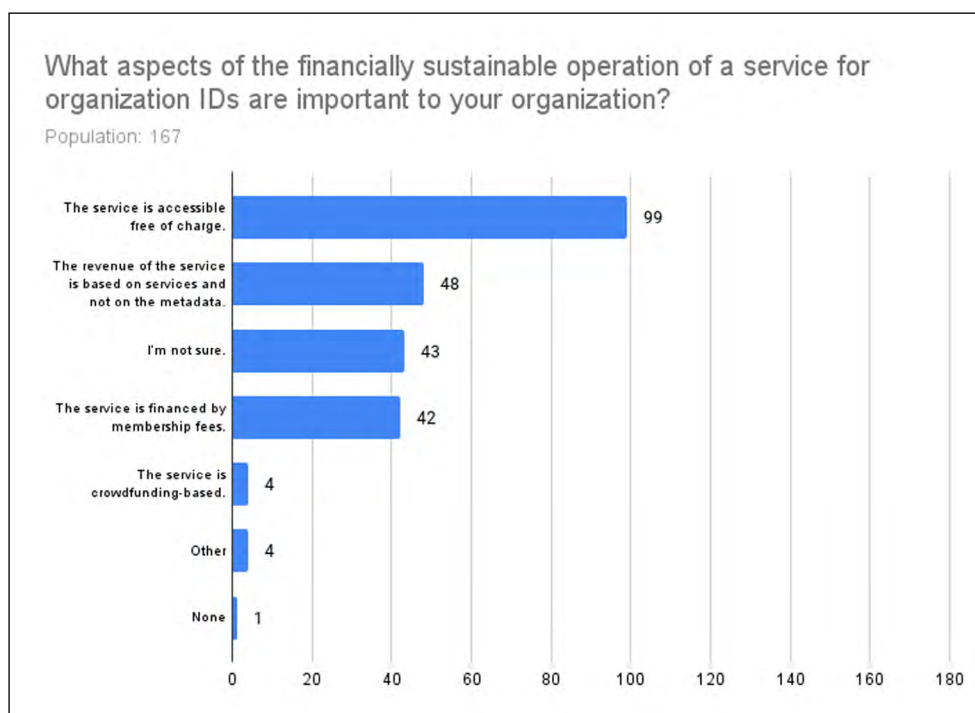


Figure 23 Distribution and number of responses to the question, 'What aspects of the financially sustainable operation of a service for organization IDs are important to your organization?'

In response to the question, ‘Would fee-based membership be an option for your organization in the case of a service for organization IDs?’, the majority of respondents (n = 103) chose the option ‘I’m not sure.’ Thirty-five respondents answered ‘yes,’ and 21 answered ‘no.’ The question was answered by 159 institutions (see Figure 24).

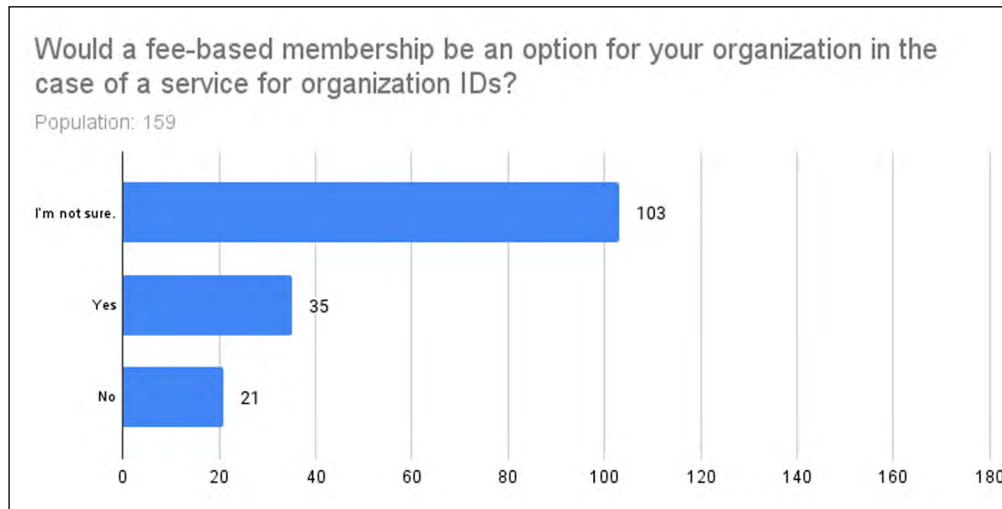


Figure 24 Distribution and number of responses to the question, ‘Would a fee-based membership be an option for your organization in the case of a service for organization IDs?’

