



# Landscape Analysis of PID Practices in NFDI

Report on Survey “PID Landscape in the NFDI”

D1.1

Sara El-Gebali

15 January 2025

## Imprint

<b>Authored by</b>	Sara El-Gebali ( <a href="https://orcid.org/0000-0003-1378-5495">https://orcid.org/0000-0003-1378-5495</a> )
<b>Contributions by</b>	Jana Böhm ( <a href="https://orcid.org/0009-0004-9802-113X">https://orcid.org/0009-0004-9802-113X</a> ), Stephanie Hagemann-Wilholt ( <a href="https://orcid.org/0000-0002-0474-2410">https://orcid.org/0000-0002-0474-2410</a> ), Torsten Kahlert ( <a href="https://orcid.org/0009-0003-3264-5006">https://orcid.org/0009-0003-3264-5006</a> ), Steffi Genderjahn ( <a href="https://orcid.org/0000-0002-8912-184X">https://orcid.org/0000-0002-8912-184X</a> ), Marc Lange ( <a href="https://orcid.org/0000-0002-7742-3867">https://orcid.org/0000-0002-7742-3867</a> ),
<b>Date of publication</b>	15 January 2025
<b>Version</b>	Version 1.0
<b>DOI</b>	<a href="https://doi.org/10.5281/zenodo.14652295">https://doi.org/10.5281/zenodo.14652295</a>
<b>License</b>	This work is licensed under <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
<b>Deliverable</b>	D1.1

### About

PID4NFDI (<https://base4nfdi.de/projects/pid4nfdi>) is the basic service for persistent identifiers in development for the German National Research Data Infrastructure (NFDI). PID4NFDI is part of and funded through Base4NFDI.

Funded by DFG as part of NFDI. DFG Grant Number: 521466146



# Contents

1. Preamble	1
1.1. Introduction to PID4NFDI and Survey Motivation	1
2. Overview	2
2.1. Response Rate	2
2.2. Respondents	5
2.3. Structure of Survey	6
3. Findings and Key Insights	7
3.1. Governance Challenges	7
3.2. PID Management and Staffing Issues	8
3.3. Challenges in PID Policy and Licensing	12
3.4. Affiliation with PID Providers	14
3.5. Overview of Services and Infrastructures	17
3.6. Maturity Vs. PID Decision-making Roles	18
3.7. Strategic Decision-making Roles	19
3.8. Advisory Roles	19
3.9. Infrastructure Types and Purpose	22
3.10. Interoperability and Metadata Alignment Needs	24
3.11. Training and Outreach	31
3.12. Engagement with PID Training	31
3.13. Barriers to Training Engagement	32
3.14. Training Formats and Preferences	33
3.15. Demand for Specific Training Topics	34
3.16. Organizational Support for Metadata Completeness	36
3.17. Outreach Activities for PID Awareness	37
4. Expanding PID Adoption in NFDI: Challenges, Opportunities, and the Role of the PID Coordination Hub	40
4.1. Methodological Limitations	40
4.2. Challenges and Opportunities in PID Adoption	41
4.3. Trends in PID Registration and Future plans	42
4.4. The Need for Improved Coordination and Collaboration	44
5. Summary and Conclusion	45
5.1. Key Takeaways	45
5.2. The Role of the PID Coordination Hub in Addressing Challenges	47
5.3. Implications for the NFDI Landscape	49
6. Supplemental material	49
7. Appendix	50
7.1. PID4NFDI Awareness, Involvement, and Support Expectations	50
7.2. Landscape of Services/Infrastructures in NFDI	54

## Landscape Analysis of PID Practices in NFDI

7.3. PID Provider Affiliation by Organization	57
7.4. Resource Types and Trends	62
7.5. Metadata	64
7.6. Interoperability and API Usage	64
7.7. Training	67
7.8. PID Management Support	69

## 1. Preamble

A digital persistent identifier (PID) is a globally unique and resolvable string of characters that is permanently and unambiguously linked to a digital object. Accompanying the PID is comprehensive metadata that describes the object in detail, often including bibliographic information and data about its origin, creation, and subsequent use.

PIDs are essential for effective data management and adhering to the FAIR principles (Findable, Accessible, Interoperable, and Reusable)<sup>1</sup>.

By capturing connections among research resources, contributors, institutions, projects, and other entities, PIDs greatly enhance the dissemination, quality assurance, and reproducibility of research.

Their implementation enriches metadata, reduces administrative overhead, and ultimately improves the discoverability and interoperability of research data worldwide, benefiting both data producers and users.

### 1.1. Introduction to PID4NFDI and Survey Motivation

The survey<sup>2</sup>, titled "PID Service Landscape in the NFDI," was conducted during the initial funding phase of the PID4NFDI project in 2024. The main objective of PID4NFDI is to establish a foundational PID service for Germany's National Research Data Infrastructure (NFDI).

PID4NFDI is a collaborative initiative led by DataCite, the Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG), the Helmholtz Open Science Office, and the Technische Informationsbibliothek (TIB), Leibniz Information Centre for Science and Technology. The consortium works closely with experienced PID infrastructure users, to leverage their insights and amplify the impact of PID adoption within NFDI workflows. For more information about the

---

<sup>1</sup> Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

<sup>2</sup> The full question catalog is available at <https://doi.org/10.5281/zenodo.1432774>

foundational work that paved the way for this project, please refer to the relevant publications<sup>3 4 5</sup>.

Central to this initiative is the development of the PID Coordination Hub, a comprehensive solution designed to standardize PID practices, improve governance, and support alignment across different consortia and organizations. This integrated approach ensures greater consistency and adoption throughout the NFDI landscape.

The survey aimed to evaluate the current level of PID adoption and integration among NFDI consortia and to understand their future needs. By identifying these needs and pinpointing existing gaps, the PID4NFDI team seeks to develop a PID service that aligns with the evolving requirements of the NFDI community.

One of the primary challenges within NFDI is effective coordination among stakeholders. This requires fostering a shared understanding of each other's initiatives to facilitate collaboration beyond organizational and consortium boundaries. Building awareness of PID4NFDI's goals is crucial for the project's success.

The project highlights the significance of PIDs and promotes their timely adoption to support FAIR data management.

## 2. Overview

### 2.1. Response Rate

The survey, conducted from April 9th to May 12th, 2024, gathered 34 complete responses from a diverse set of organizations and NFDI consortia. Although it does not capture the entire scope of all PID integration activities across NFDI, the survey

---

<sup>3</sup>Bingert, Sven, Jan Brase, Felix Burger, Britta Dreyer, Stephanie Hagemann-Wilholt, Paul Vierkant, and Philipp Wieder. 'Concept for Setting up the Persistent Identifier Services Working Group in the NFDI Section "common Infrastructures"'. Zenodo, 30 April 2022. <https://doi.org/10.5281/zenodo.6507760>.

<sup>4</sup>Schrader, Antonia C., Daniel Arend, Janete Bach, Kirsten Elger, Sandra Göller, Stephanie Hagemann-Wilholt, Rolf Krah, et al. 'Workshop on Pids Within NFDI'. Zenodo, 13 February 2023. <https://doi.org/10.5281/zenodo.7635905>.

<sup>5</sup>Schrader, Antonia C., Stephanie Hagemann-Wilholt, and Andreas Czerniak. 'Isn't a Number and a URL Enough? Why Pids Matter and Technical Solutions Alone Are Not Sufficient.'. Presented at the 1st Conference on Research Data Infrastructure (coRDI), Karlsruhe, Germany, 8 September 2023. <https://doi.org/10.5281/zenodo.8328807>.

provides a representative snapshot due to the broad participation of consortia and key infrastructure stakeholders (**Figure 1**). Notably, 12% of the respondents represented multiple consortia, adding further depth to the survey insights.

The primary audience for the survey consisted of providers and managers of research data infrastructures developed or operated by the NFDI consortia. Invitations were sent directly to respondents or through consortium leads.

Among the respondents, 97% (33 out of 34) expressed their willingness to participate in follow-up interviews to provide further insights into PID practices, challenges, and needs.

All consortia represented in the survey have at least one respondent willing to participate, reflecting a positive attitude towards collaboration and mutual support.

## Landscape Analysis of PID Practices in NFDI

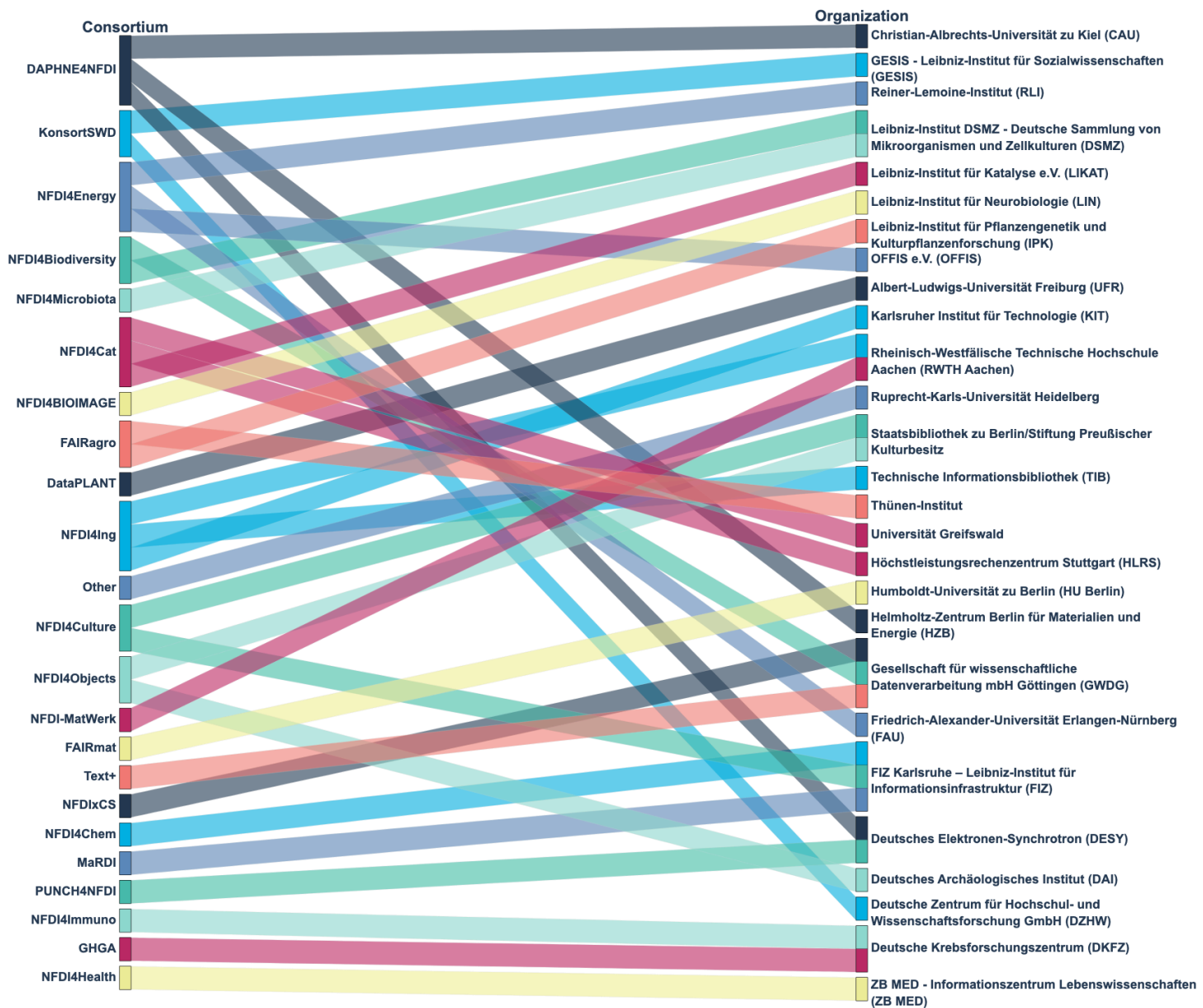


Figure 1 illustrates the consortia involved (on the left) and their associations with various organizations (on the right), highlighting the extensive reach of the survey within the NFDI landscape.



## Landscape Analysis of PID Practices in NFDI

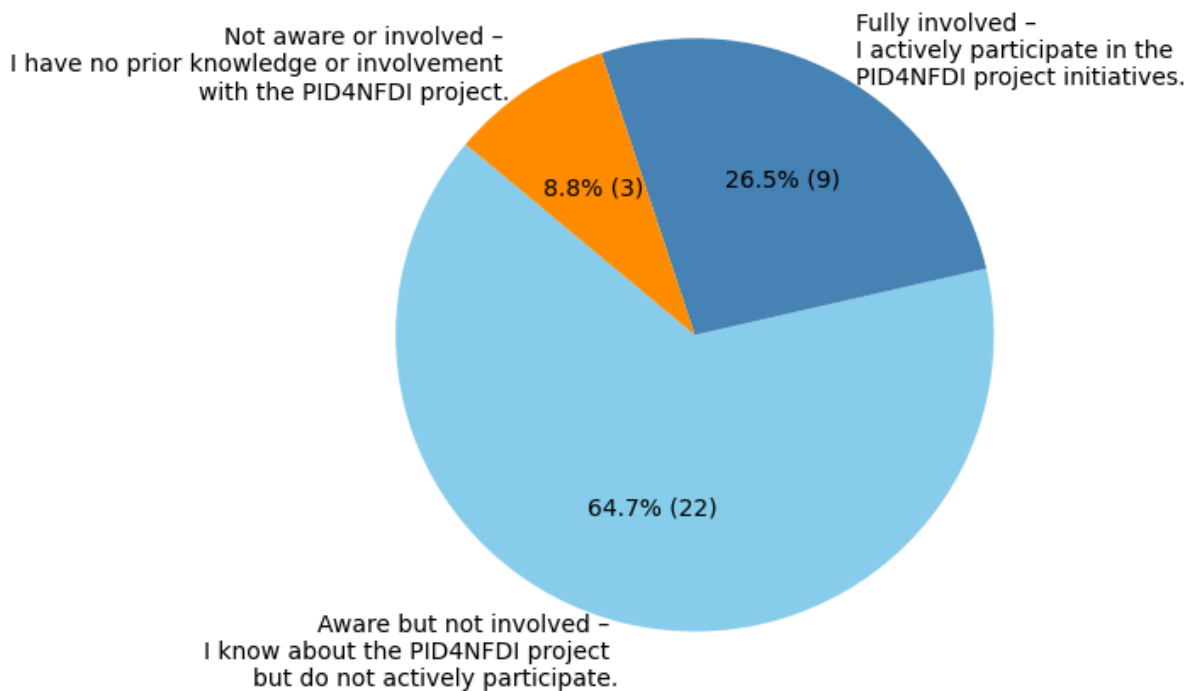


Figure 2 displays the involvement of survey respondents in the PID4NFDI project.

Taken together, **Figures 1 and 2** highlight respondents' engagement with the PID4NFDI project, revealing that 91.2% of participants are either directly involved in or aware of the project. Many respondents are actively participating as use-case partners, reflecting the strong visibility of PID4NFDI even at this early stage.

### 2.2. Respondents

We have categorized the roles described by the respondents as “leadership and coordination”, “technical implementation and development”, “research data management” and “other roles”, allowing respondents to fall into multiple categories.

Several individuals occupy leadership roles such as task area leaders (8 mentions), (co-)spokespersons (5 mentions), and task/project managers (4 mentions). These roles oversee consortia activities, lead task areas, and manage projects within consortia. Many respondents are also involved in technical aspects, including co-developing services and platforms (6 mentions), managing platforms/repositories (3 mentions), and leading technical measures. Tasks include developing and maintaining tools for data management, simulation services, and scientific workflows. In research data management, roles include data managers, project managers, and data stewards.

## Landscape Analysis of PID Practices in NFDI

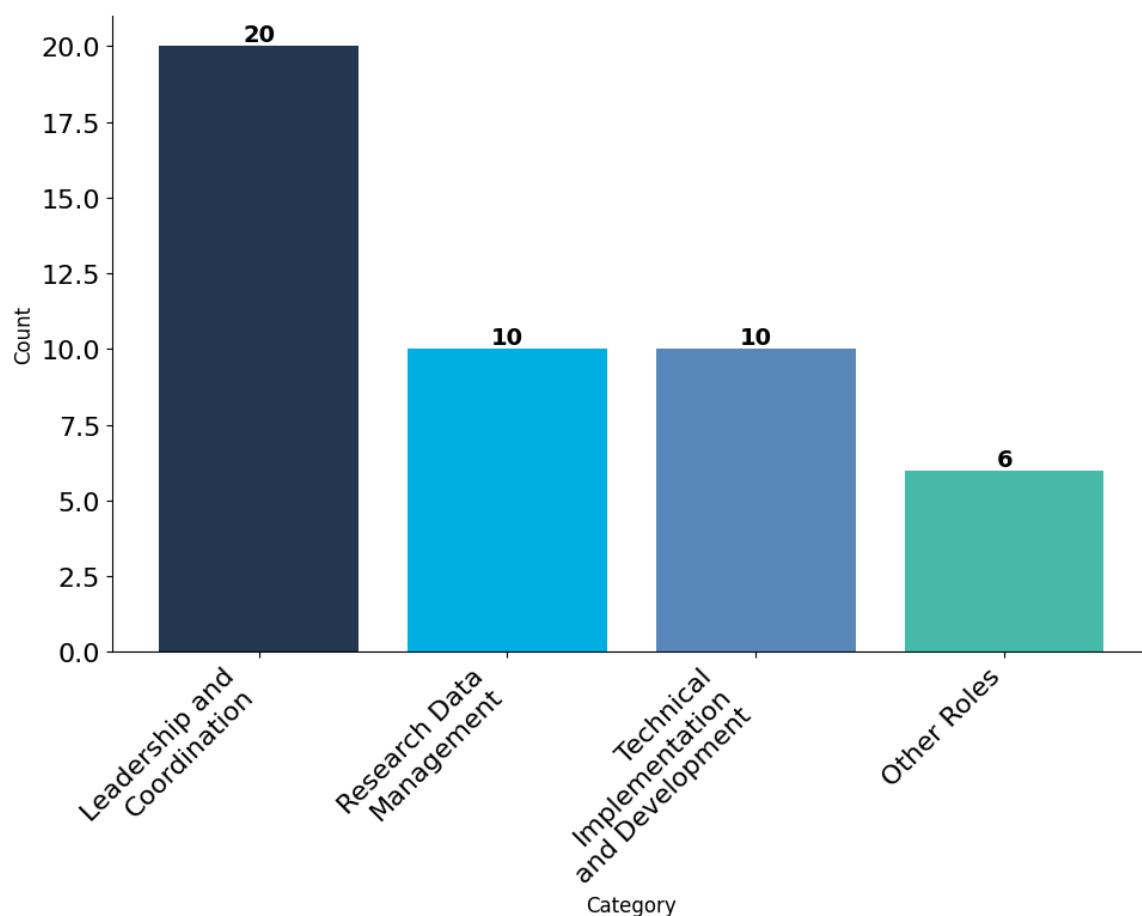


Figure 3 displays the distribution of categorized roles of respondents.

Overall, the survey successfully engaged individuals in a variety of roles and with diverse areas of expertise.

### 2.3. Structure of Survey

The core of the survey was structured into ten main question groups. In addition to the (1) background of the respondents, we asked about (2) their awareness of, involvement in and support expectations of PID4NFDI to identify support needs and collaboration opportunities. Furthermore, we asked for (3) a detailed description of the infrastructures/services in use or in preparation for the respective consortia; followed by questions on (4) current and planned PID system integrations. The key focus of the questions related to PID4NFDI's main areas of work: (5) metadata practices, (6) enabling Open Science, (7) technical interoperability, (8) governance issues, (9) outreach and (10) training and support.

## 3. Findings and Key Insights

### 3.1. Governance Challenges

The persistence of identifiers depends largely on how they are managed. Along with technical interoperability and metadata harmonization, governance criteria are a critical factor for a sustainable integration and use of PIDs. Therefore, we asked about the existence of formal governance structures for and main challenges of the management of PIDs.

A key finding of the survey is the lack of formal governance structures dedicated to PID management in many organizations. This gap directly impacts PID implementation and makes it challenging to maintain a cohesive strategic direction across NFDI. **Figure 4** shows responses to the question, "Does your organization have a formal governance structure dedicated to PID management?" The data shows that many organizations lack such governance, leading to inconsistencies in PID implementation.

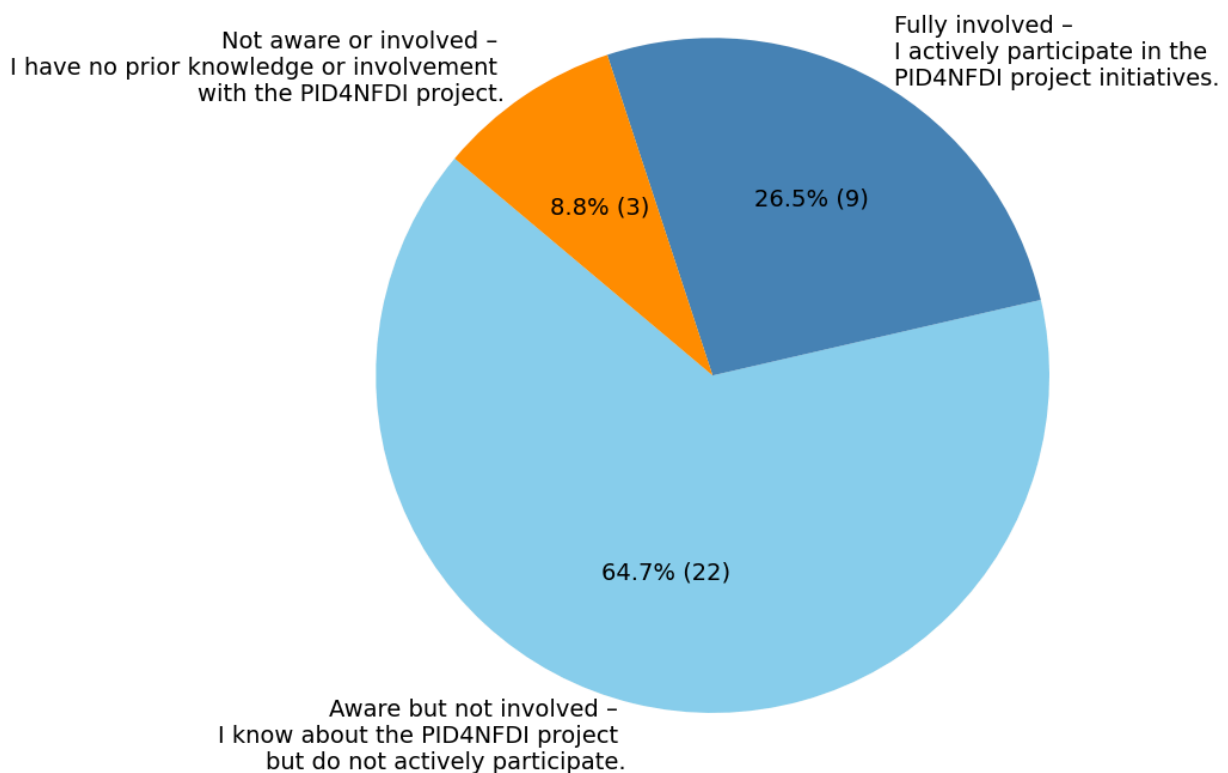


Figure 4: Visualization of responses regarding the presence of formal governance structures or committees for PID policy management and oversight within organizations.