At the end of the survey, respondents were asked, ‘For which of the following resources/products do you think further persistent IDs should be developed?’ Taking the ratings ‘Very high need’ and ‘High need’ together, the greatest desideratum was for a PID for publishers (n = 85). The second highest need was seen for PIDs for projects (n = 79). The least need was seen for PIDs for samples (n = 26). PIDs for research infrastructures (n = 70), conferences (n = 58), research software (n = 44), and research instruments (n = 42) were in the middle of this range. Multiple selection was possible. One hundred and thirty-six institutions answered all the response options offered in the question. However, due to the question structure (not a mandatory question, matrix structure, and the lack of a ‘None’ option), the populations of the individual response options varied (see Figure 25). These differences were taken into account in the analysis.

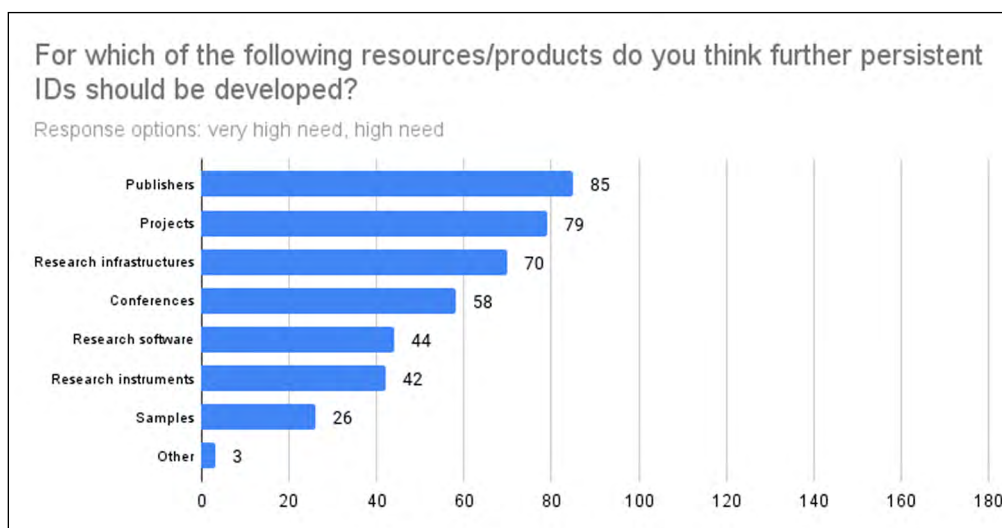


Figure 25 Distribution and number of responses to the question, ‘For which of the following resources/products do you think further persistent IDs should be developed?’

5. ANALYSIS

5.1. EXPERIENCE WITH ORGANIZATION IDS

The GND proved to be the best-known organization ID among the scientific institutions in Germany that participated in the survey. This is not surprising; thanks to the German National Library (DNB), the library networks of the German-speaking countries, and other library services, the GND has a high level of adoption in its application as a national identifier for scholarly authors (inter alia through the mandatory deposit of dissertations with the DNB; see Figure 5).

5.2. USE

The fact that the use of organization IDs is differentially distributed across the areas of an organization is not necessarily related to the topic of organization IDs per se, but rather fundamentally depends on whether or not the respective infrastructures or systems exist at the institution. In the case of the question on the use of organization IDs, the distribution of the answers to the response option ‘We are not using it, but we are planning to use it’ was similar to that in the case of the already existing use of organization IDs, whereby the increase as a result of the planned implementation in the various areas constitutes a doubling or trebling of the level of use (see [Figure 10](#)).

With regard to the implementation of organization IDs, it can be noted that, on balance, there were more implementations with open-source software solutions. However, this is probably attributable to their generally high level of adoption (see [Figure 11](#)).

The level of integration of organization IDs into institutional guidelines was rather low. This might be related, on the one hand, to a low level of awareness of the necessity of organization IDs in policies. On the other hand, the distribution of the response option ‘Does not exist’ suggests that these guidelines simply do not exist at many institutions (see [Figure 12](#)).

In sum, a heterogeneous picture emerged with regard to the technical implementation of organization IDs in information infrastructures and their anchoring in policies of scientific institutions in Germany. A wide range of organization IDs are used in different areas and software solutions, whereby the GND is currently the most widely adopted organization ID and the ROR ID is the organization ID whose use is most frequently planned. At present, organization IDs are anchored only to a small extent in guidelines of scientific institutions in Germany. However, analogous to the technical implementation, an increased integration of organization IDs is also to be expected here.

5.3. NEEDS

It is to be expected that the clear vote in favor of the coverage of the organization structure down to the third level is related to the complex organizational structures, for example, of universities (see [Figure 14](#)). Moreover, it is apparent that the representation of relationships between organizations in the metadata of organization IDs is an important desideratum for scientific institutions in Germany (see [Figure 15](#)).

The distribution of the answers to the question as to what relationships between persons and organizations should definable shows that relationships between organizations and persons (affiliation use case) are deemed to be particularly important in the case of organization IDs. Thus, the majority of the respondent institutions expressed a desire that the affiliation of the researchers with the institution should be definable (see [Figure 17](#)).