### 3.2. PID Management and Staffing Issues

The survey also included a question to further assess the roles of individuals specifically involved in PID decision-making within their organizations categorized into various roles:

- **Strategic Decision-Maker:** Directly responsible for making final decisions on PID implementation strategies and policies.
- **Advisory Role:** Providing recommendations or expert advice influencing PID strategy and policy decisions.
- **Technical Implementation and Support:** Involved in the technical aspects of PID system integration, support, and maintenance.
- **Policy and Standards Development:** Participating in the development of policies and standards related to PID usage and management.
- **Operational Management:** Overseeing the day-to-day operations and management of PID systems.
- **Data Management and Curation:** Managing and curating data that uses PIDs, ensuring best practices in metadata and data linkage.
- **Research and Development:** Engaged in researching new PID technologies or methodologies for organizational adoption.

## Landscape Analysis of PID Practices in NFDI

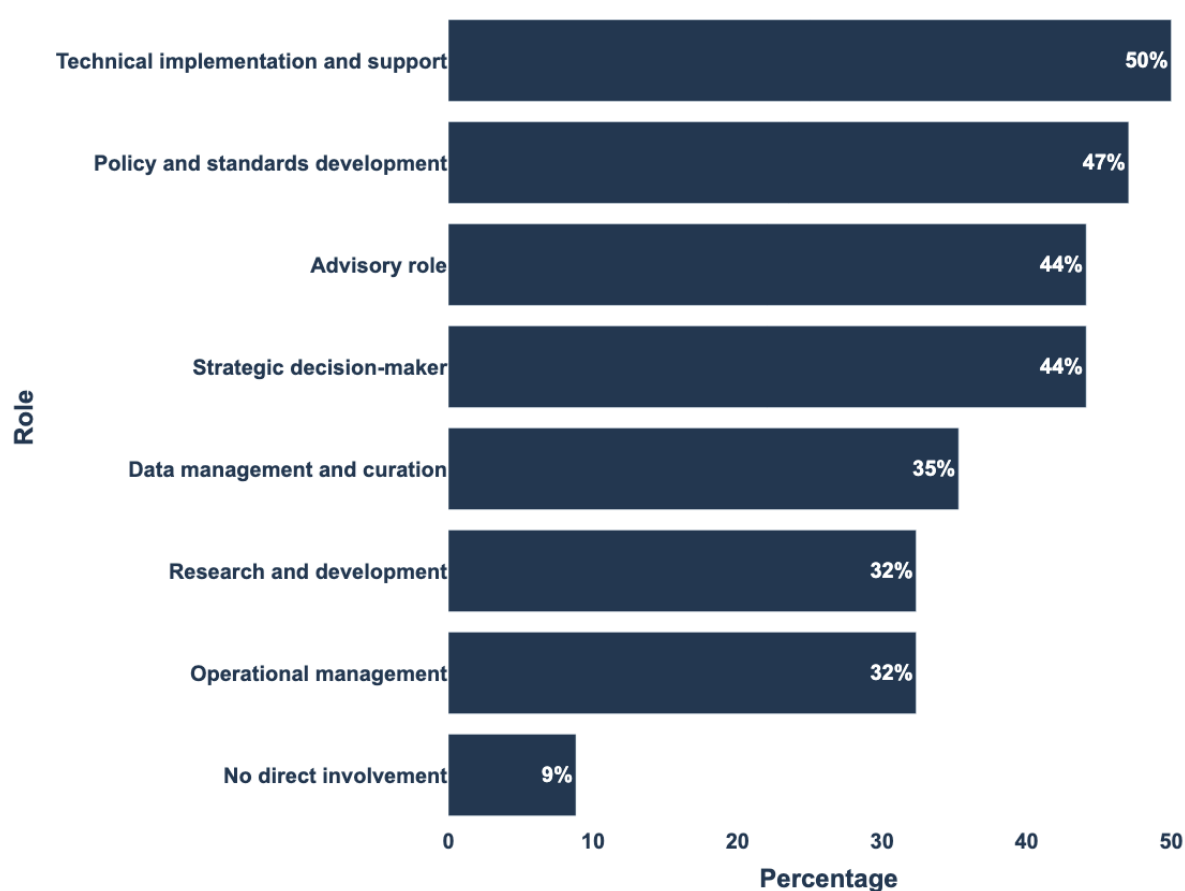


Figure 5 displays PID decision-making roles.

The survey revealed that respondents often held multiple roles, with the highest involvement observed in technical implementation and support, policy and standards development, advisory roles, and strategic decision-making (**Figure 5**).

When asked about the primary challenges faced in PID governance, a predominant issue highlighted was the lack of dedicated budget, limiting the ability to secure long-term financial resources crucial for PID management (**Figure 6**). This financial instability is further exacerbated by the reliance on project-based funding, complicating efforts to maintain consistent governance and operational continuity.

## Landscape Analysis of PID Practices in NFDI

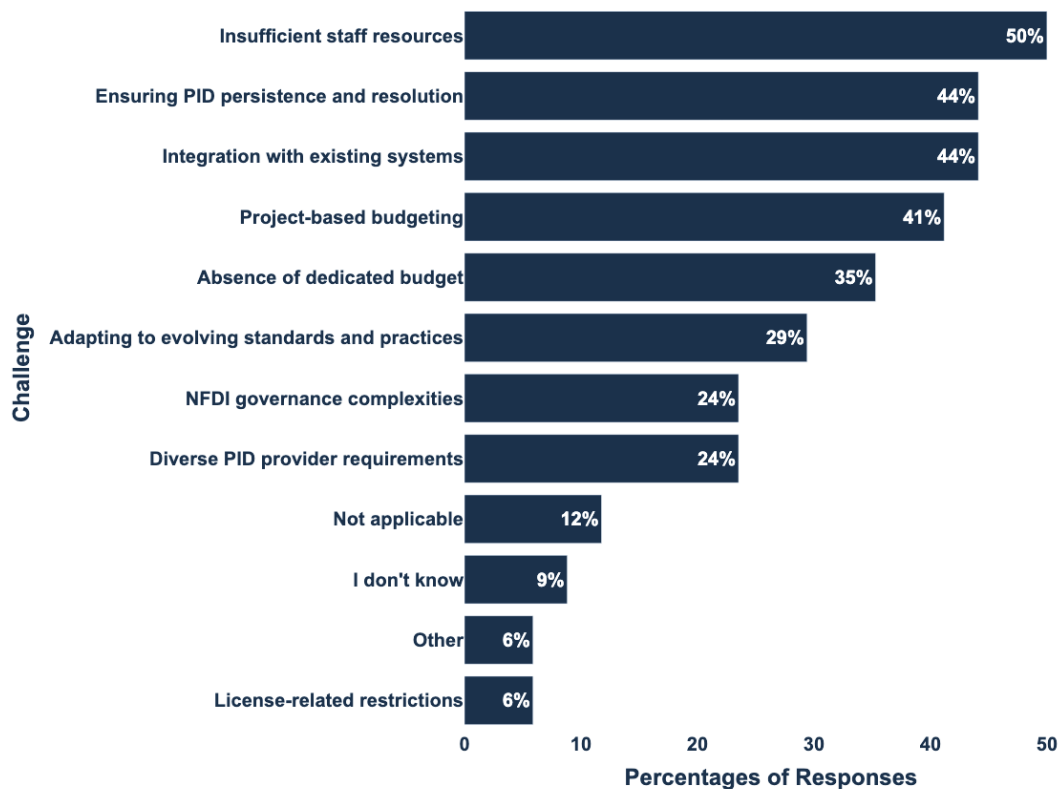


Figure 6 illustrates challenges in PID Management.

The sustainability of PID initiatives is substantially constrained by several interrelated factors identified across various organizational roles. Insufficient staffing resources emerged as one of the most critical barriers, reported by 50% of respondents. Limited personnel hinder the effective integration of PIDs into existing systems and impede adaptation to evolving standards and practices (**Figure 6**). Ensuring PID persistence and resolution remains an ongoing challenge, crucial for maintaining data integrity and accessibility in the long term. These challenges were especially emphasized by respondents in Technical Implementation and Support roles, as well as those in Advisory roles (**Figure 7**).

## Landscape Analysis of PID Practices in NFDI

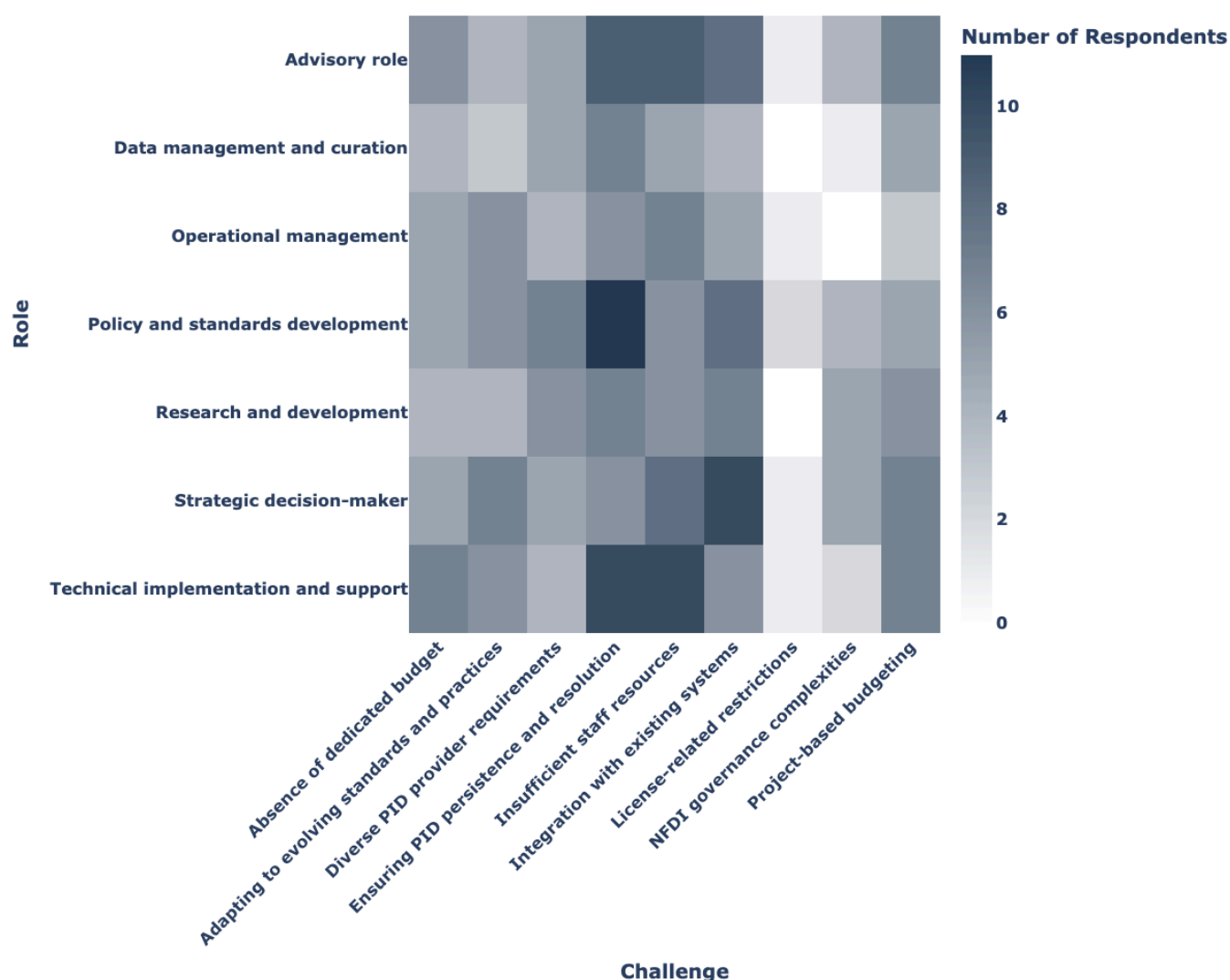


Figure 7 displays the roles in PID decision-making processes versus challenges in PID management. This heat map provides a visual correlation between specific roles and the intensity of challenges they face, emphasizing role-specific prominence of certain issues.

Furthermore, **Figure 7** illustrates that the integration with existing information systems presents additional complexities, particularly for Strategic Decision-Makers, Advisory roles, and those involved in Policy and Standards Development. Aligning diverse metadata requirements across NFDI's complex structure, with its various communities and standards, adds substantial administrative and financial burdens. These challenges often conflict with broader community or infrastructural needs.

The survey further highlights difficulties stemming from the diverse requirements of PID providers, each governed by unique membership models and registration terms. This variability increases the operational burden on NFDI consortia, particularly when PID-supported services are distributed across multiple teams.

This fragmentation results in governance inconsistencies and diverging practices. Moreover, the distinct metadata requirements of each PID necessitate mapping to the existing NFDI infrastructure while accommodating specific community needs.

Collectively, these issues contribute to a fragile environment for PID management, especially considering the intended persistence and reliability of these identifiers. Addressing financial and staffing gaps and improving alignment with PID provider requirements is crucial to ensure PIDs support FAIR data practices across NFDI.

### 3.3. Challenges in PID Policy and Licensing

Organizations also face issues with establishing PID policies due to the absence of clear chains of responsibility. While Open Science policies are common, often spearheaded by library or institutional initiatives, PID-specific policies are critically less developed. The PID Coordination Hub could play a key role in driving focused advocacy and supporting the development of such policies.

Licensing of content and metadata is one of the crucial factors for making research outputs such as research data and software FAIR and enabling their sharing and re-use. To support findability, at least the bibliographic core metadata describing the content should be available and accessible under a free license<sup>6</sup> to further promote the realization of Open Science.

The majority of respondents reported using open licenses, with CC BY and CC0 being the most prevalent across repositories and services (**Figure 8**). Free-text responses highlighted additional practices, such as the use of customized licenses, data usage contracts, and specialized frameworks.

One notable example is the RADAR Metadata Schema<sup>7</sup>, which offers flexibility through a comprehensive array of licensing options. These include various Creative Commons licenses, public domain dedications, and licenses tailored for software. This diversity in licensing reflects efforts to accommodate the specific needs of different communities and data types.

---

<sup>6</sup> Restrictions due to data protection, copyright, competition law and security aspects must be taken into account.

<sup>7</sup> FAIRsharing.org: RADAR Metadata Schema, DOI: [10.25504/FAIRsharing.e26f92](https://doi.org/10.25504/FAIRsharing.e26f92).



## Landscape Analysis of PID Practices in NFDI

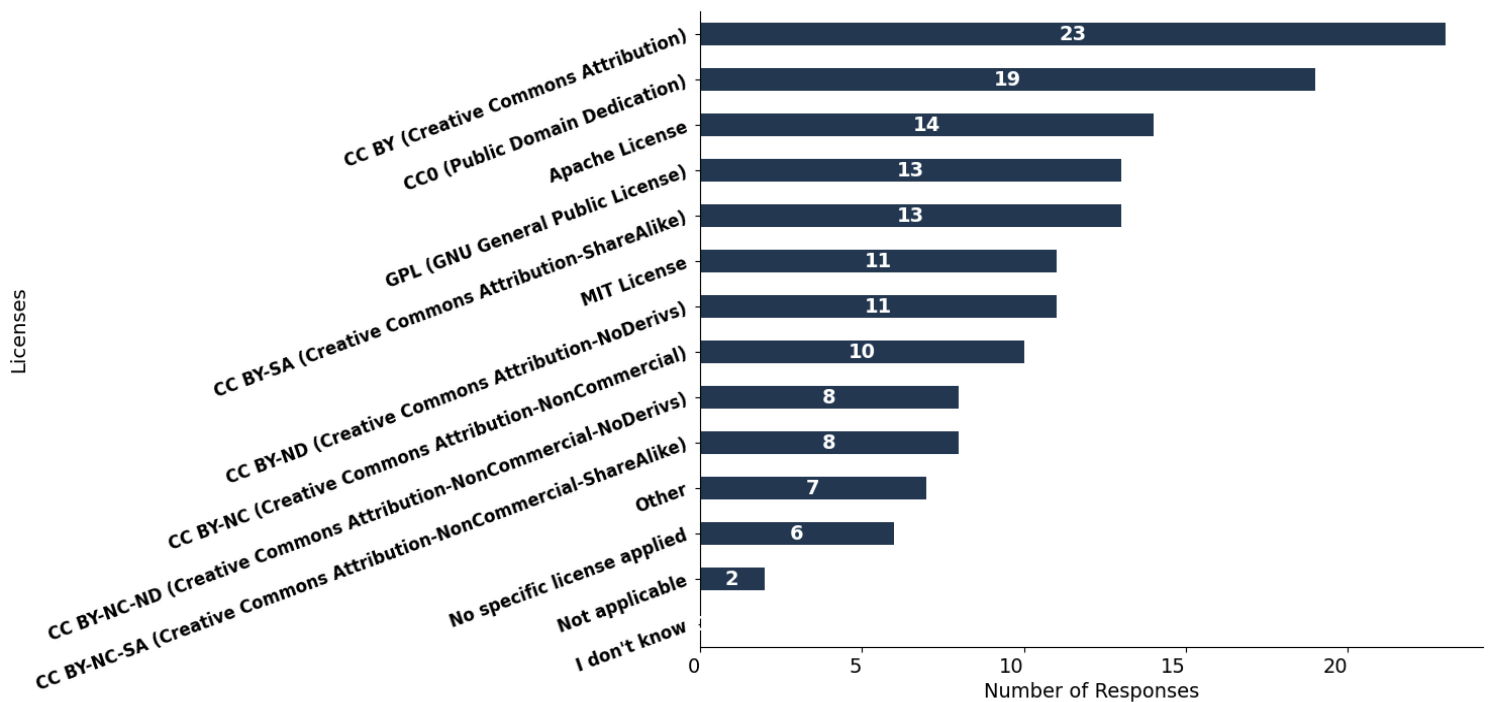


Figure 8 displays trends in openness and use of permissive licenses.

According to survey data, 50% of respondents reported that their institution has established policies or guidelines to govern the use of licenses for research outputs, which include datasets, publications, and software (**Figure 9**). This highlights the widespread normalization of licensing practices that are essential for promoting open research.

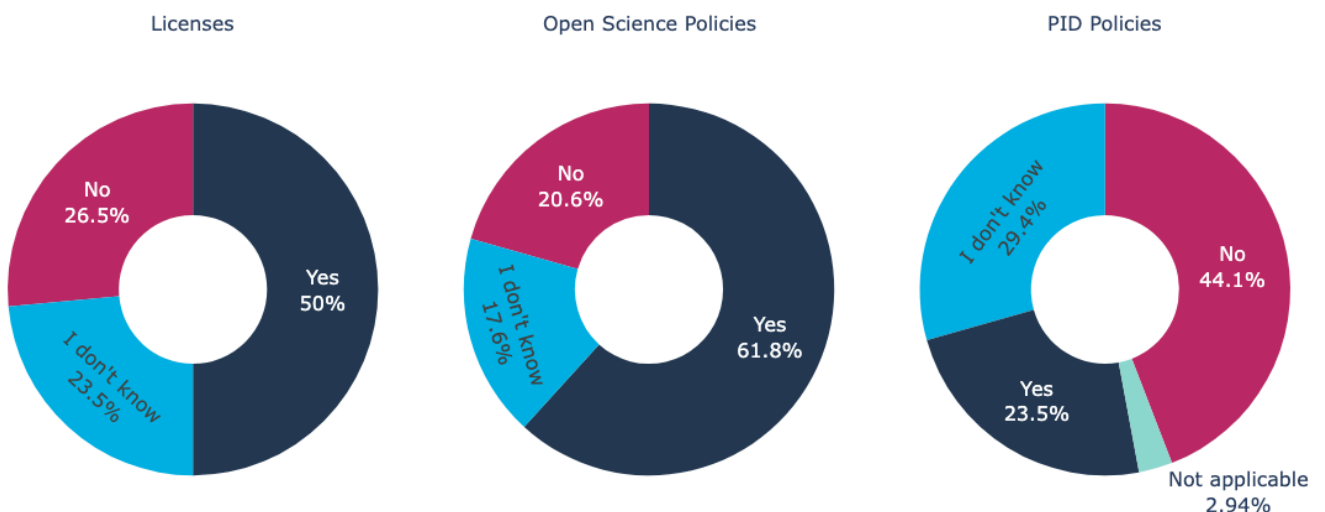


Figure 9 displays to which extent institutional policies or guidelines are utilized regarding the implementation of licenses, the adoption of Open Science practices, and the use of PIDs.

Further supporting this trend, a substantial 62% of respondents confirmed that their institutions have formal policies or guidelines in place specifically aimed at promoting and supporting Open Science initiatives.

However, despite the widespread adoption of licensing and Open Science policies, the implementation of PID policies remains inconsistent. Only 24% of respondents reported having formal PID policies, while 29% were unaware of such requirements within their institution. This suggests that PID policies are either lacking or insufficiently communicated, highlighting apparent gaps compared to the broader adoption of licensing and Open Science practices (**Figure 9**)

These results are also in line with responses where only 18% of respondents acknowledged that their institution has a formal governance structure for overseeing PID policies, while 56% reported lacking such mechanisms (**Figure 4**).

A qualitative analysis to better understand these obstacles could provide valuable insights and help promote the open availability of metadata for research resources.