In the case of the question, ‘What metadata should be assignable to an organization ID?’ it is noteworthy that most of the metadata fields desired by the majority of respondents addressed the topic of the organization name. By contrast, the other characteristics of an organization, such as its legal form, hierarchical relationships, or the description of the institution, were considered to be of only secondary importance (see [Figure 18](#)).

Responses to the question, ‘What name form of an organization should be the preferred and primarily displayed name form that is presented together with the organization ID?’ show that the respondents considered that decision-making authority over the name form should primarily lie with the institution itself (see [Figure 19](#)).

In sum, it can be stated with regard to the metadata of organization IDs that the relationships between organizations and the relationships between organizations and persons (affiliation use case) are considered particularly important. In the case of the organization-related details desired by the majority of respondents, the metadata focus on name forms, which in turn should be determined by the institutions themselves. Because the name form may only be an isolated and nonrepresentative example of the desire to curate the metadata—that is, to maintain the metadata record of the organization—the participants were also asked directly about curation. Here it became clear that the majority of the participating institutions considered

that responsibility for the maintenance of the metadata of their records should lie with the organizations themselves (see [Figure 20](#)). Although the majority of respondents reported that the maintenance of the metadata of the institution should be carried out by the library, the relatively high percentage of respondents who also considered that the administration department should be responsible appears surprising (see [Figure 21](#)).

The distribution of responses to the question as to what aspects of the governance of the service for organization IDs were important to the organization shows that importance is attached to openness and a not-for-profit approach (see [Figure 22](#)).

With regard to the question as to whether a fee-based membership would be an option for the participating organizations in the case of a service for organization IDs, what is interesting about the distribution of the responses is that there was no clear picture of the mood in one or another direction (see [Figure 24](#)). Thus, the acceptance of a membership-based service for organization IDs in Germany remains an open question.

Overall, the questions on the subject of governance show that for the respondent institutions, a service for organization IDs should be open and freely accessible, although the survey participants were not sure whether a fee-based membership would be an option for their institutions.

In the case of the final question as to the resources or products for which further persistent IDs should be developed, the results suggest that the heterogeneity of the reported needs is attributable to the different and discipline-specific application scenarios or research fields of the various institution types (see [Figure 25](#)).

6. DISCUSSION AND RECOMMENDATIONS

The results of the present survey provide a comprehensive view of the current state of knowledge and use of organization IDs at scientific institutions in Germany. Organization IDs are considered useful or are already used in diverse application scenarios—mainly, however, to unambiguously name an institution and to link it to persons affiliated with the institution or to other organizations (affiliation use case). Both familiarity with and the use of different organization IDs are very heterogeneous. As the level of use is considerably lower than the level of familiarity, there is currently great potential for adoption. This future potential is seen in open PIDs such as the ROR ID.

The findings about the needs with regard to relationships, metadata, and the curation of organization IDs will also prove very helpful when it comes to designing metadata schemas for organization IDs. It is noteworthy that the majority of respondents considered that decision-making authority over the metadata (e.g., name forms) should lie with the institution itself. This is in turn reflected in the aspiration that the institutions themselves should curate the metadata of their organization IDs. In view of the fact that respondents expressed a preference for granularity of coverage down to the third level of the organizational structure (e.g., institutes), this aspiration appears to be justified because the documentation of the complex and volatile structures of an institution can be permanently and efficiently kept up to date in the metadata of an organization ID only by the institution itself.

Openness, free accessibility, and a not-for-profit orientation were the core needs with regard to the governance and operation of a service for organization IDs. However, it was unclear to the respondent institutions whether and how such a service could be sustainably financed.

To promote the use of organization IDs in Germany, the following recommendations of ACTION were identified: **A**wareness **C**onnectivity **I**ntegrati**O**n **N**etworking. As a first step, awareness of the usage, value, and benefits of an organizational identifier needs to be raised within the research community. This will lay the foundation for the following step of fostering connectivity among persistent identifiers. PIDs can only unleash their full potential when they are connected to each other. Open infrastructures should therefore enable the interlinking of PIDs by fostering interoperability. In turn, this will increase integration of PIDs in information infrastructures of research organizations and universities. In addition

to technical integration, policies play an important role in the spread of organization IDs. Proof can be found in countries where national PID strategies are in place or being worked on.¹⁵ Finally, networking of all community stakeholders involved is essential to a sustainable future of every PID.

7. CONCLUSION AND OUTLOOK

Looking at the findings of the survey, it is determined that ACTION is needed to make organization IDs an integral part of the research life cycle. The survey is one step in contributing to the visibility of open organization identifiers and shows that a current and future need is seen for PIDs for further resources and products. To this issue even further, the project partners of ORCID DE have submitted a grant proposal for the project PID Network Germany to the DFG that has been accepted for funding.