### 3.4. Affiliation with PID Providers

Respondents were asked about their relationship with PID providers, focusing on membership status, affiliated providers, and the type of membership (for DataCite users). A majority (68%) of respondents indicated that their organization is a member of a PID provider, either directly or through a consortium. In contrast, 24% reported that they were not members, and 9% were unsure.

## Landscape Analysis of PID Practices in NFDI

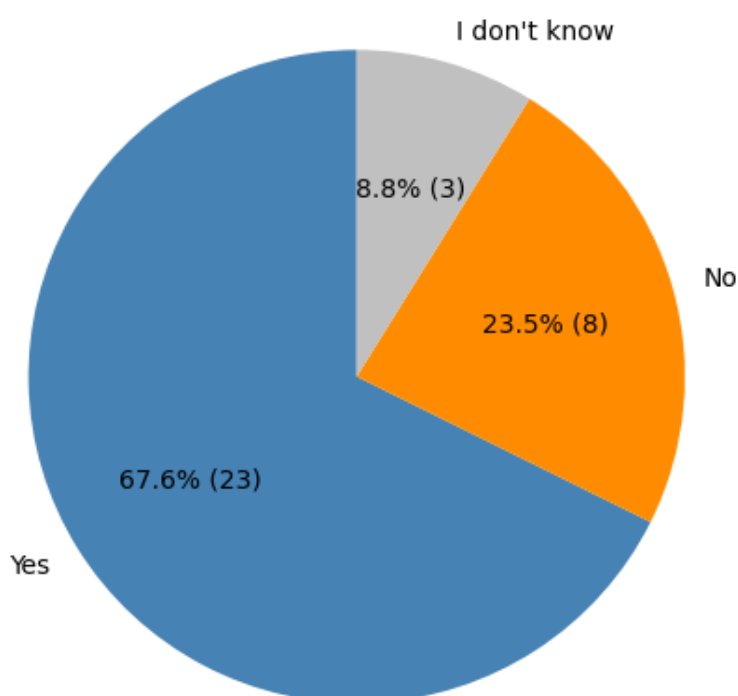


Figure 10 shows organizational Membership to PID Providers (Directly or via a Consortium, e.g., Crossref, DataCite, ePIC).

Those respondents who are members of PID providers were also asked which specific PID providers their organization is affiliated with.

Among those who reported memberships, DataCite emerged as the most commonly affiliated provider, cited by 57% of respondents. Other notable providers included ORCID (20%) and ePIC (10%). Under the “Other” category, providers such as da|ra, EGA, and GND were each mentioned by 3.3% of respondents.

The interpretation of “membership” differs across PID providers, complicating direct comparisons. For example, DataCite members are at the user level, while ePIC memberships are structured at the provider level. These differences influence how respondents perceive and report their affiliations.

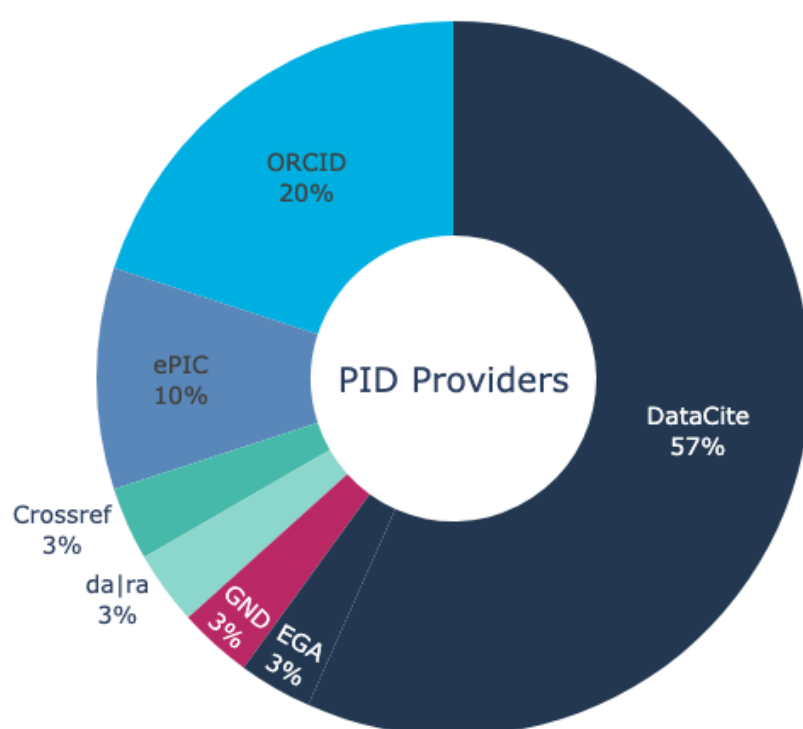


Figure 11 illustrates the distribution of PID provider affiliations. “Other” field included da|ra, GND and EGA.

Notable inconsistencies emerged between reported affiliations and actual practices. For instance, some respondents did not list DataCite as a provider, yet had repositories registered with DataCite. These discrepancies highlight the need for clearer definitions and guidance to improve the accuracy of such reporting (**Figure 12**).

## Landscape Analysis of PID Practices in NFDI

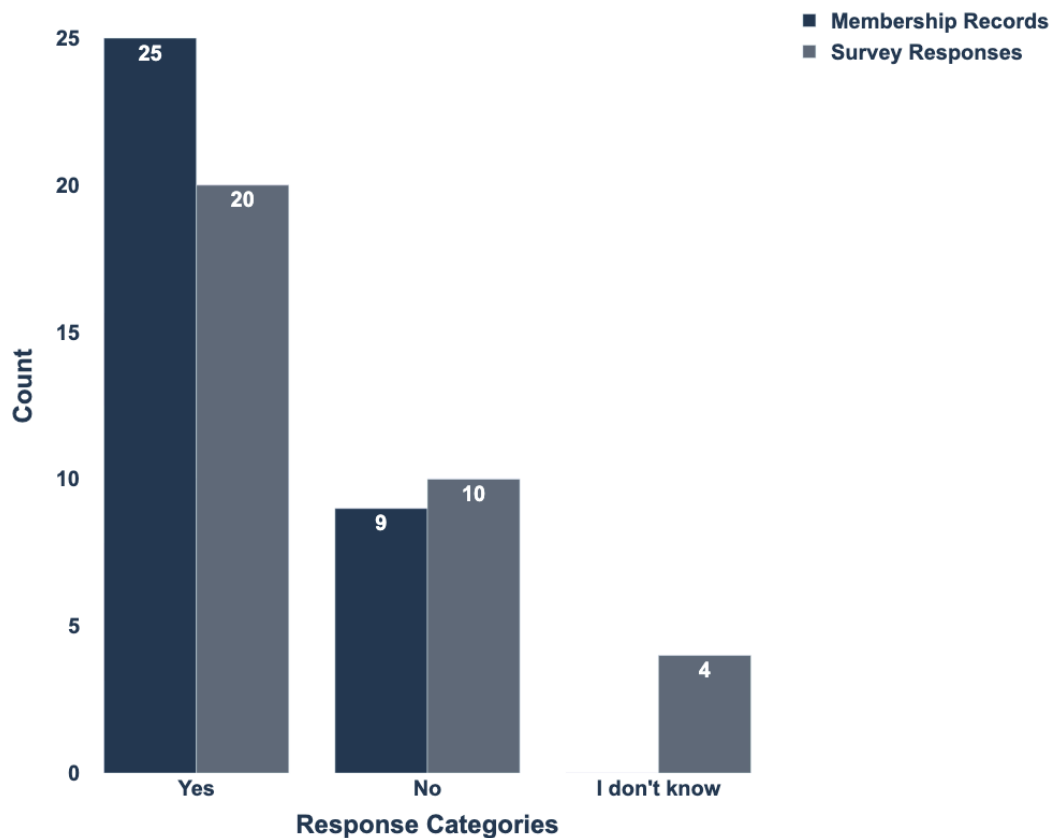


Figure 12 displays DataCite membership records vs survey responses.

### 3.5. Overview of Services and Infrastructures

The survey provided valuable insights into the range of services and infrastructures managed or utilized within NFDI. Respondents were invited to describe up to three services or infrastructures, revealing a broad spectrum of operational maturity and involvement across consortia.

Among the 34 respondents, participation varied in terms of service descriptions:

- 2 did not provide any service descriptions.
- 22 described one service.
- 8 described two services.
- 2 provided the maximum of three service descriptions.

After consolidating overlapping entries, where two services were described by multiple respondents, a total of 42 unique services were identified, offering a rich landscape of the consortia's operational ecosystem.

The infrastructures described span varying levels of maturity (**Figure 13**):

- 39% are already in mature operation, serving as established components of their respective infrastructures.
- 16% remain in the planning phase, reflecting the continuous evolution of services.
- 14% are currently in active development, while 16% are classified as prototypes, representing the innovation and experimentation underway.
- 11% have been recently released, signaling new contributions to the ecosystem.

Further details on the distribution of maturity phases across consortia are provided in the appendix **Figure S5a**.

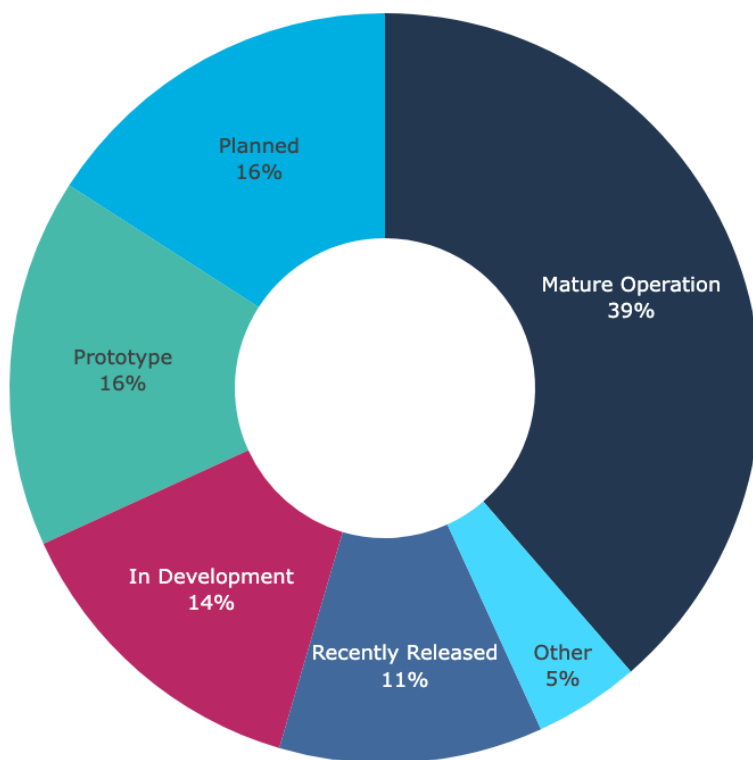


Figure 13 displays the current Operational Status of Infrastructures and Services.

### 3.6. Maturity Vs. PID Decision-making Roles

A clear connection emerges between infrastructure maturity and the roles involved in its management (roles illustrated in **Figure 5**). Mature operations demonstrate discernible engagement across all roles, with particularly strong associations with

technical, advisory, and strategic roles, reflecting their critical contribution to well-established systems (**Figure 14**). However, roles like Data Management and Curation as well as Research and Development are less prevalent in these contexts.

### 3.7. Strategic Decision-making Roles

Mature, prototype, in-development, and recently released infrastructures are closely aligned with roles in strategic decision-making, advisory functions, technical implementation, and operational management. Strategic staff play a critical role in these phases, particularly in the earlier stages of infrastructure development. Their frequent presence in both mature and in-development infrastructures suggests they often take on multiple responsibilities, including overseeing PID management. As infrastructures evolve, these responsibilities become more specialized, enabling technical staff to focus on specific tasks such as integration and implementation.

### 3.8. Advisory Roles

In contrast, planned infrastructures are predominantly linked to policy and advisory roles, reflecting the preparatory nature of these phases. Similarly, infrastructures with “other” operational statuses tend to correlate with advisory roles, underscoring the versatile function of these positions across various stages. Among all roles, advisory roles are particularly notable for their strong association with both mature infrastructures and prototypes, highlighting their integral contribution to infrastructure progression and refinement.

## Landscape Analysis of PID Practices in NFDI

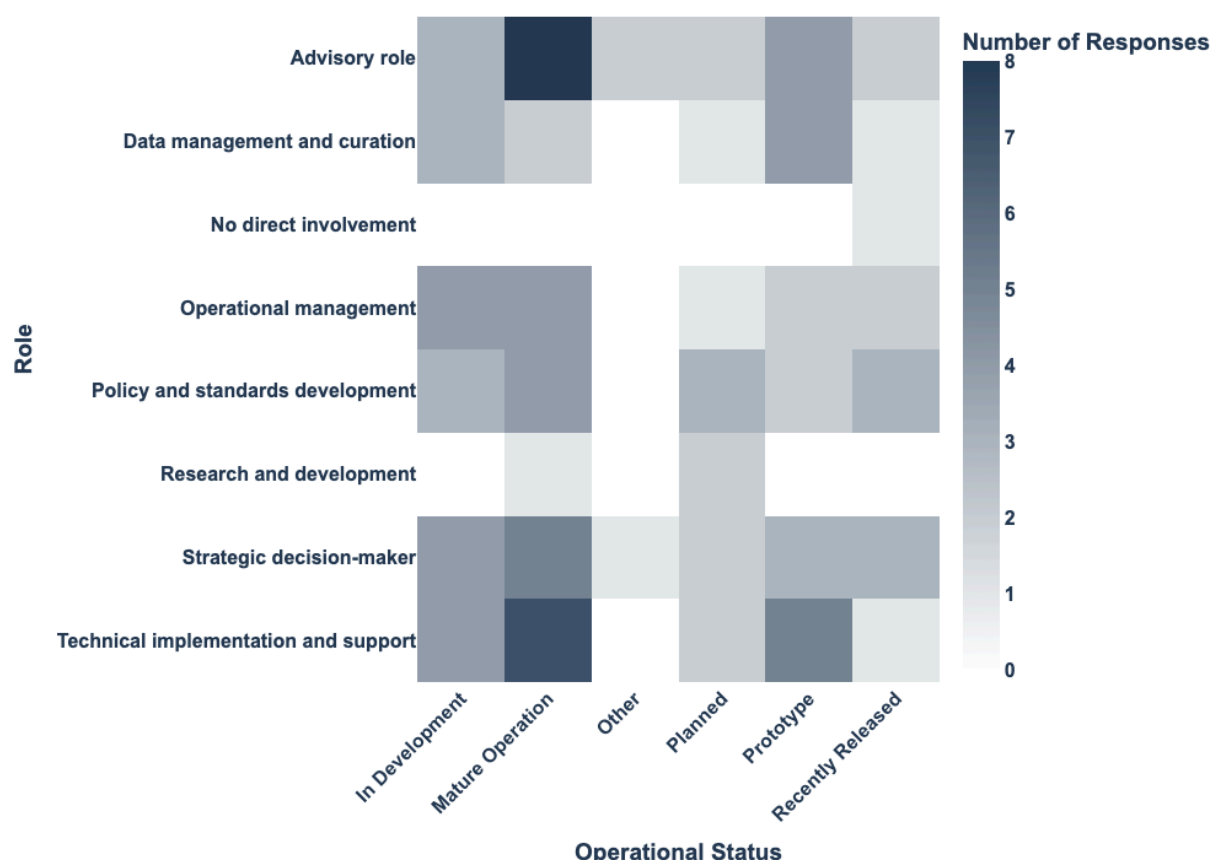


Figure 14 displays a heat-map highlighting distinct patterns between infrastructure maturity and the roles involved.

The operational status of an infrastructure introduces distinct challenges that vary depending on its level of maturity (**Figure 15**). Mature infrastructures face a comprehensive array of challenges including:

- **Organizational challenges**, such as ensuring PID persistence and resolution, managing project-based budgeting, and addressing insufficient staffing and resources.
- **Technical challenges**, such as integrating with existing systems, maintaining consistent PID resolution, and adapting to evolving standards and best practices.

Interestingly, some challenges, such as insufficient resources, the absence of a dedicated budget, and reliance on project-based funding, persist across all levels of maturity. These issues highlight systemic barriers that affect infrastructures regardless of their operational status.



On a technical level, integrating with existing systems and ensuring PID persistence are recurring pain points across all stages of development, particularly in mature and in-development infrastructures. This emphasizes the need for sustained support to address these foundational challenges.

Collectively (**Figure 14 and Figure 15**) the interplay between operational maturity, roles and challenges highlights the need for accessible and robust PID management, particularly as organizations look to scale their PID registration activities.

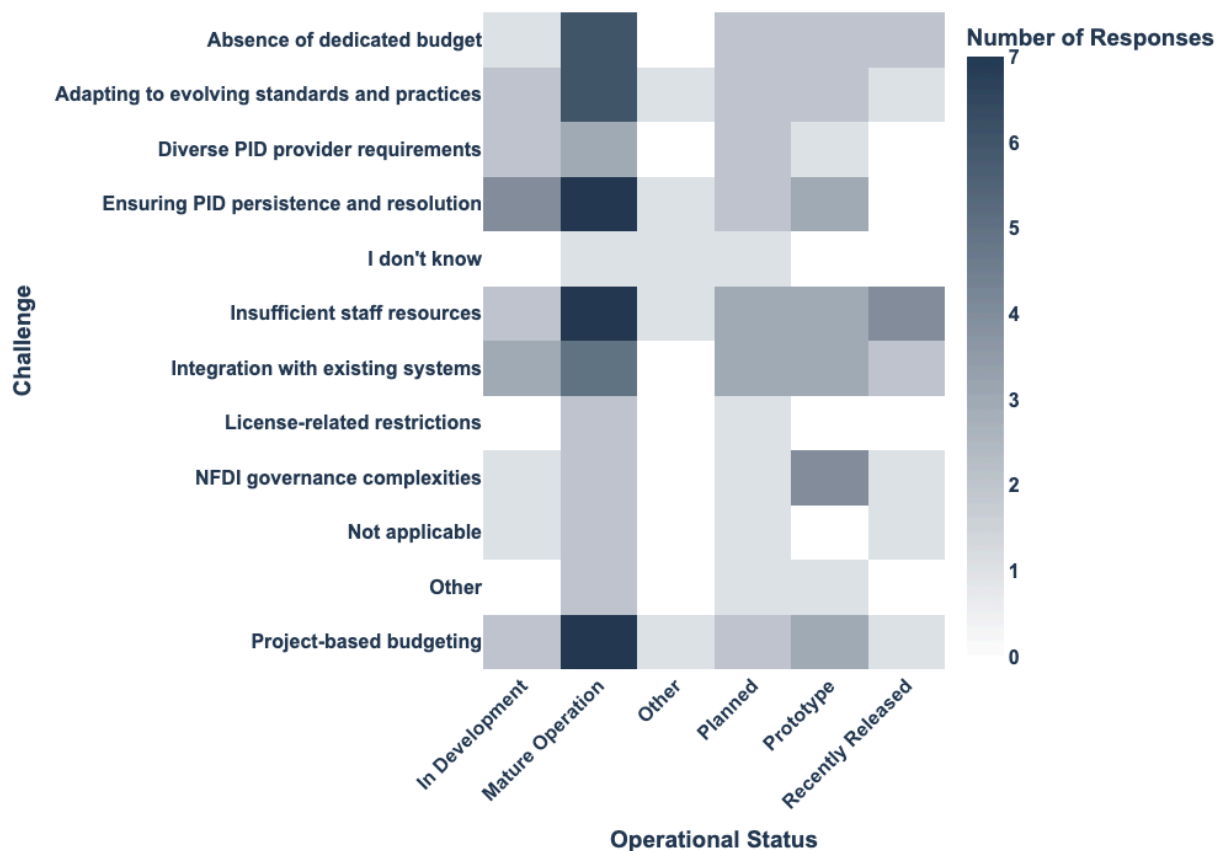


Figure 15 displays a heat-map highlighting distinct patterns between challenges versus operational status.

The PID Coordination Hub is uniquely positioned to address these challenges by providing targeted support frameworks, thereby enabling infrastructures to scale their PID registration activities effectively and sustainably.

These frameworks can help infrastructures bridge gaps in staffing and training while equipping them with the necessary tools and expertise to meet evolving standards and best practices.

### 3.9. Infrastructure Types and Purpose

The variety in infrastructures is considerable, with many organizations relying on cloud-based platforms such as Amazon's Simple Storage Service (S3), OpenStack, and the Open Telekom Cloud, indicating a mix of global and European-based solutions. Open source solutions are also popular, with multiple references to Dataverse, Fedora, Coscine, and ICAT. Tools like RDMO and DaCHS are commonly used, emphasizing the community's preference for customizable, open source, community-driven software.

Furthermore, there is a noticeable share of infrastructures based on custom-built or in-house developed solutions, reflecting the diverse and tailored approaches to infrastructure within the community (**Figure 16**).

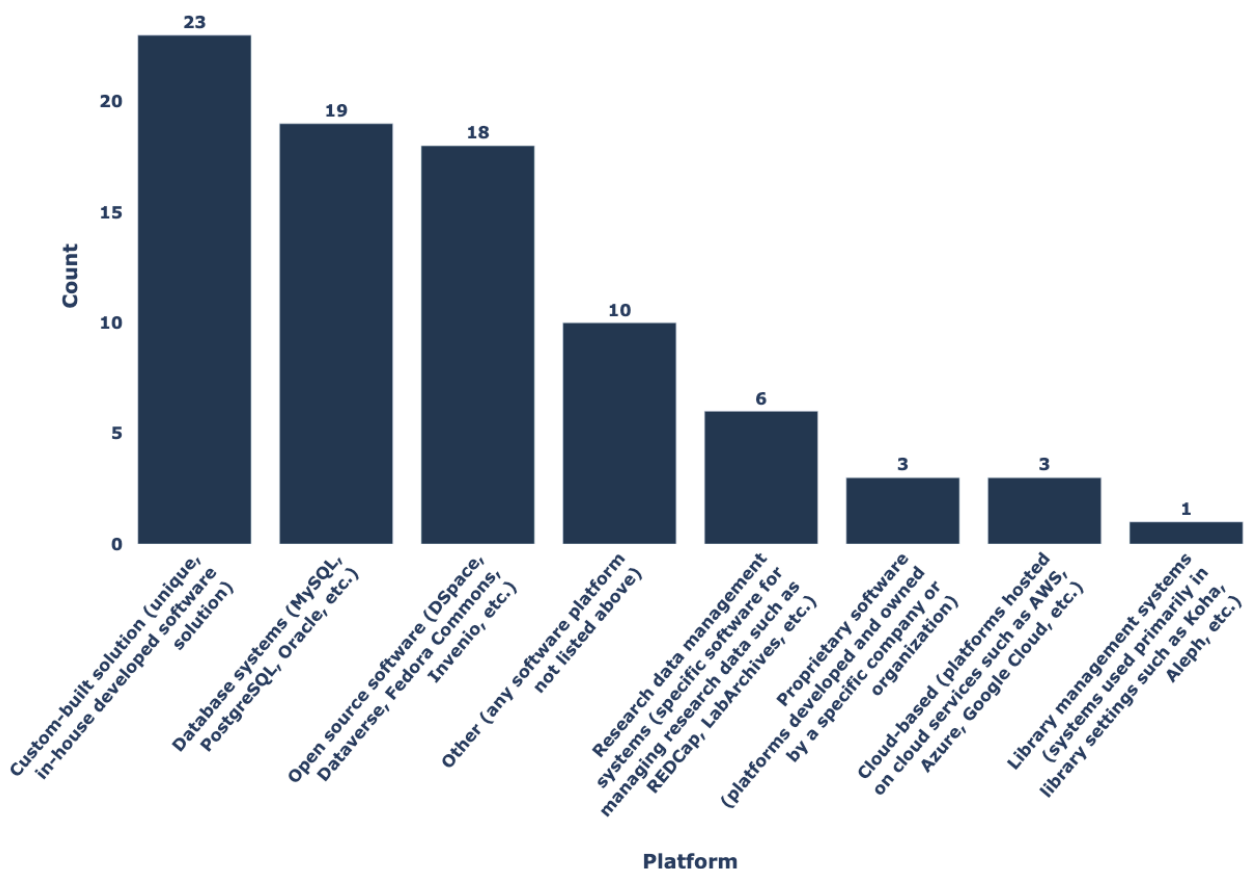


Figure 16 displays responses detailing the types of described infrastructure, specifically the distribution of software platforms, based on a multiple-response question.

Despite this diversity, the underlying purpose behind the development of infrastructures remains consistent across organizations. This alignment in purpose

indicates that, regardless of the technological diversity, the adoption of PIDs serves as a unifying framework for achieving these goals (**Figure 17**).

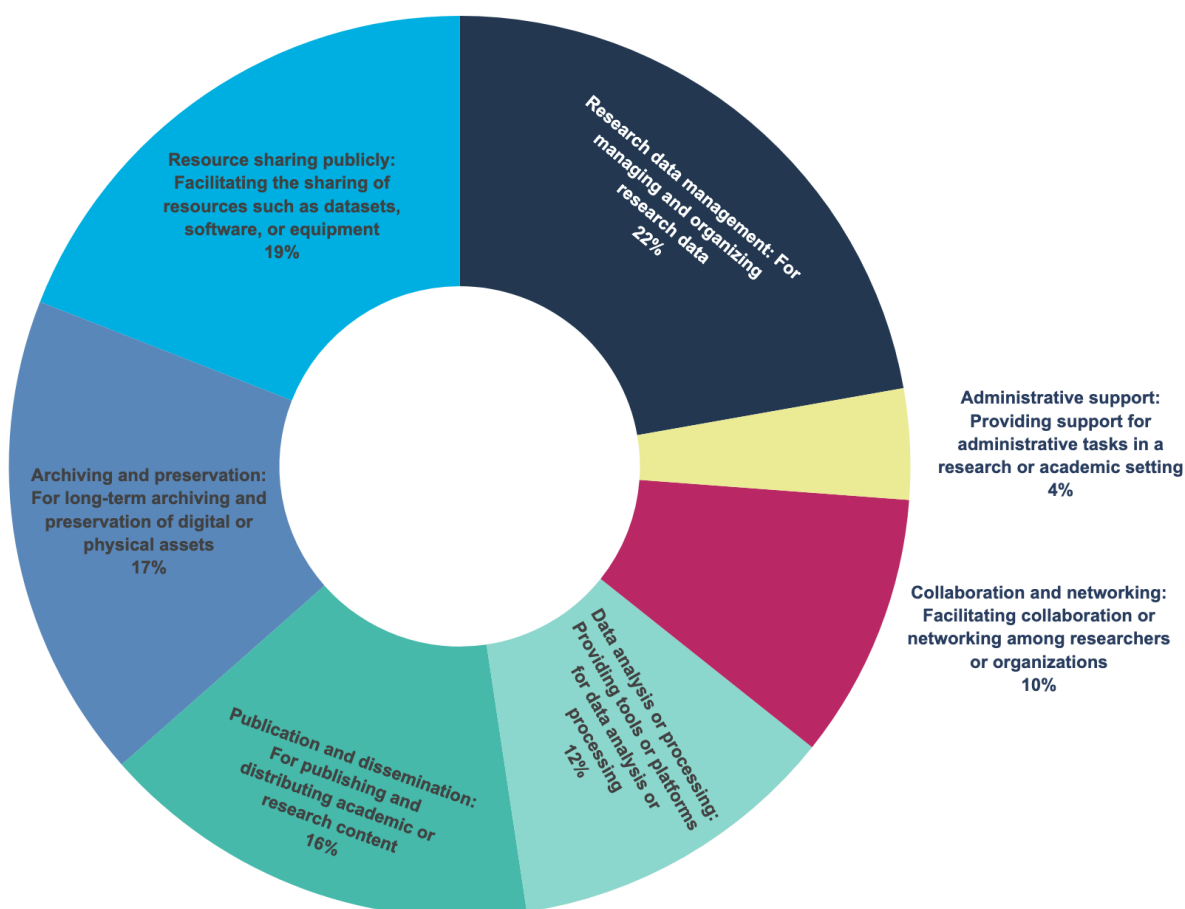


Figure 17 displaying the reported primary purposes of infrastructures and services.

The main purposes of infrastructures and services within the surveyed organizations reveal a strong focus on supporting key activities:

- **Research Data Management (22%):** By far the most cited purpose, highlighting the critical role of infrastructures in organizing and managing research data.
- **Resource Sharing (19%):** Facilitating the open exchange of datasets, software, and other resources to promote collaboration and innovation.
- **Archiving and Preservation (17.5%):** Ensuring the long-term safeguarding of digital and physical assets, a foundational aspect of sustainable data practices.

- **Publication and Dissemination (15.9%):** Supporting the publishing and distribution of academic content, underscoring the importance of accessible and shareable outputs.

Other purposes, including administrative support, collaboration and networking, and data analysis or processing, were also represented.

The shared challenges of PID management underscore the need for a unified, top-down strategy. Documenting best practices and lessons learned from mature infrastructures can help consortia replicate successful models while avoiding common pitfalls, ensuring all NFDI consortia can advance towards a higher level of maturity in PID management.

### 3.10. Interoperability and Metadata Alignment Needs

The survey responses highlighted several technical challenges and opportunities for improvement in interoperability and metadata alignment. A recurring theme was the need for enhanced alignment with metadata standards, particularly the Data Catalog Vocabulary (DCAT) in compliance with W3C guidelines. Respondents frequently emphasized that improving DataCite's alignment with DCAT would facilitate more seamless integration with other tools and services, such as Electronic Lab Notebooks (ELNs) and Data Management Plans (DMPs). This alignment would also make resources more directly accessible and interoperable across platforms.

The widespread adoption of the DataCite Metadata Schema (**Figure 18**) underscores its central role in enabling PID registration and fostering interoperability across repositories and services. Its popularity presents a unique opportunity to position it as a unifying framework for metadata alignment within the NFDI community. By improving its compliance with standards such as DCAT, the DataCite schema could become a more powerful tool for streamlining integrations and reducing redundancies across platforms.

## Landscape Analysis of PID Practices in NFDI

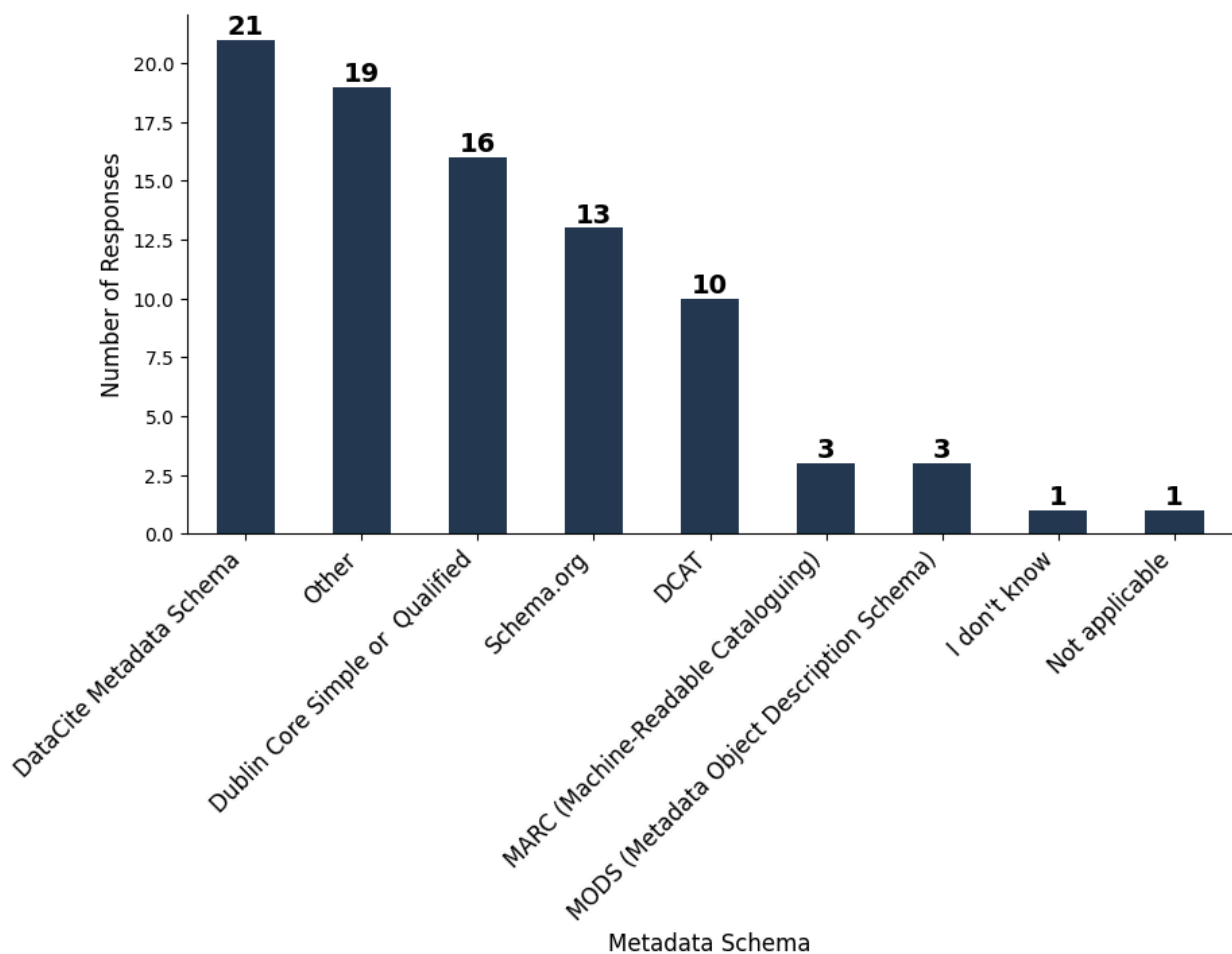


Figure 18 displays the frequency and types of metadata schemas currently used in infrastructure services for PID registration.

Additionally, there was a strong call for consolidating metadata standards across PID providers and data repositories to streamline interoperability efforts. A unified approach, centered around the harmonization of DataCite Metadata Schema with other standards, would significantly ease the workload for institutions that interact with multiple platforms. Respondents highlighted specific needs, such as:

- Improved mappings between metadata schemas like Schema.org and DCAT.
- Integration of RDF to enable the creation of knowledge graphs and triple stores, enhancing the semantic richness and connectivity of data.

These improvements would promote better interoperability between repositories and data services, ultimately enhancing the discoverability and usability of resources.

Additional free-text responses also pointed to the widespread use of custom-built solutions for metadata management, signaling a lack of standardized frameworks widely adopted across the community. This fragmentation creates additional challenges for interoperability and alignment.

A common suggestion was the centralization of metadata schema registration, which would enhance visibility for mappings, crosswalks, and alignments. To address this need, the PID Coordination Hub could play a pivotal role by:

- **Establishing a Data Type Registry (DTR):** A centralized tool for registering and managing metadata schemas, improving accessibility and interoperability.
- **Facilitating a Metadata Schema and Crosswalk Registry (MSCR):** A repository for mapping and extending existing schemas, making it easier for organizations to navigate metadata standards and align their systems.

These initiatives would support institutions in managing their metadata more efficiently, reducing fragmentation and fostering a more cohesive approach to interoperability.

### 3.10.1. Granular Metadata and Controlled Vocabulary

The need for more granular metadata and controlled vocabularies emerged as a key priority among respondents. Many emphasized the need for metadata at lower granularity levels to document attributes such as elements within variables, audiovisual data, and text.

Addressing these requirements would involve updating existing standards and guidelines to accommodate a broader spectrum of resource types. Several respondents highlighted the importance of developing more granular metadata systems, including sub-units within the Research Organization Registry (ROR) or utilizing community-driven contributions like Wikidata IDs for defining more specific entities.

As **Figure 19** illustrates, DataCite metadata is also a popular method for establishing and recording links between research outputs, emphasizing its role as a foundational tool for resource connectivity. Leveraging this popularity to develop

more granular metadata systems could address specific discipline-based needs while maintaining a unified framework.

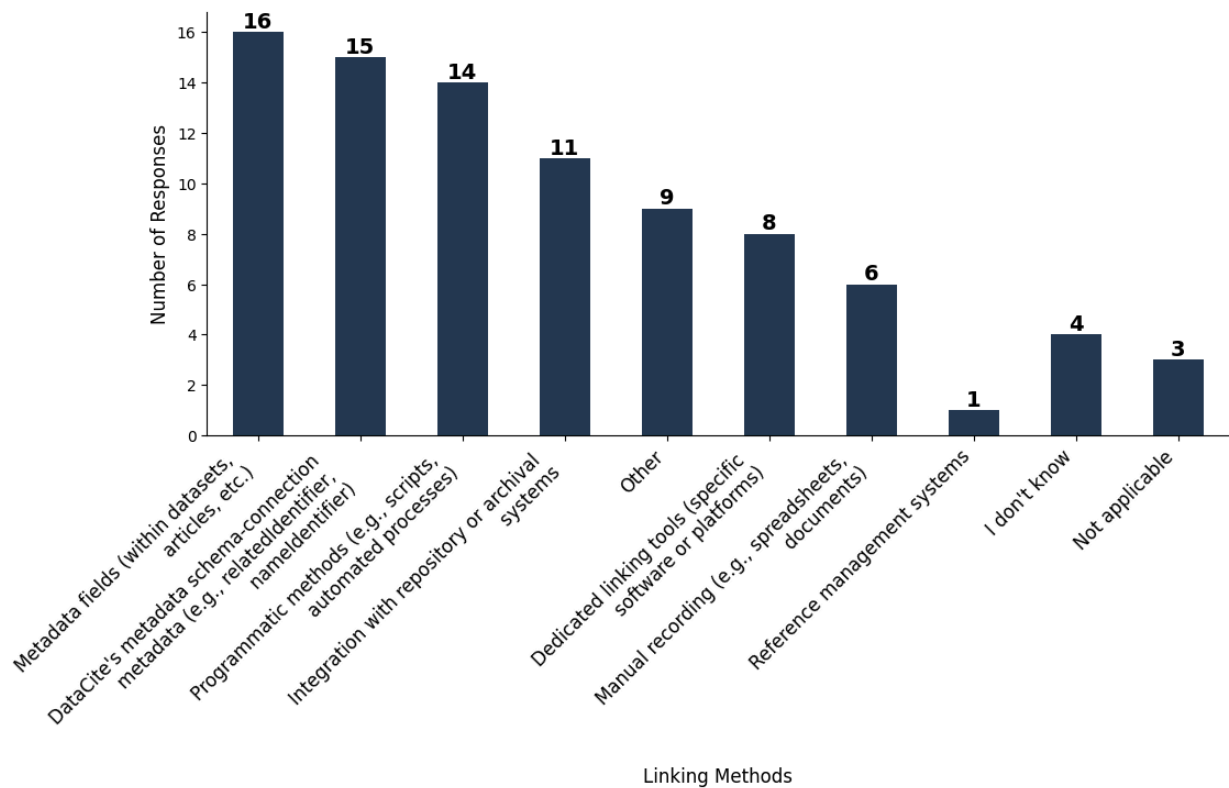


Figure 19 displays the methods used to establish and record links between various research outputs.

This approach addresses the need to integrate discipline-specific metadata into the broader DataCite schema, a recurring theme among respondents managing both generic and specialized repositories. Linking specialized metadata through related identifiers in DataCite was highlighted as a way to standardize workflows and streamline processes, benefiting all NFDI participants.

Developing a general NFDI-wide strategy for metadata granularity and controlled vocabulary integration would also address key challenges related to metadata alignment. Establishing guidelines and best practices for linking discipline-specific metadata with DataCite would ensure a consistent, scalable approach, further enhancing resource discoverability and usability across repositories on a global scale.

### 3.10.2. Interoperability Through APIs and Integration

The majority of respondents (56%) indicated plans to enhance interoperability in the near future (**Figure 20**). These efforts focus on areas such as API integration, ORCID integration, and exposing data as FAIR Digital Objects.

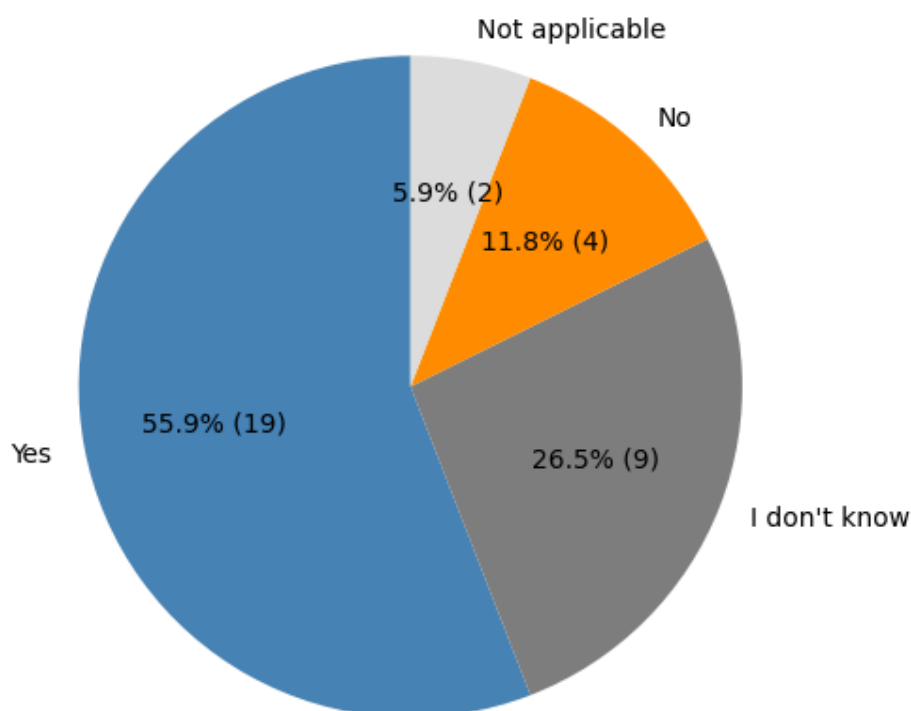


Figure 20 shows upcoming plans for enhancing technical interoperability

Survey responses underscored the increasing reliance on APIs to improve interoperability and metadata alignment across repositories (**Figure 21**). Many institutions are actively developing and integrating APIs into their workflows, utilizing both DataCite APIs and repository-specific APIs to establish seamless connections between systems. Additionally, some respondents highlighted efforts to implement FAIR Data Points and FAIR Digital Objects, aiming to elevate data management practices and align more closely with the principles of FAIR data.



## Landscape Analysis of PID Practices in NFDI

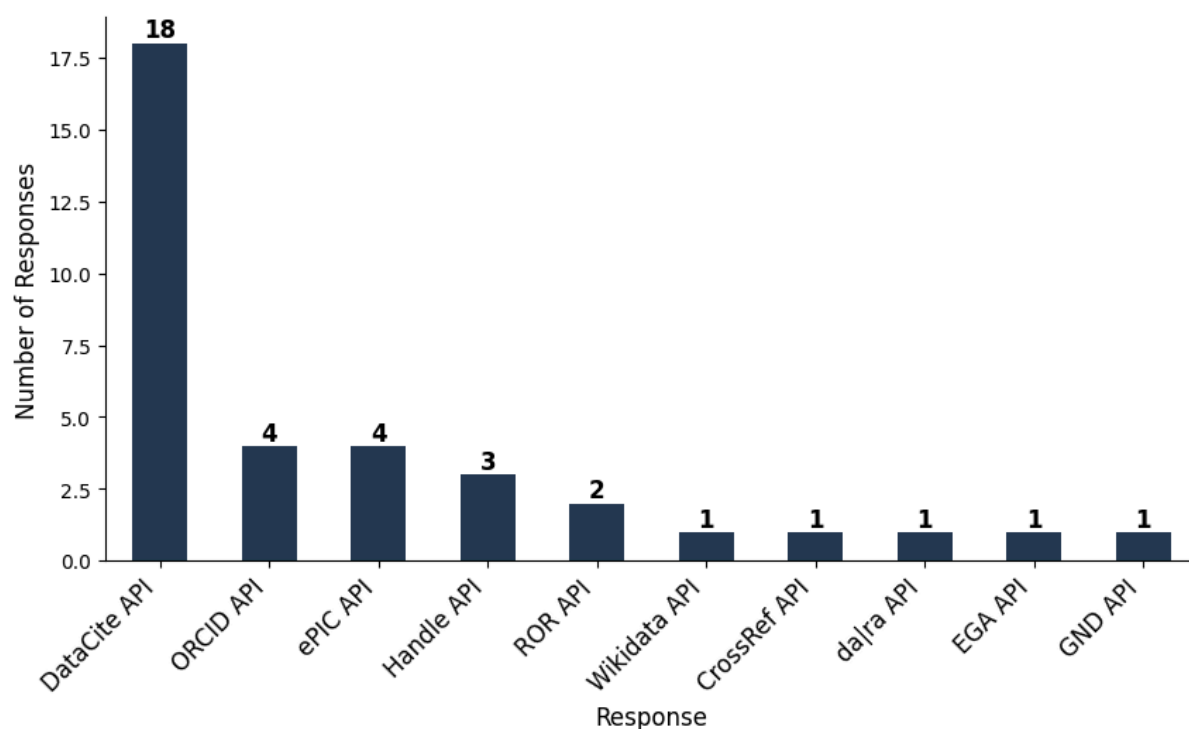


Figure 21 illustrates the use of APIs for interacting with PID services. Figure S13 in the appendix provides a detailed overview of which consortia utilize specific APIs.