Starting in 2023, the project PID Network Germany will foster the networking of the PID community through a variety of events, the development of PID metadata guidelines, and the creation of a national PID strategy.

The authors would like to thank all the institutions and their representatives who participated in the survey.

8. RESEARCH DATA

The research data from this survey have been made publicly accessible in anonymized form via Zenodo (Vierkant 2021).

ACKNOWLEDGEMENTS

The authors thank the reviewers for their helpful suggestions. Parts of this article have already been published in German (Vierkant et al. 2022).

FUNDING INFORMATION


The writing of this paper was supported by a grant from the German Research Foundation (project number: 429866380). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.


COMPETING INTERESTS

PV works for DataCite, one of the three organizations that run the Research Organization Registry (ROR). All other authors have no competing interests.

AUTHOR AFFILIATIONS

Paul Vierkant  orcid.org/0000-0003-4448-3844
DataCite, DE

Antonia Schrader  orcid.org/0000-0001-7080-634X
Helmholtz Association, Helmholtz Open Science Office, DE

Heinz Pampel  orcid.org/0000-0003-3334-2771
Helmholtz Association, Helmholtz Open Science Office, DE

¹⁵ Examples are the Australian 'Persistent identifiers policy.' Last checked on 25 November 2022, <https://zenodo.org/record/6529768#.Y4DbSi9XZhE>, and the UK 'national persistent identifier research strategy'. Last checked on 25 November 2022, <https://www.jisc.ac.uk/rd/projects/a-national-persistent-identifier-research-strategy>.

- Bertelmann, R**, et al. 2019. ORCID DE 2—Konsolidierung der ORCID-Informationsinfrastruktur in Deutschland [ORCID DE 2—Consolidation of the ORCID information infrastructure in Germany]. [Last accessed 14 October 2021]. DOI: <https://doi.org/10.2312/lis.20.01>
- Bertelmann, R**, et al. 2021. ORCID DE 2—Consolidation of the ORCID information infrastructure in Germany. Available at <https://gepris.dfg.de/gepris/projekt/429866380?language=en> [Last accessed 6 September 2021].
- Bilder, G, Brown, J** and **Demeranville, T**. 2016. Organisation identifiers: Current provider survey. [Last accessed 14 October 2021]. DOI: <https://doi.org/10.5438/4716>
- Cruse, P, Haak, L** and **Pentz, E**. 2016. Organization identifier project: A way forward. [Last accessed 14 October 2021]. DOI: <https://doi.org/10.5438/2906>
- Fuchs, C, Pampel, H** and **Vierkant, P**. 2017. ORCID in Deutschland—Ergebnisse einer Bestandsaufnahme im Jahr 2016 [ORCID in Germany—Results of an assessment of the status quo in 2016]. *O-bib. Das offene Bibliotheksjournal* [The Open Library Journal], Ed. VDB 4(2): 35–55. DOI: <https://doi.org/10.5282/o-bib/2017H2S35-55>
- ORCID DE**. 2021. Available at <https://www.orcid-de.org> [Last accessed 6 September 2021].
- ORCID Inc**. 2021. About ORCID. Available at <https://info.orcid.org/what-is-orcid/> [Last accessed 6 September 2021].
- Rothfritz, L**. 2018. Persistent identifier für wissenschaftliche Einrichtungen [Persistent identifiers for scientific institutions]. Bachelor's thesis, Potsdam University of Applied Sciences. Available at <https://nbn-resolving.org/urn:nbn:de:kobv:525-20147>.
- Vierkant, P**. 2021. Forschungsdaten und Fragebogen zur 'Umfrage zum Bedarf und Nutzung von Organisations-IDs an Hochschulen und außeruniversitären Forschungseinrichtungen in Deutschland' [Research data from and questionnaire for the 'Survey on the need for and use of organization IDs at higher education institutions and non-university research organizations in Germany']. DOI: <https://doi.org/10.5281/zenodo.5467980> [Last accessed 10 June 2022].
- Vierkant, P, Schrader, A** and **Pampel, H**. 2022. Organisations-IDs in Deutschland—Ergebnisse einer Bestandsaufnahme im Jahr 2020. *Bibliothek Forschung und Praxis*, 46(1): 191–215. [Last accessed 10 June 2022]. DOI: <https://doi.org/10.1515/bfp-2021-0089>

TO CITE THIS ARTICLE:

Vierkant, P, Schrader, A and Pampel, H. 2022. Organization IDs in Germany—Results of an Assessment of the Status Quo in 2020. *Data Science Journal*, 21: 19, pp. 1–23. DOI: <https://doi.org/10.5334/dsj-2022-019>

Submitted: 10 June 2022

Accepted: 27 November 2022

Published: 22 December 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.