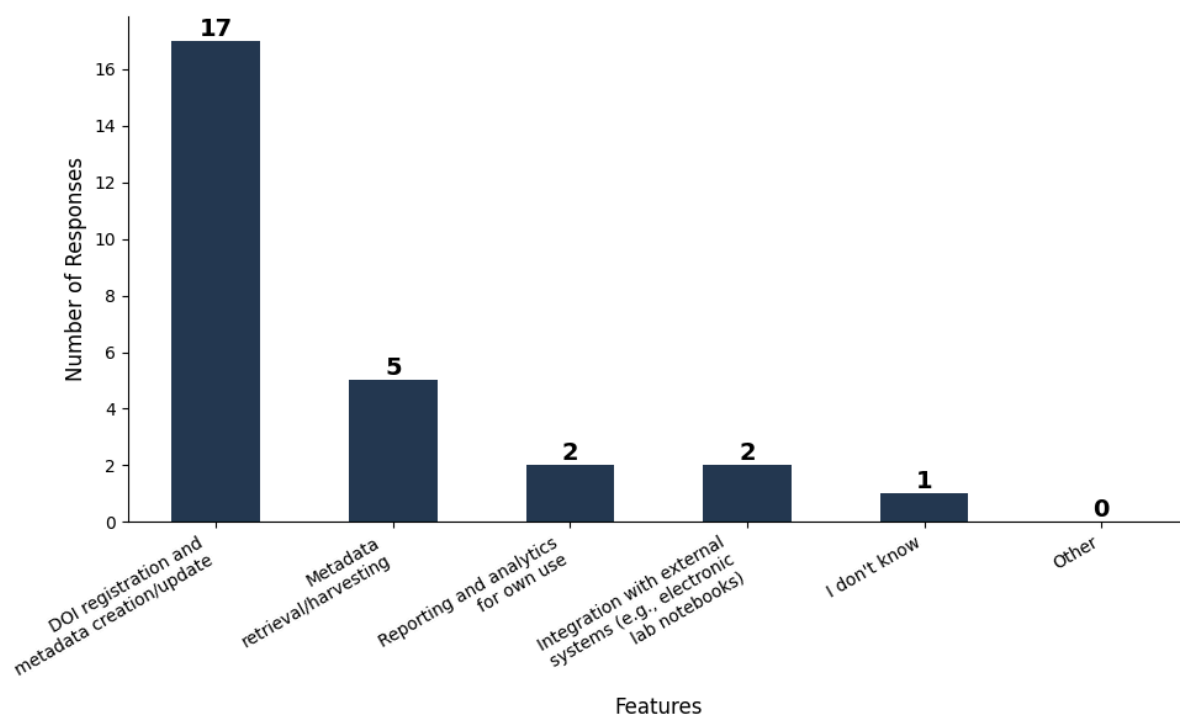
Additional responses highlighted a strong emphasis on integrating ORCID within Single Sign-On (SSO) systems. This integration simplifies user authentication processes while enabling seamless data sharing and interoperability across platforms.

These initiatives highlight the strategic importance of APIs and identity systems in addressing technical challenges and fostering greater collaboration within the research data ecosystem.

The survey results, depicted in **Figure 22a**, highlight the widespread adoption of APIs for interacting with PID services, including tasks such as creating, updating, and retrieving PID data. Among these, the DataCite API emerged as the most widely used, as shown in **Figure 22b**. This popularity is likely tied to the extensive membership and adoption of DataCite among survey respondents (**Figure 12**).

## Landscape Analysis of PID Practices in NFDI

a)



b)

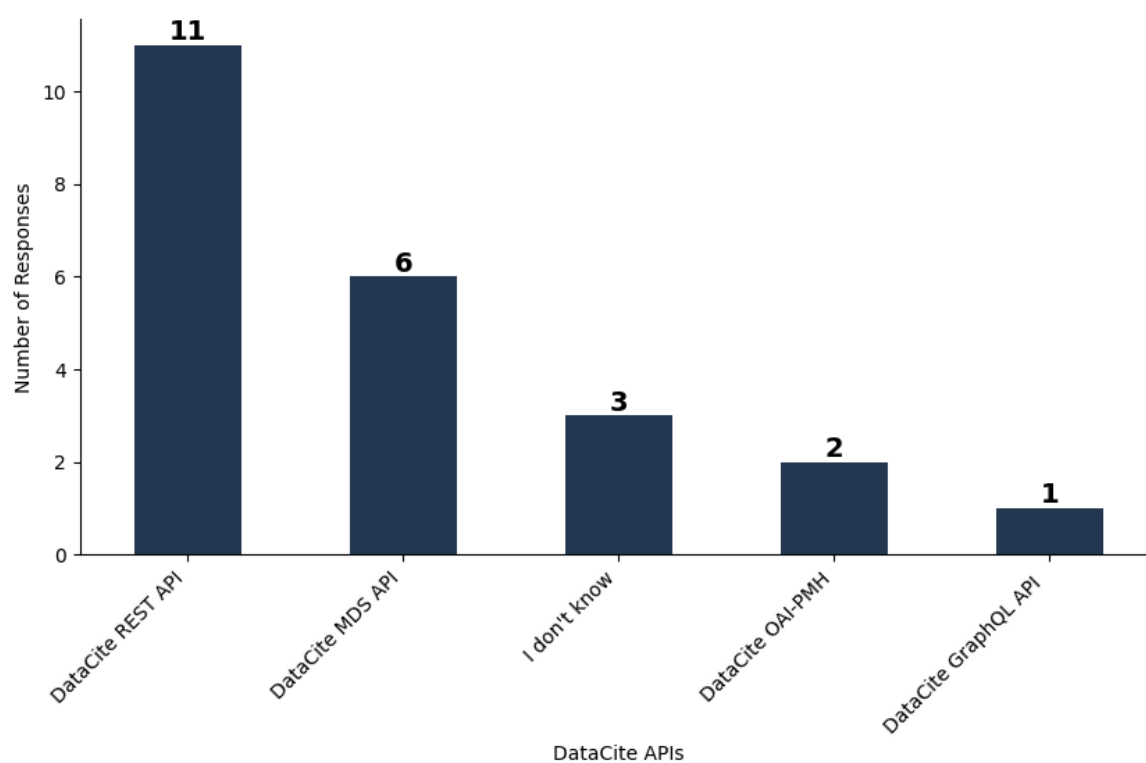


Figure 22 displays a) most frequently used features of DataCite APIs. b) usage of DataCite APIs.

The growing popularity of APIs, particularly those offered by DataCite, aligns with another key trend identified in the survey: an increasing interest in training on API development and utilization. This reflects a growing recognition among institutions of the critical role APIs play in enhancing interoperability, streamlining workflows, and driving innovation in PID management practices.

Enhancing API expertise will empower organizations to fully leverage APIs capabilities, enabling improved metadata alignment, reducing administrative burdens, and driving innovation in PID management practices.

To support this growing demand, the PID Coordination Hub is positioned to facilitate comprehensive training and resources tailored to API use in PID management. By providing hands-on workshops, documentation, and collection of integration tools, the PID Coordination Hub aims to bridge the gap between technical capabilities and operational needs to strengthen interoperability across platforms.

### 3.11. Training and Outreach

The quality of metadata for **PIDs** plays a pivotal role in effective research data management. However, achieving high-quality metadata heavily depends on researchers' understanding of its importance.

### 3.12. Engagement with PID Training

Survey responses reveal limited and uneven engagement with PID training efforts across the community (**Figure 23**). Over half of the respondents (19 out of 34) reported not having attended any PID training courses in the past two years, indicating a gap in awareness and skill development.

Among those who participated in training, DataCite emerged as the most prominent provider, with 13 out of 34 respondents attending its courses. This reflects DataCite's central role within the NFDI landscape. In contrast, training from other providers, including ORCID, the Handle System, ePIC, GND, ROR, URNs, and Wikidata ID, saw considerably lower participation rates.

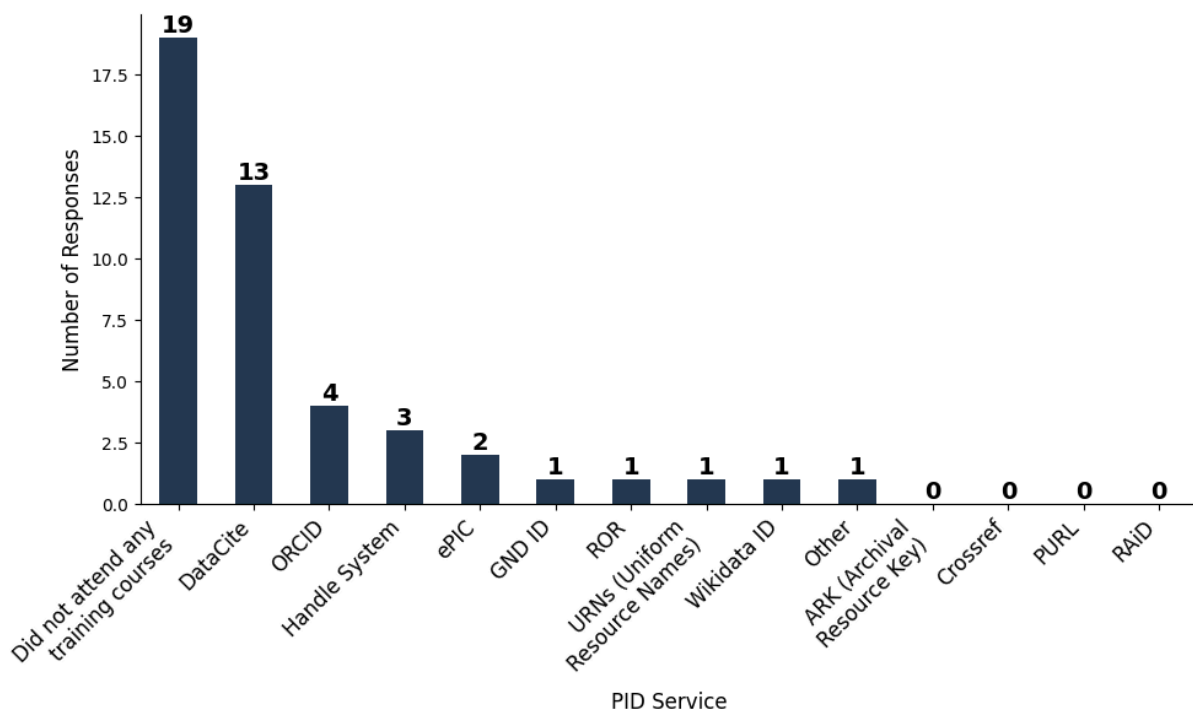


Figure 23 displays engagement with PID Training efforts.

### 3.13. Barriers to Training Engagement

The survey did not explicitly identify the reasons behind low participation in PID training, but several potential factors can be inferred:

- **Limited Availability of Training Programs:** A lack of accessible, well-publicized courses may reduce participation.
- **Resource Constraints:** Time and budget limitations could prevent individuals from attending training sessions.
- **Perceived Irrelevance:** Some respondents may view PID training as unnecessary for their specific roles or tasks.

These barriers align with broader outreach findings, where respondents highlighted clear expectations for support (**Figure 24**). Notably, half of the respondents identified training as an essential area, emphasizing a critical need for targeted intervention.

Bridging these gaps is vital to increasing training participation and, in turn, enhancing the overall quality of metadata management.

The PID Coordination Hub is uniquely positioned to bridge these gaps by:

- Expanding training opportunities to ensure wide accessibility.
- Customizing content to align with diverse roles and institutional needs.
- Promoting the importance of metadata quality through outreach and advocacy efforts.

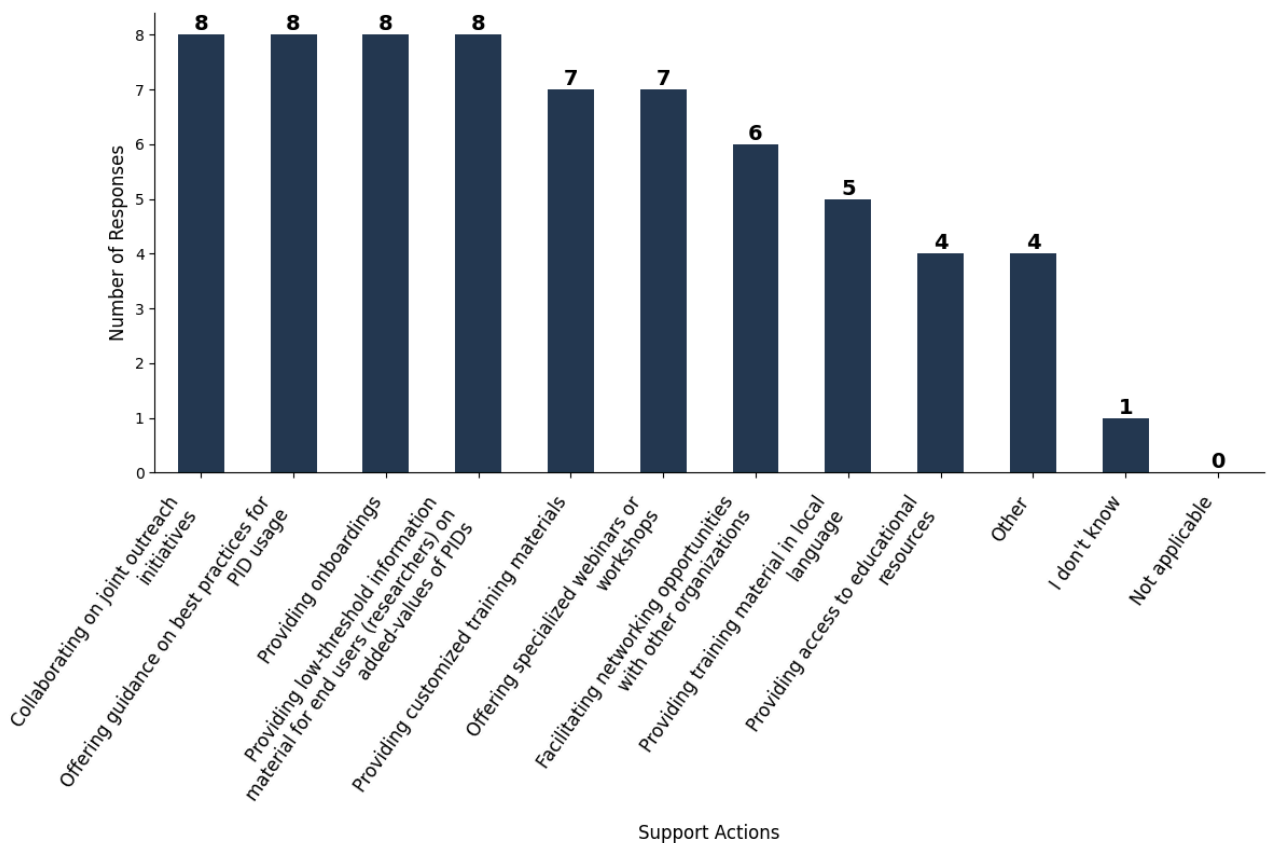


Figure 24 displays support expectations from PID Providers for organizational training and outreach efforts.

### 3.14. Training Formats and Preferences

Webinars emerged as the most popular training format among respondents (**Figure 25**). Their appeal likely stems from the interactive and lively nature of these sessions, which provide opportunities for direct engagement and tailored support through Q&A segments. Following webinars, support documentation was the second most preferred format, reflecting respondents' appreciation for asynchronous learning options that allow them to learn at their own pace and according to their schedules.

This preference for a combination of synchronous (webinars) and asynchronous (documentation) training formats highlights the importance of offering diverse learning opportunities. A flexible approach that includes both live interaction and self-paced resources ensures that PID-related training is accessible, adaptable, and effective in meeting the varied needs.

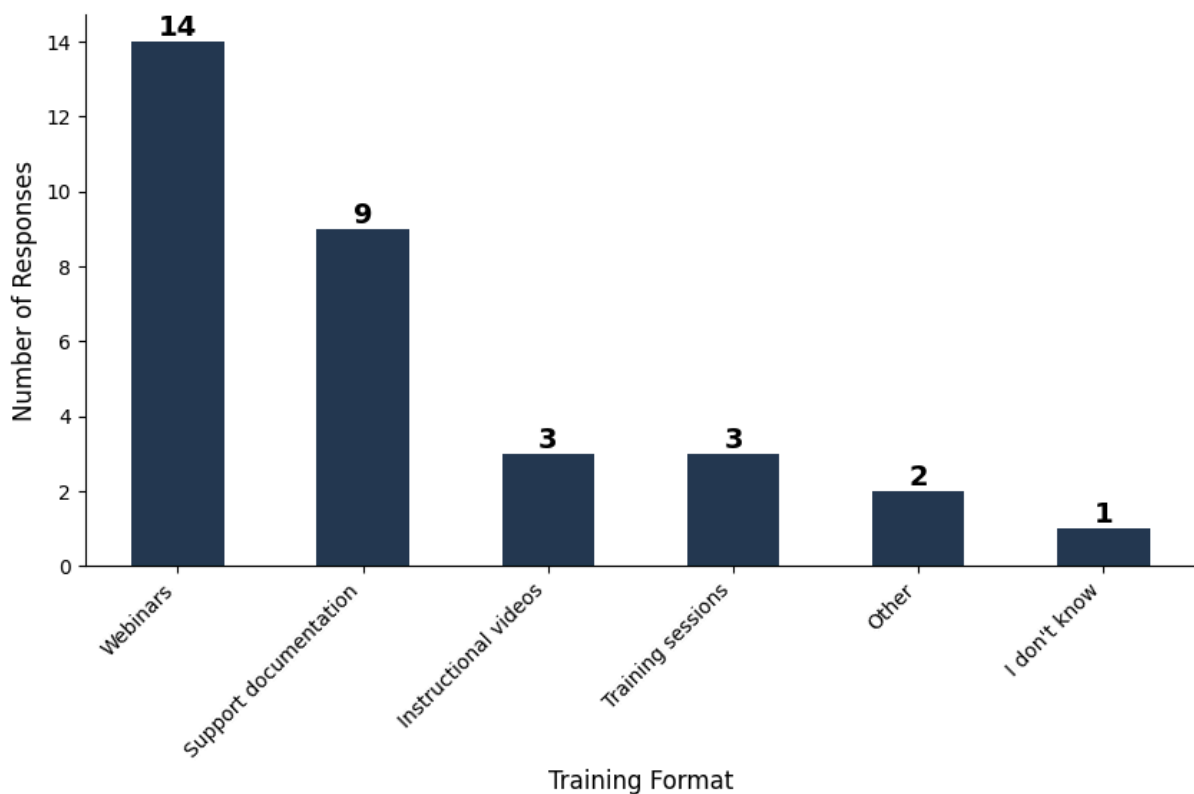


Figure 25 illustrates respondents preferred training formats. Figure S15 a) in the appendix provides an overview of attended formats stratified by PID providers.

### 3.15. Demand for Specific Training Topics

Respondents expressed prominent interest in several specific training topics (**Figure 26 and Figure 27**). Metadata quality and metadata standards emerged as the most sought-after areas, reflecting a clear need to address knowledge gaps that directly impact the comprehensiveness and reliability of metadata.

There was also notable interest in foundational PID concepts, underscoring the need for training that caters to a range of expertise levels. This demand supports the ongoing efforts of the PID Coordination Hub to provide both beginner level and advanced training content, ensuring inclusivity and progression for all participants.

## Landscape Analysis of PID Practices in NFDI

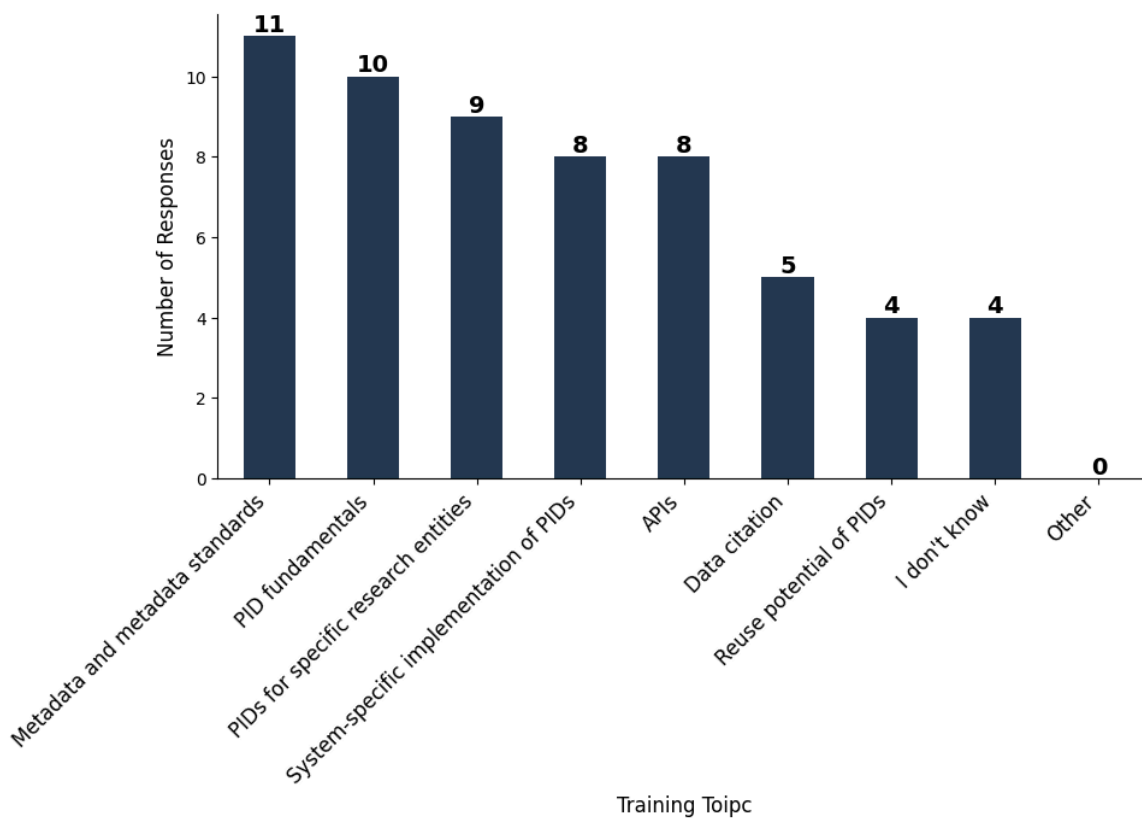


Figure 26 displays the estimated attendance for various training topics. Figure S15 b) in the appendix provides an overview of attended topics stratified by PID providers.

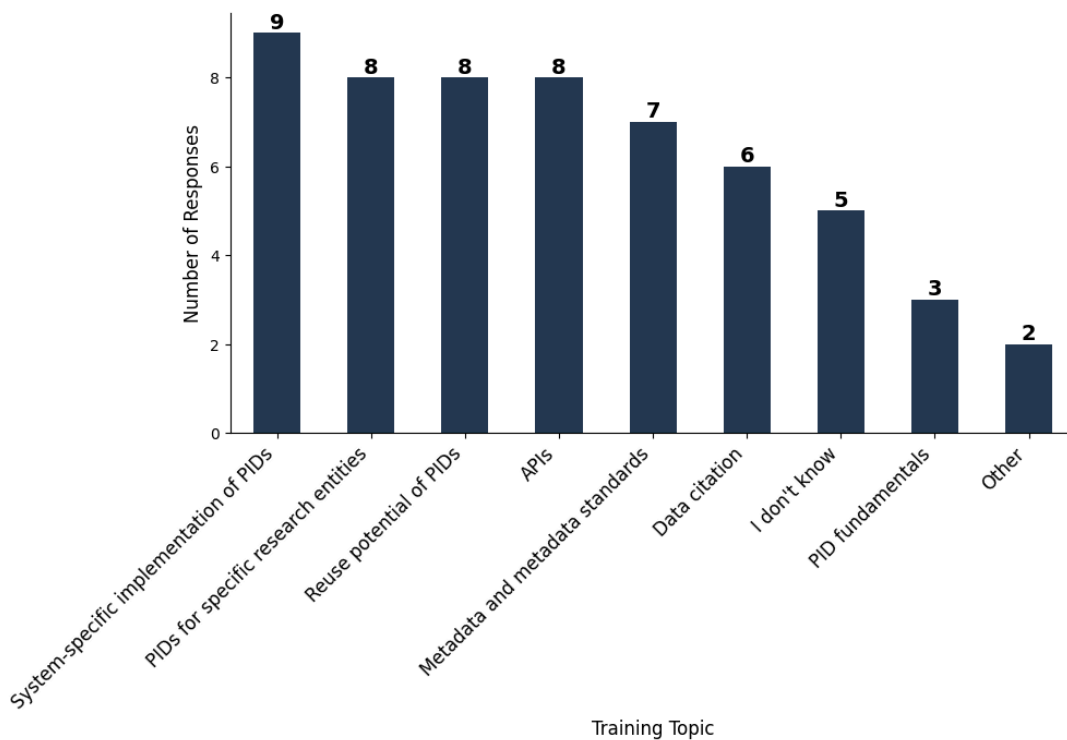


Figure 27 lists the training topics in demand. Figure S15 c) in the appendix provides a detailed overview of desired topics stratified by PID providers.

The survey revealed a particularly high interest in DataCite-related training, compared to other PID providers. This preference may reflect the respondent group's specific background and DataCite membership popularity.

Given their experience, respondents expressed a stronger need for advanced training across all areas, with slightly less emphasis on fundamentals. This aligns with the assumption that the target audience already possesses a baseline familiarity with PID concepts, seeking to refine their expertise. Offering specialized and in-depth training opportunities would address this demand, helping participants enhance their technical and operational capabilities in PID management.

### 3.16. Organizational Support for Metadata Completeness

The survey revealed evident gaps in organizational support for metadata completeness (**Figure 28**). More than half of the respondents either did not know whether their organization provided training on metadata completeness or reported that no such training existed. Less than half indicated that such training was available, either internally or through a combination of internal and external resources.

This uneven provision of training highlights a critical need for a centralized and consistent resource. The PID Coordination Hub is well-positioned to address this gap by offering standardized training programs, ensuring that all consortia members are equally equipped to achieve metadata completeness.

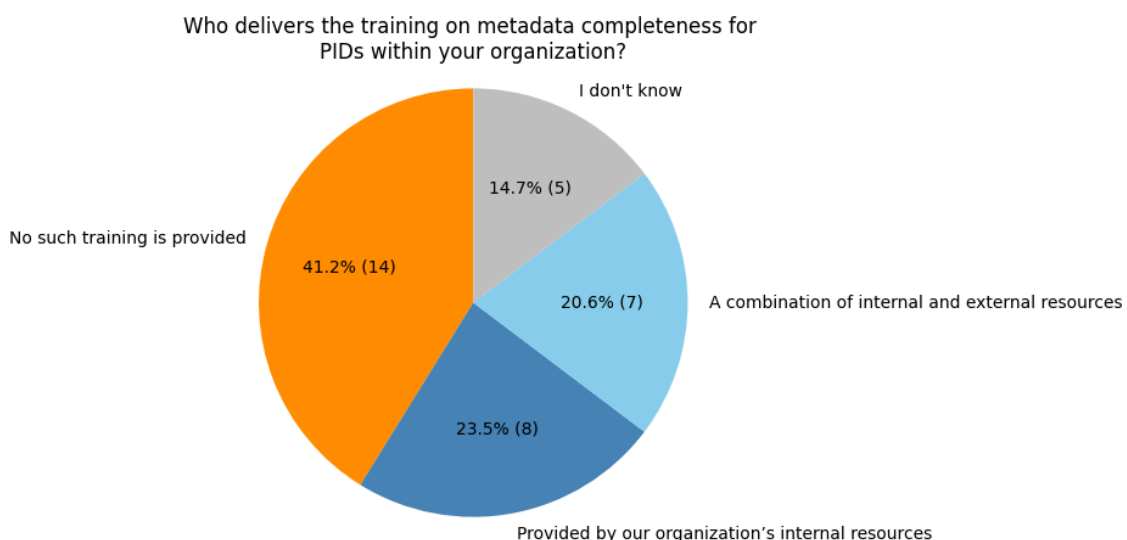


Figure 28 describes if metadata completeness training is delivered and by which entity.



### 3.17. Outreach Activities for PID Awareness

The survey also investigated outreach activities aimed at promoting PID usage and education. Half of the respondents indicated that their organizations conduct outreach efforts (**Figure 29**). Specifically, respondents from 14 organizations (representing 14 NFDI consortia) confirmed the existence of such activities. In contrast, respondents from 8 organizations (representing 10 NFDI consortia) reported no outreach activities, while respondents from 7 organizations (representing 5 NFDI consortia) were uncertain.

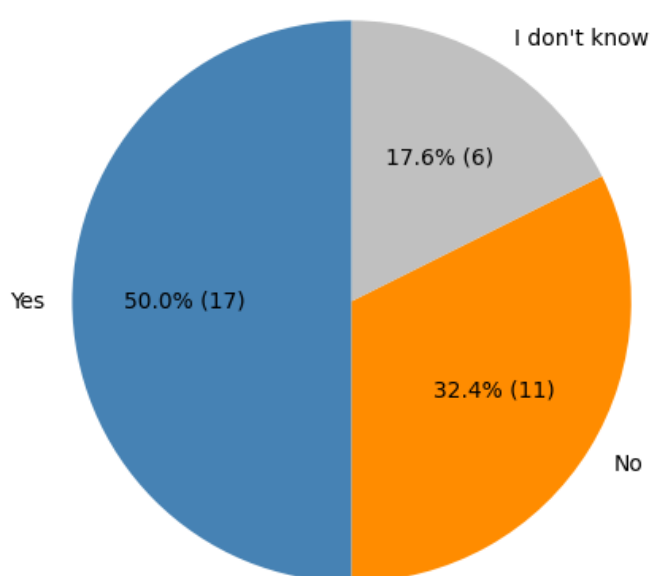


Figure 29 illustrates the percentages of undertaken outreach activities for PID awareness.

Among the 17 respondents who reported active outreach efforts, researchers (15 out of 17) and data curators or librarians were the primary audiences targeted (**Figure 30**). Outreach was less frequently directed toward administrators, IT professionals, policymakers, or the general public.

## Landscape Analysis of PID Practices in NFDI

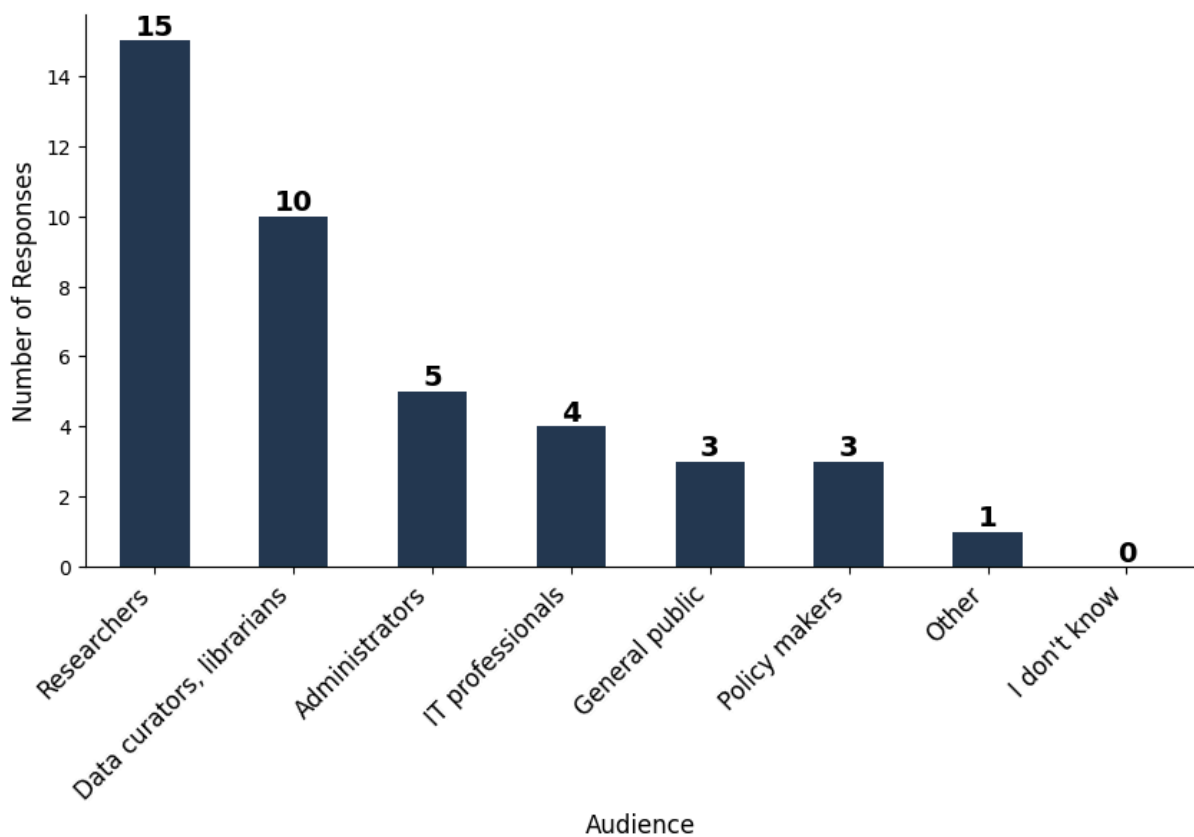


Figure 30 illustrates the primary target audience for outreach activities on the topic of PIDs.

Outreach activities commonly take the form of workshops and conferences, with formats evenly split across in-person, online, and hybrid events. Communication channels included institutional websites, blogs, and, to a lesser extent, email newsletters and social media. The platforms most frequently mentioned were Twitter/X and LinkedIn (each cited three times), with Mastodon, YouTube, and BlueSky receiving single mentions. Additionally, some organizations collaborate with others in their outreach efforts, further broadening their reach (**Figure 31**).

## Landscape Analysis of PID Practices in NFDI

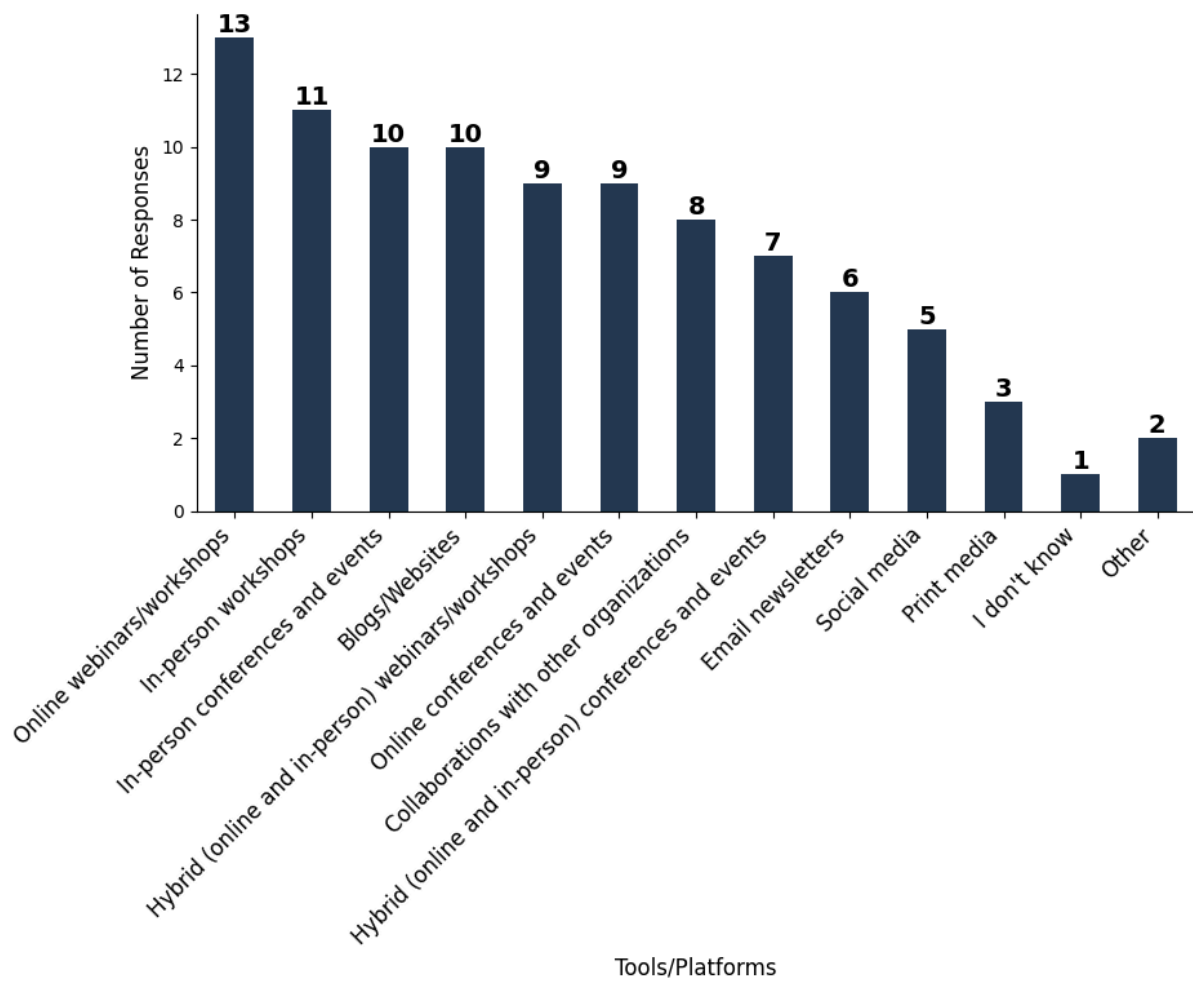


Figure 31 lists channels and tools used for outreach activities on the topic of PIDs.

The diversity of responses regarding preferred support for outreach from PID providers underscores the need for tailored approaches to meet specific organizational requirements. For the PID Coordination Hub, this highlights the importance of flexible outreach strategies and collaborations with stakeholders to maximize the effectiveness and reach of these efforts.

## 4. Expanding PID Adoption in NFDI: Challenges, Opportunities, and the Role of the PID Coordination Hub

The survey conducted among the NFDI consortia reveals key challenges and opportunities in the adoption and management of persistent identifiers. While many respondents are actively minting PIDs across their services using a range of providers, including 16 different PID systems, the fragmented nature of these efforts has led to inefficiencies and missed opportunities for greater cohesion and impact.

### 4.1. Methodological Limitations

Before delving into the next section, it is important to acknowledge certain methodological limitations of the survey:

- **Limited Service Descriptions:** Questions 16–27 did not request general service descriptions from respondents, such as the area of research, types of resources that can be registered with a repository, or general service capabilities. While this is not an issue when a URL is provided (as details can be looked up), it poses a problem for services marked as “planned” or “in development” without an accompanying URL. This leaves gaps in understanding the objectives of these services.
- **Metadata Schema Associations:** The survey did not directly ask which metadata schema is applied in each service. Currently, we can only associate metadata schemas with consortia and organizations, not with specific services.
- **Terminology Clarification:** Respondents indicated that terms like “organization,” “institution,” and “consortium” should have been more clearly defined within certain questions to minimize confusion.

## 4.2. Challenges and Opportunities in PID Adoption

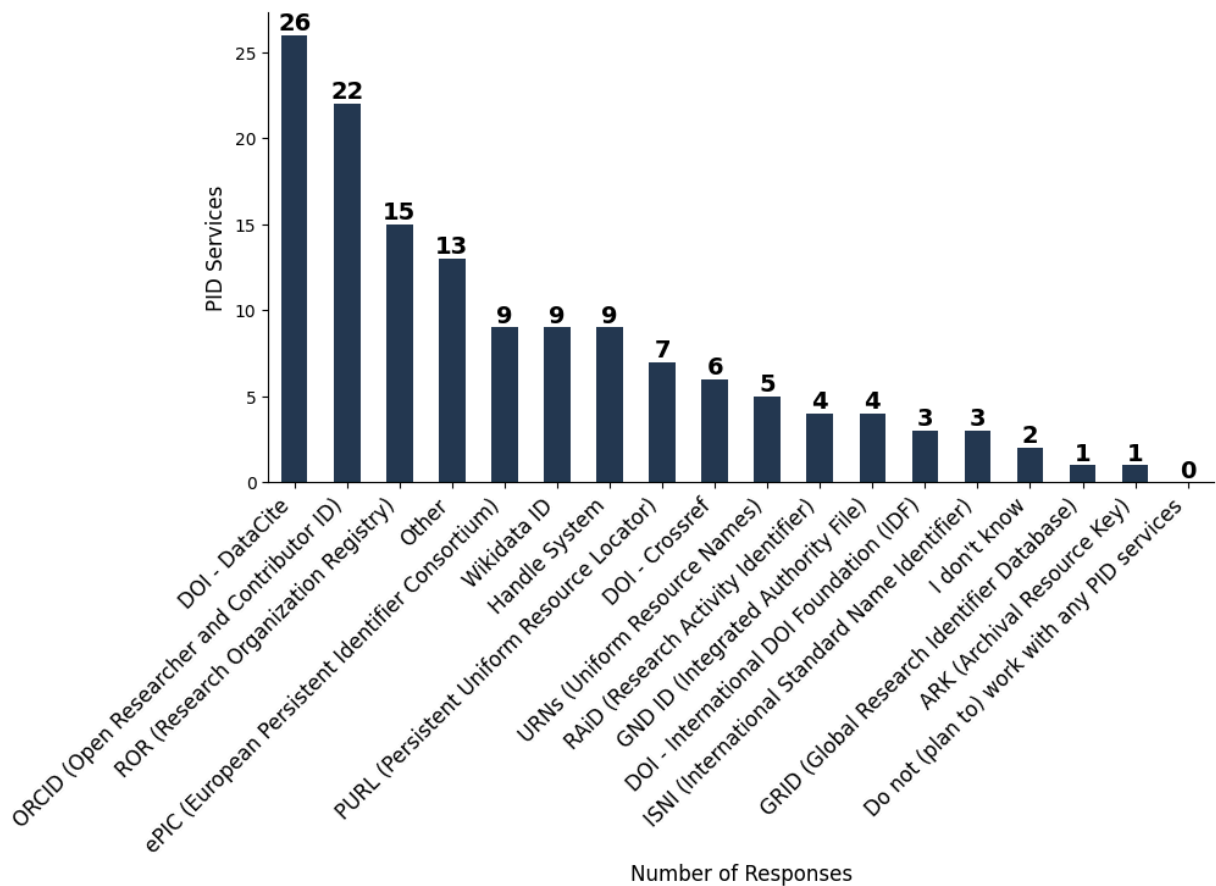


Figure 32 displays the distribution of PID providers that respondents (plan to) work with.

The survey results highlight that DataCite emerges as the most popular choice for PID services, with 76% of respondents reporting working with it (**Figure 32**). ORCID and ROR closely follow in popularity, with 65% and 44% of respondents working with them, respectively. ePIC and Handle are less common choices (26%).

The duration of engagement with these PID providers varies considerably (**Figure 33**), leading to inconsistencies in satisfaction levels and uneven familiarity with best practices.

## Landscape Analysis of PID Practices in NFDI

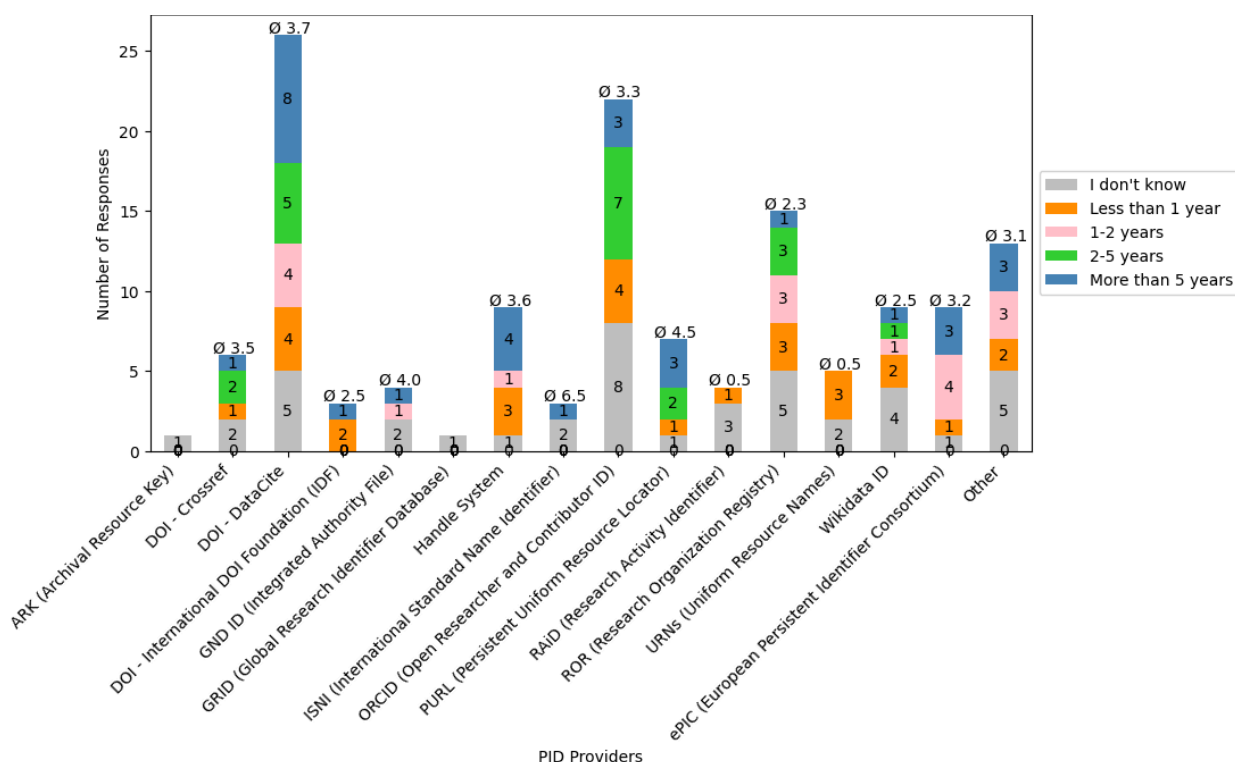


Figure 33 illustrates the duration of engagement with each PID provider across all survey responses. Averages are based on weighted categorical responses: "Less than 1 year" = 0.5, "1–2 years" = 1.5, "2–5 years" = 3.5, "More than 5 years" = 6.5.

For effective PID management, improving internal communication is crucial. This could involve:

- Documenting use cases and sharing lessons learned from projects.
- Clearly identifying responsible individuals, such as those overseeing DOI registration accounts or assisting with specific PID processes.

Such steps would prevent valuable institutional knowledge from being lost and ensure more consistent PID practices across teams and projects.

### 4.3. Trends in PID Registration and Future plans

The survey indicates a visible trend toward registering more PIDs, highlighting a growing awareness of the benefits of PID management in expanding research resource discoverability and interoperability.

While DataCite, ORCID, and ROR are the most popular choices, there is notable interest in other systems like the Handle System, PURL, and ePIC. The correlation

## Landscape Analysis of PID Practices in NFDI

between interest and the number of PIDs already registered (**Figure 34**) or planned for annual registration (**Figure 35**) suggests a preference for using these systems for managing large volumes of PIDs, typically over 10,000.

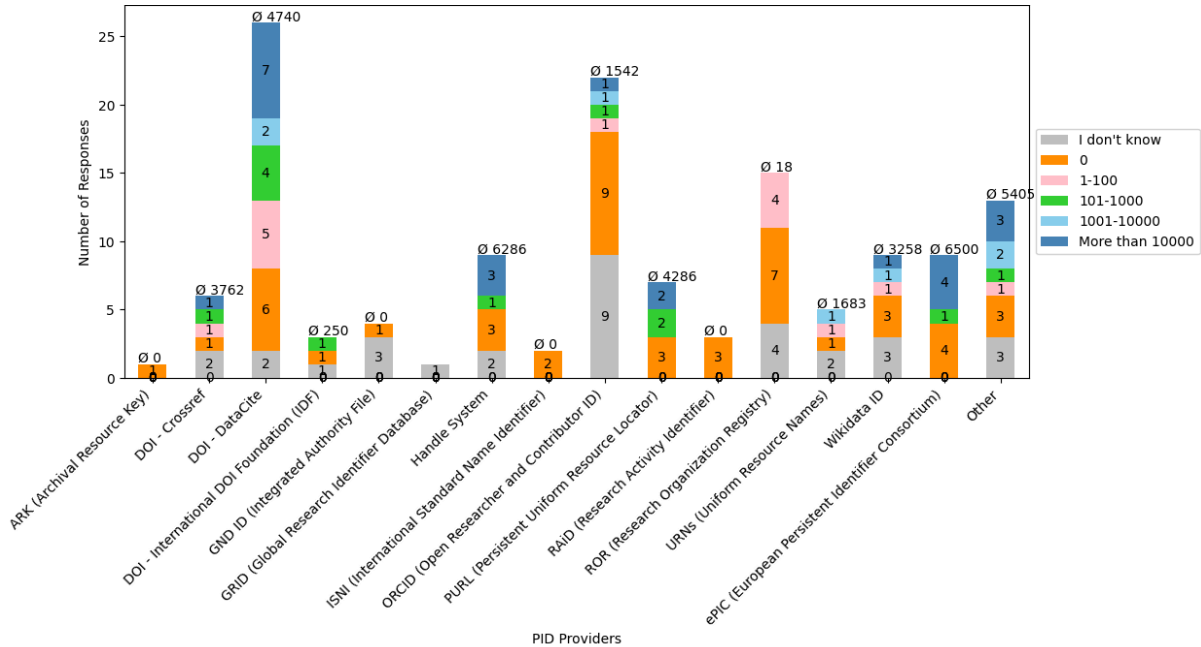


Figure 34 displays the estimated number of PIDs already registered with each PID provider. Averages are based on weighted categorical responses: "0" = 0, "1-100" = 50, "101-1000" = 500, "1001-10000" = 5000, "More than 10000" = 14500.

Many organizations express a pressing need to mint more PIDs, as shown by plans for large-scale PID minting (**Figure 35**). Additional trends observed are summarized in the appendix.

## Landscape Analysis of PID Practices in NFDI

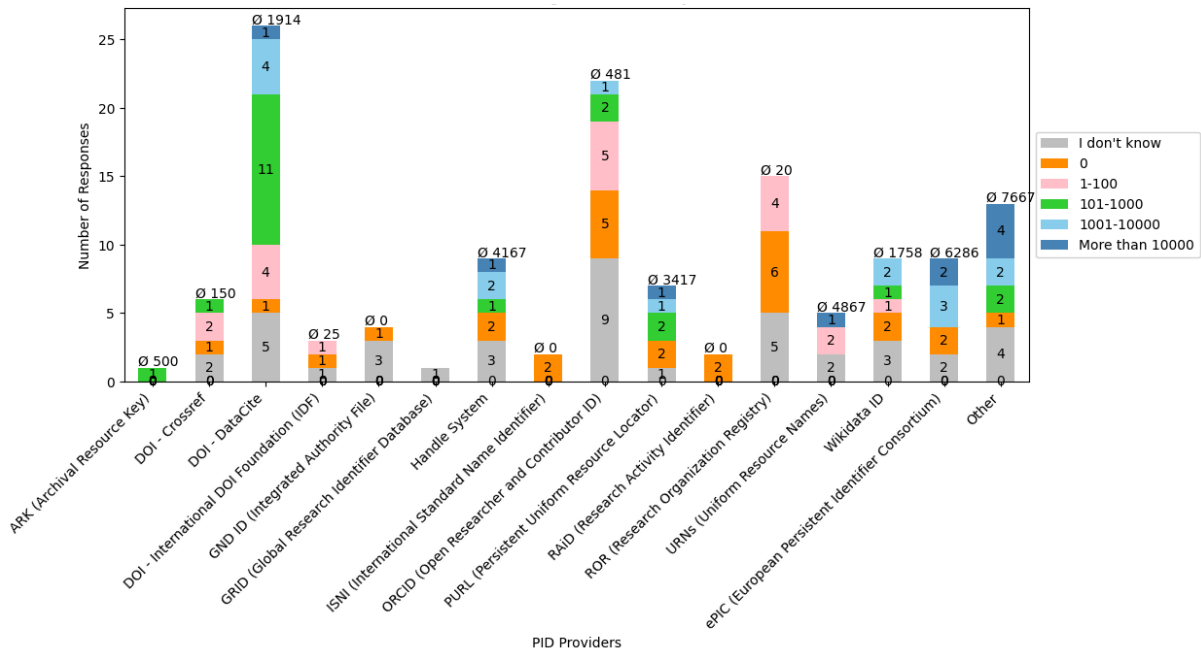


Figure 35 Perceived trends in the number of registered PIDs. Averages are based on weighted categorical responses: "Decreased" = -2, "Remained the same" = 0, "Slightly increased" = 1, "Moderately increased" = 2, "Significantly increased" = 3.

### 4.4. The Need for Improved Coordination and Collaboration

A significant observation from the survey is the lack of coordinated efforts in interoperability initiatives across different organizations. There is a clear demand for collaborative opportunities such as NFDI consortia where institutions working towards similar goals can share insights, tools, and best practices.

Examples of shared objectives:

- Metadata Mappings and Alignments:** Some organizations across different consortia share objectives in metadata mappings and alignments. Formalized sharing of lessons learned could be facilitated.
- ORCID Integration and FAIR Digital Objects:** Institutions focusing on ORCID integration and those advancing FAIR Digital Objects could benefit from coordinated efforts to align with PID service providers and ensure consistency.

By promoting collaboration and facilitating the exchange of knowledge and resources among institutions with shared goals, the PID Coordination Hub can play



a pivotal role in bridging these gaps. This collaborative approach will enhance interoperability and standardization efforts across NFDI.

## 5. Summary and Conclusion

### 5.1. Key Takeaways

#### 5.1.1. PID4NFDI Awareness and Expectations

The survey data reveals widespread awareness of the PID4NFDI project, attributed to prior outreach efforts and engagement activities. Respondents expressed clear expectations for support in areas such as best practices, standards development, cross-consortium collaboration, technical assistance, and training. Additionally, specific needs such as API development, migration pathways, and improvements to existing PID standards highlight the diverse requirements across consortia.

#### 5.1.2. Collaborative PID Implementation

PID implementation efforts are seen as collaborative opportunities, with common goals including setting up shared services, aligning standards across disciplines, and integrating metadata into generalized schemas like DataCite.

#### 5.1.3. Services and Infrastructures in Use

Respondents provided descriptions of 44 services, many of which were repurposed from prior projects or institutional contexts. A significant portion of these services is used across multiple consortia or integrated into broader frameworks like the European Open Science Cloud (EOSC). New services continue to be developed to address evolving needs.

#### 5.1.4. PID Providers and Membership Decisions

Integration with PID providers such as DataCite, ORCID, and RAiD varies across organizations, with many services in the development or prototype stages. Satisfaction levels with PID providers are generally positive, with technical compatibility, alignment with organizational needs, and ease of integration identified as key factors influencing membership decisions.

### 5.1.5. Resource Types and Trends

The NFDI community registers a wide range of resources, including datasets, software, metadata, and emerging categories such as instruments and physical objects. Trends such as granular PID use in social sciences and semantic web technologies illustrate the growing complexity and scope of PID applications.

### 5.1.6. Metadata and Interoperability

Metadata practices emphasize adherence to standards, with most respondents participating in discussions to improve quality and interoperability. RESTful APIs are the most widely used standard for integration with PID systems, and suggestions for improvement include better support for non-DOI PIDs and automatic metadata extraction.

### 5.1.7. Training and Support

Training remains an essential aspect of PID implementation, with most respondents attending sessions occasionally. DataCite is a leading provider of training on metadata and standards, reflecting its central role in harmonization efforts. For PID management support, technical assistance is the most sought-after service, followed by billing inquiries and best practices guidance.

Further information is available in the Appendix which provides a comprehensive overview with supplemental figures and their corresponding analyses.

The findings highlight the diverse needs, practices and expectations related to PID management and implementation within the NFDI community. The data underscores the importance of tailored solutions, collaboration, and targeted support to meet the varying demands of the NFDI consortia and their members. Efforts like PID4NFDI are well-positioned to address these challenges by fostering standardization, enhancing interoperability, and providing the necessary resources and training.

## 5.2. The Role of the PID Coordination Hub in Addressing Challenges

Through its unique role in bridging disparities between NFDI consortia, the PID Coordination Hub has the potential to alleviate the burdens of effective PID registration and management by delivering structured and consistent support.

By focusing on developing infrastructure that is both stable and scalable while remaining inclusive and supportive, the PID Coordination Hub can foster an environment where all organizations, regardless of their current maturity level, are able to mint, manage, and utilize PIDs effectively. This effort will ensure that even organizations with fewer internal resources can participate fully in the benefits of PID management.

To achieve this, the Hub can implement several key initiatives:

### 1. Offer Training on Best Practices

- **Standardized Training Programs:** Establish training focused on critical topics such as metadata quality, API integration, and PID fundamentals. This will support organizations in building the necessary competencies for effective PID management.
- **Interactive and Flexible Formats:** Prioritize live, interactive training sessions like webinars for both beginners and experts. To cater to advanced users, include specialized webinars on topics like API integration for specific research contexts.
- **Supporting Materials:** Provide documentation and resources for asynchronous learning to accommodate diverse needs and schedules.

### 2. Develop Tiered Training Pathways

- **Foundational Courses:** Offer introductory courses for newcomers to PID management.
- **Advanced Workshops:** Conduct advanced workshops focusing on technical aspects such as interoperability and metadata standards.

- **Consistency in Knowledge:** This tiered approach ensures consistency in PID knowledge across NFDI, addressing current training disparities.

### **3. Establish a Knowledge-Sharing Platform**

- **Institutional Learning Base:** Create platforms where use cases, lessons learned, and responsible roles are documented. This will prevent the loss of valuable institutional knowledge.
- **Promote Consistency:** Encourage consistent PID practices across teams, projects, organizations, and consortia by sharing insights and experiences.

### **4. Provide Recommendations for an NFDI-wide PID Governance Framework**

- **Clear Governance Structures:** Develop guidelines to ensure consistent PID adoption across all organizational levels.
- **Reduce Disparities:** A governance framework will reduce disparities in satisfaction and performance between different groups.

### **5. Facilitate Collaboration and Alignment**

- **Metadata Schemas and Knowledge Graphs:** Serve as a platform for stakeholders to work together on aligning metadata schemas and developing knowledge graphs.
- **Advance FAIR Principles:** Support the pursuit of FAIR (Findable, Accessible, Interoperable, and Reusable) principles across the NFDI community.
- **Scale Efforts Community-Wide:** By consolidating and sharing progress made by individual organizations, the Hub can scale these efforts across the entire NFDI community.

### **6. Centralized Metadata Support**

- **Training Resources on Metadata Practices:** Provide centralized resources focusing on metadata completeness and quality, which are essential for the reliability and discoverability of research outputs.

- **Address Organizational Gaps:** Fill gaps in organizational support for metadata practices, enhancing overall data quality.

### 5.3. Implications for the NFDI Landscape

By implementing these initiatives, the PID Coordination Hub will:

- **Cultivate a Cohesive Strategic Direction:** Contribute to a more unified approach to PID adoption across the entire NFDI landscape.
- **Enhance Interoperability:** Foster an integrated and effective approach to PID management, improving interoperability among research outputs.
- **Empower Organizations:** Enable all organizations, regardless of their current maturity level, to fully realize the benefits of effective PID management.
- **Promote Sustainability:** Enhance the long-term sustainability and discoverability of research outputs through high-quality metadata and consistent PID practices.

In summary, the coordinated efforts facilitated by the PID Coordination Hub aims to bridge existing gaps in PID adoption and management within the NFDI. By offering comprehensive support in training, governance, collaboration, and metadata practices, the Hub will empower organizations to better address current challenges. This strategic approach should lead to a robust, unified PID ecosystem, enhancing the global impact and sustainability of research conducted within the NFDI community.

## 6. Supplemental material

‘Survey Question Catalog for D1.1 Landscape of PID Practices Within NFDI Services’ available in Zenodo <https://doi.org/10.5281/zenodo.14327774>.

## 7. Appendix

The appendix contains a collection of supplemental figures along with accompanying textual explanations, descriptions, and interpretations to provide additional insights.

### 7.1. PID4NFDI Awareness, Involvement, and Support Expectations

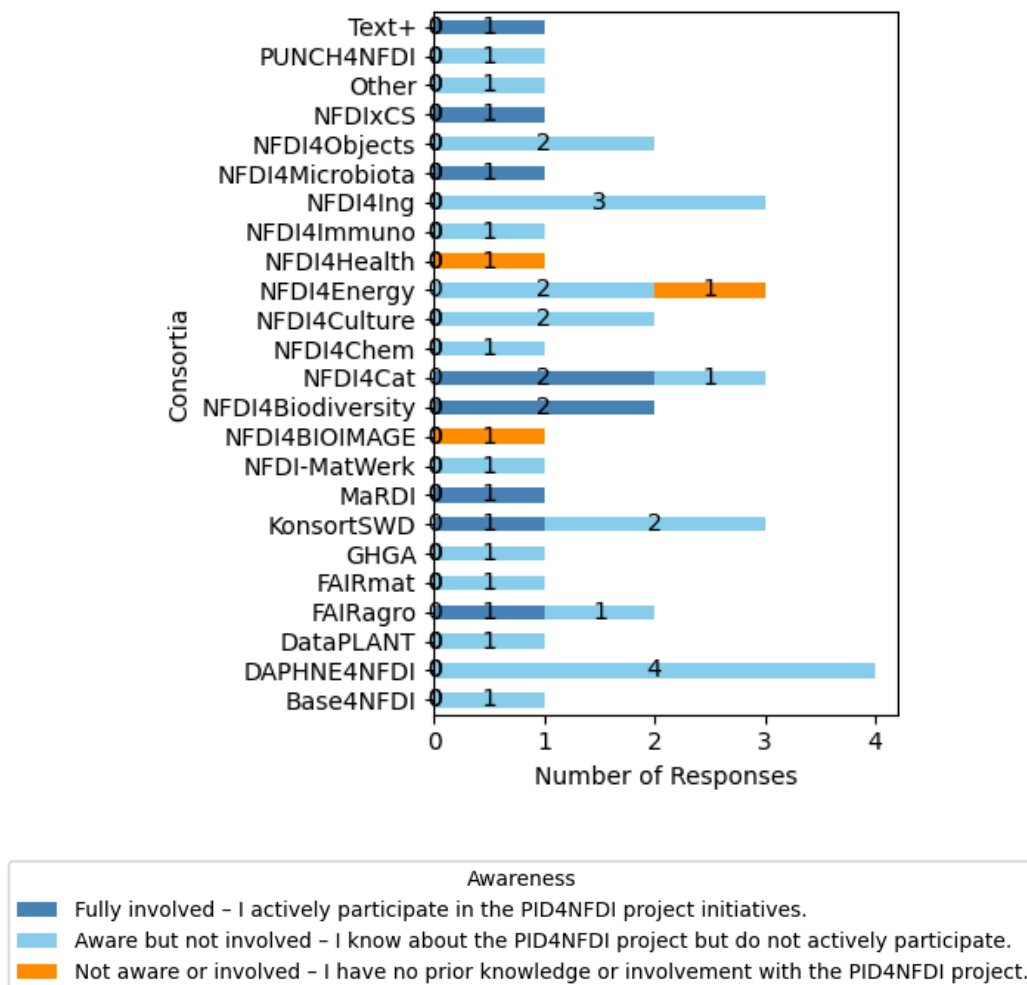


Figure S1 Involvement and Awareness of PID4NFDI project across various consortia representatives.

**Figure S1** illustrates the involvement and awareness of the PID4NFDI project among representatives from various NFDI consortia. The image indicates that PID4NFDI is already well-known by many representatives actively working with PIDs. This widespread awareness is largely due to the targeted approach taken before launching the survey: many respondents were informed about the project and the

survey beforehand, while others had been involved in earlier surveys and workshops during the preparation of the PID4NFDI initial project proposal<sup>8 9 10</sup>.

To design an effective work programme for both the initialisation and integration phase, and to ensure that all planned activities align with the needs of the NFDI community, survey respondents were asked about their support expectations from PID4NFDI. **Figure S2** reveals that the majority of respondents expect support in areas such as establishing best practices, developing standards and policies, fostering cross-consortium collaborations, and providing technical support and resources. Training sessions are also considered important by half of the respondents.

---

<sup>8</sup>Bingert, Sven, Jan Brase, Felix Burger, Britta Dreyer, Stephanie Hagemann-Wilholt, Paul Vierkant, and Philipp Wieder. 'Concept for Setting up the Persistent Identifier Services Working Group in the NFDI Section "common Infrastructures"'. Zenodo, 30 April 2022. <https://doi.org/10.5281/zenodo.6507760>.

<sup>9</sup> Schrader, Antonia C., Daniel Arend, Janete Bach, Kirsten Elger, Sandra Göller, Stephanie Hagemann-Wilholt, Rolf Krah, et al. 'Workshop on Pids Within NFDI'. Zenodo, 13 February 2023. <https://doi.org/10.5281/zenodo.7635905>.

<sup>10</sup> Schrader, Antonia C., Stephanie Hagemann-Wilholt, and Andreas Czerniak. 'Isn't a Number and a URL Enough? Why Pids Matter and Technical Solutions Alone Are Not Sufficient.'. Presented at the 1st Conference on Research Data Infrastructure (coRDI), Karlsruhe, Germany, 8 September 2023. <https://doi.org/10.5281/zenodo.8328807>.

## Landscape Analysis of PID Practices in NFDI

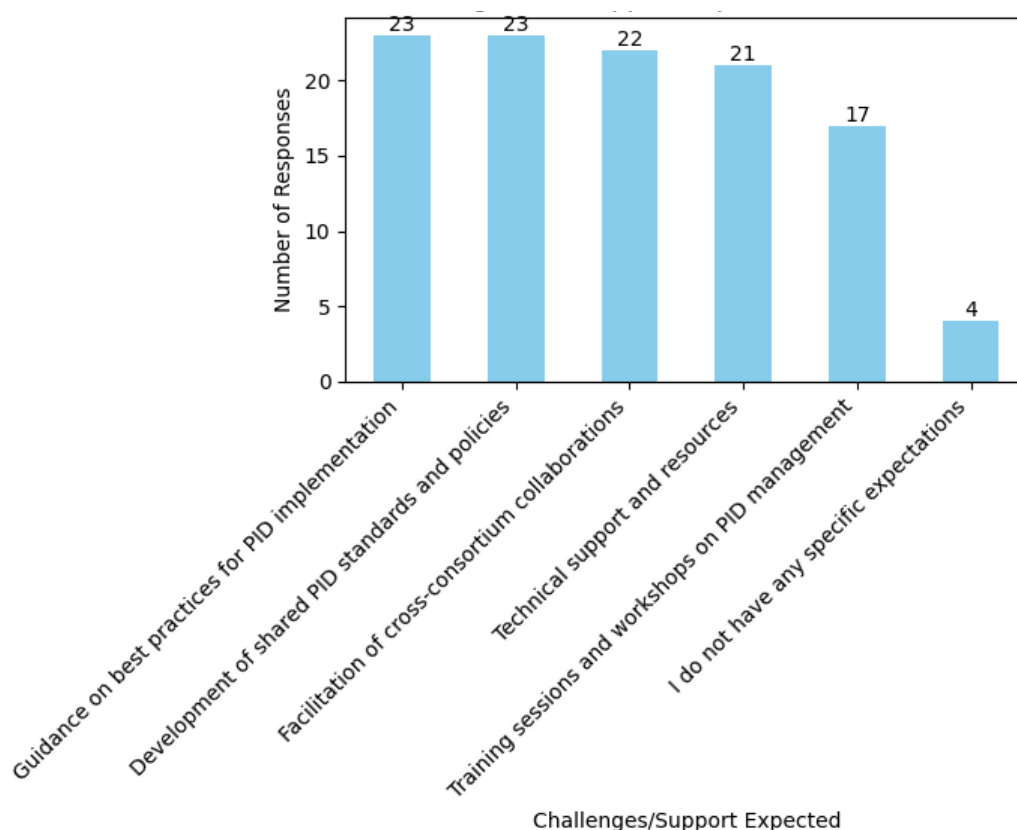


Figure S2 Support Expectations from PID4NFDI as a Cross-Consortium Base Service.

In addition to the more common support areas, other specific expectations mentioned by respondents include:

- Provision of a PID-issuing service (e.g., for samples).
- Creation and long-term financing of an API for PID registration.
- Support for implementing PIDs other than DOIs.
- Improvement of existing PID standards, such as those for Research Organization Registry (ROR).
- Guidelines for using PID citation graphs as Key Performance Indicators (KPIs).
- Provision of migration pathways for existing permanent identifiers from legacy systems.



These responses highlight the diverse and practical support needs across the NFDI consortia, underscoring the importance of the PID4NFDI project in facilitating effective PID implementation and management.

**Figure S3** highlights the collaborative nature of PID implementation efforts across different NFDI consortia. The figure shows that most surveyed consortia representatives view PID integration as a cross-consortium initiative with significant potential for collaboration. Common collaborative initiatives identified by respondents include:

- Setting up PID services that benefit multiple consortia.
- Establishing Wikibase PIDs to enhance the management of persistent identifiers.
- Improving existing PID standards to ensure better alignment across disciplines.
- Participating in Research Activity ID (RAiD) services.
- Developing approaches to integrate discipline-specific metadata into more generic schemas such as DataCite, potentially via related identifiers.

## Landscape Analysis of PID Practices in NFDI

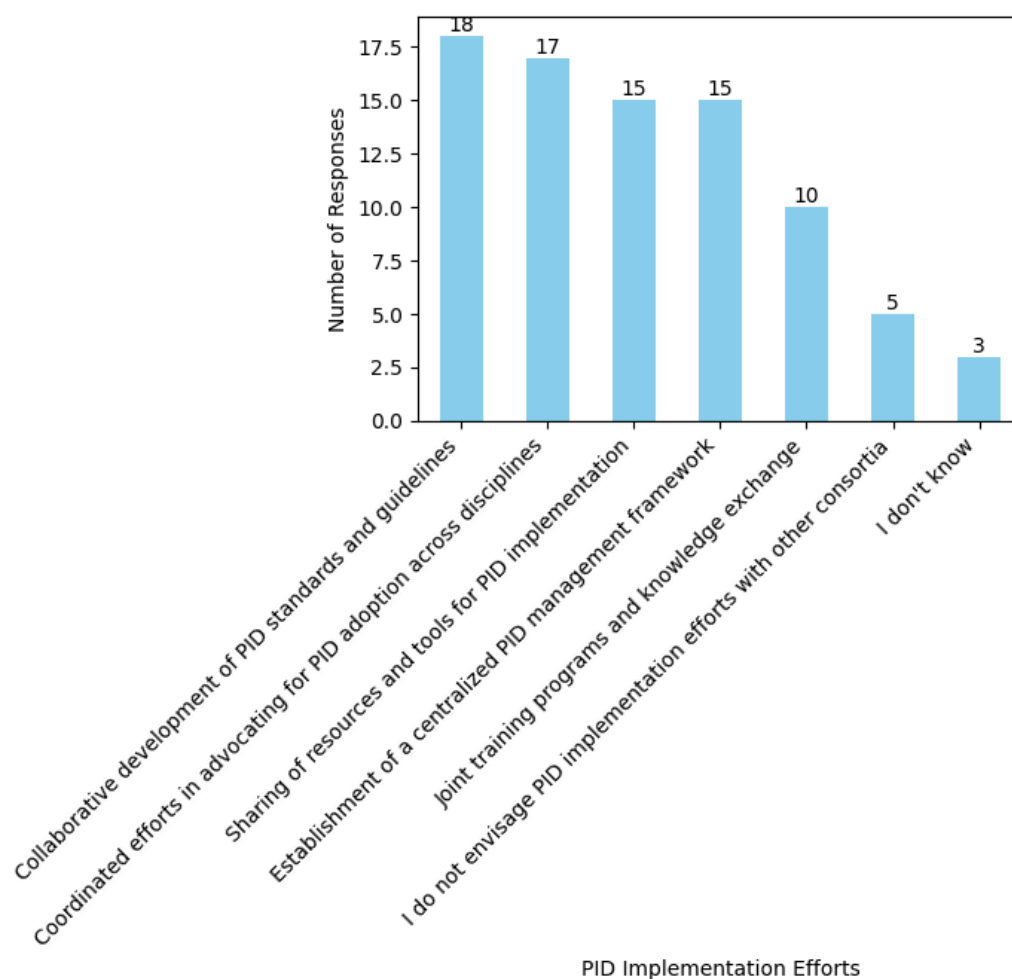


Figure S3 Shared PID implementation efforts among NFDI consortia.

## 7.2. Landscape of Services/Infrastructures in NFDI

This section gives an overview of services/infrastructures related to PIDs that are used in the NFDI communities. This might include for example PID registration services, tools for finding, archiving, processing, sharing and exchanging research data based on PIDs, as well as collaborating in shared working environments. It is important to note that this is not an exhaustive list of PID-related services but represents a selection of key services currently in use. In total, 44 service descriptions were provided, with 42 of these referring to distinct services. The majority of service descriptions were given by DAPHNE4NFDI and NFDI4Ing (5 descriptions each), followed by FAIRagro, NFDI4Cat, and NFDI4Energy (4 descriptions each), as shown in **Figure S4**.

## Landscape Analysis of PID Practices in NFDI

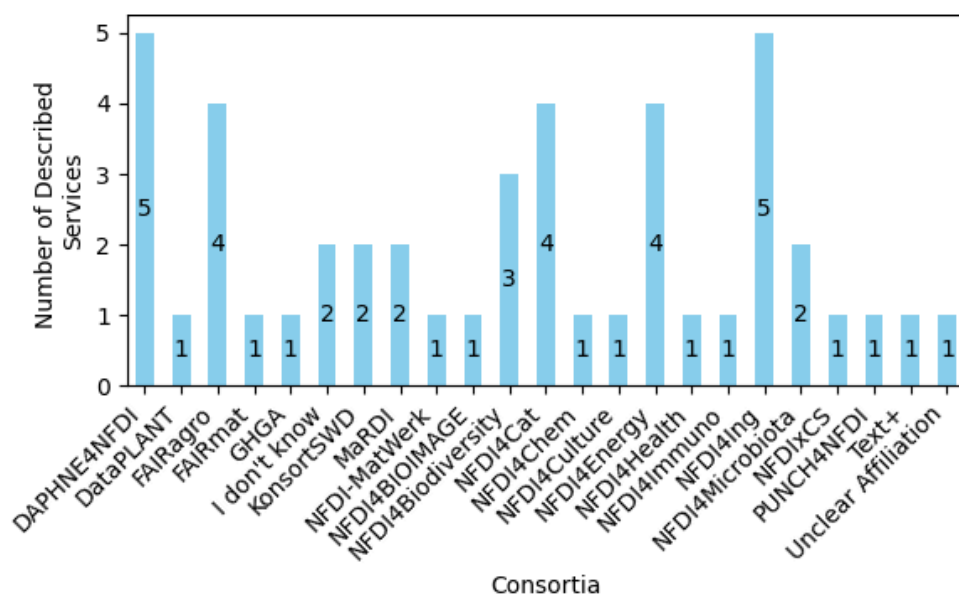
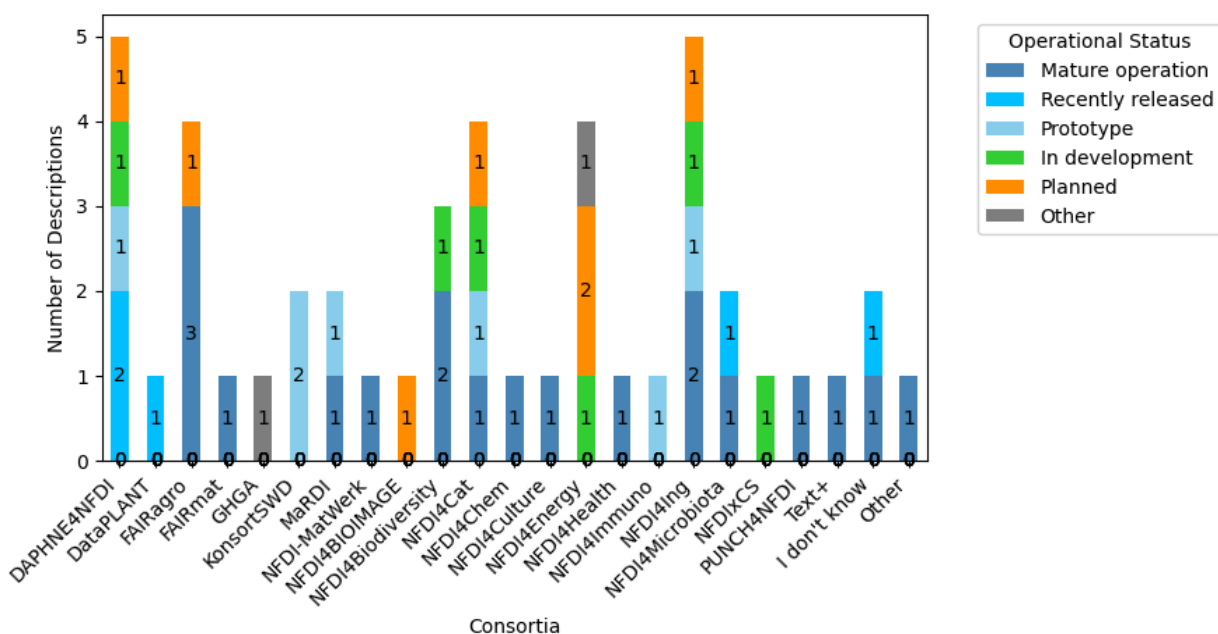


Figure S4 Number of described services for each consortium.

**Figure S5a** displays the operational status of the services and infrastructures stratified by consortia. It indicates that new services are being consistently developed across most consortia, while existing services serve a wide range of purposes (**Figure S5b**).

a)



b)

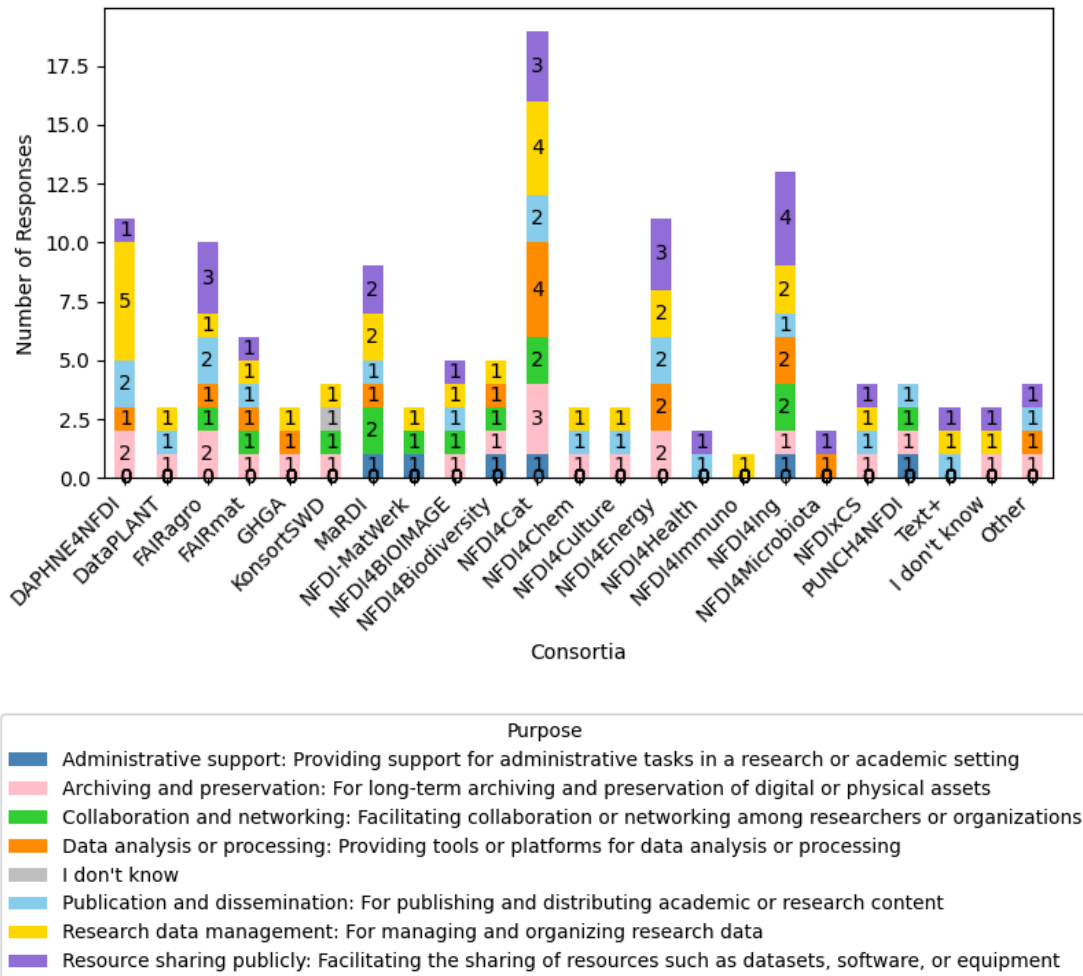


Figure S5 a) Operational status of infrastructures by consortia. b) Main purposes of the infrastructures used by each consortium.

**Figure S6** shows that 46% (20 out of 44) of the described services were in existence prior to their application in the NFDI consortia. This suggests that many services were initially developed in a different context, such as projects or institutional services, and later adapted for use within NFDI. In comparison, 41% of the described services were specifically developed for the NFDI consortia.

Some services are also used beyond their original consortia. Combining the "Yes" and "Other" 30% (13 out of 44 unique responses) of the services are used by multiple consortia, while an additional 2% are planned for cross-consortium usage, totaling 32% of services with cross-consortium applications. Additionally, 9% of services are used in the European Open Science Cloud (EOSC).

## Landscape Analysis of PID Practices in NFDI

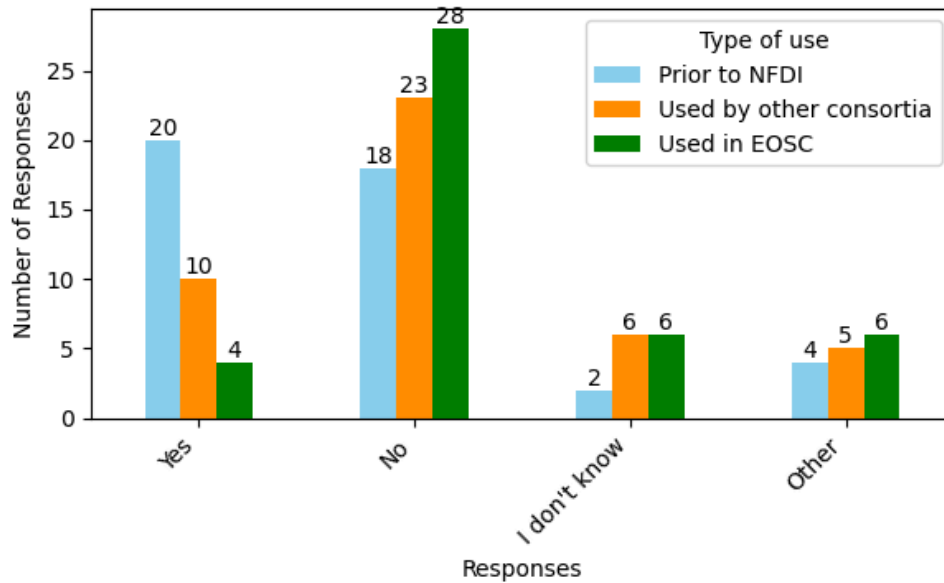


Figure S6 Existence of Services Prior to NFDI versus EOSC usage and Cross-Consortium Usage.

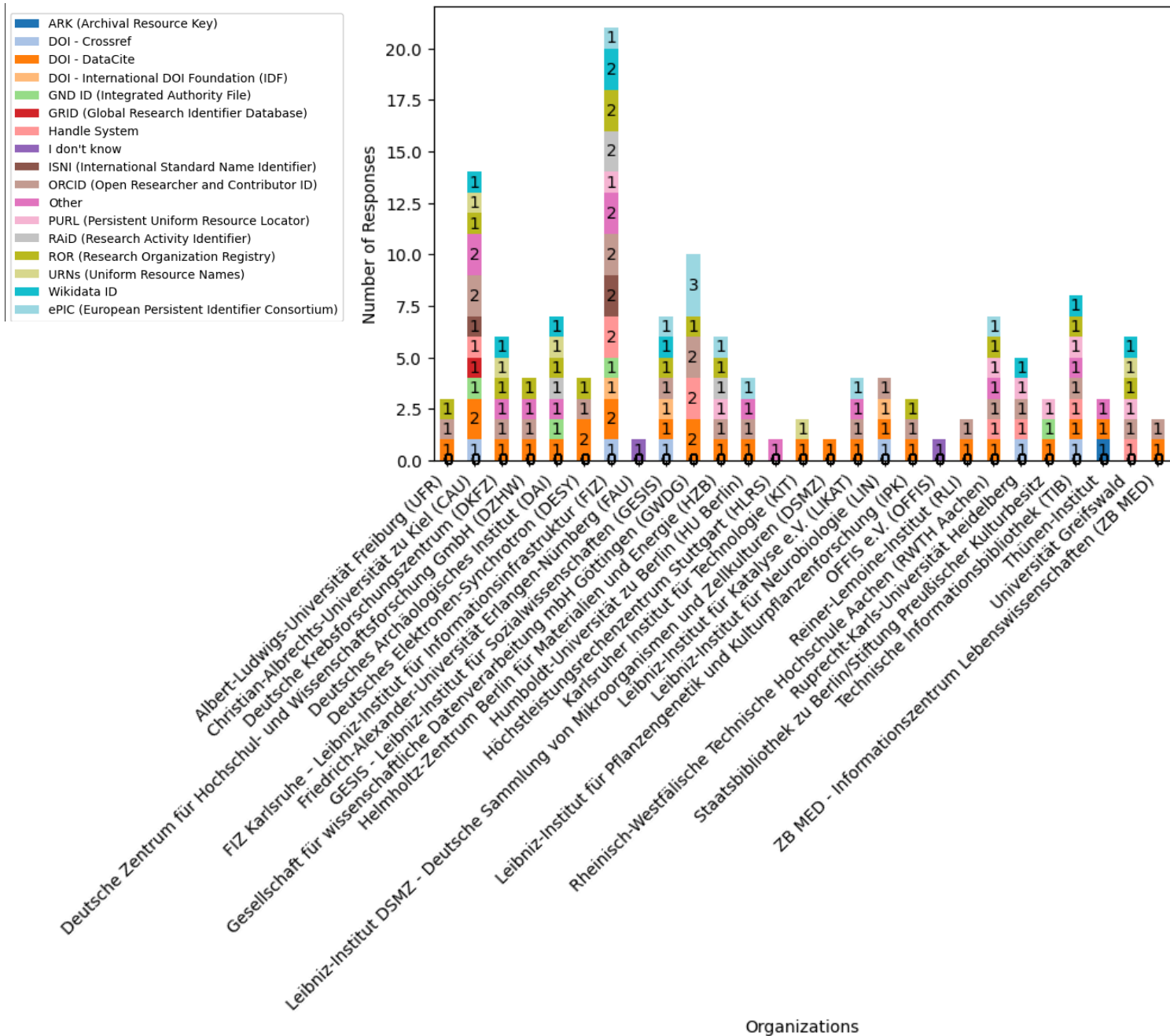
Other uses of these infrastructures, as described by respondents, mainly involve specific projects or institutional applications, further showcasing the versatility and broader impact of these services.

### 7.3. PID Provider Affiliation by Organization

**Figure S7a** provides a more detailed view of which organizations either currently work with or plan to work with specific PID providers, as indicated by the survey responses.

# Landscape Analysis of PID Practices in NFDI

a)



## Landscape Analysis of PID Practices in NFDI

b)

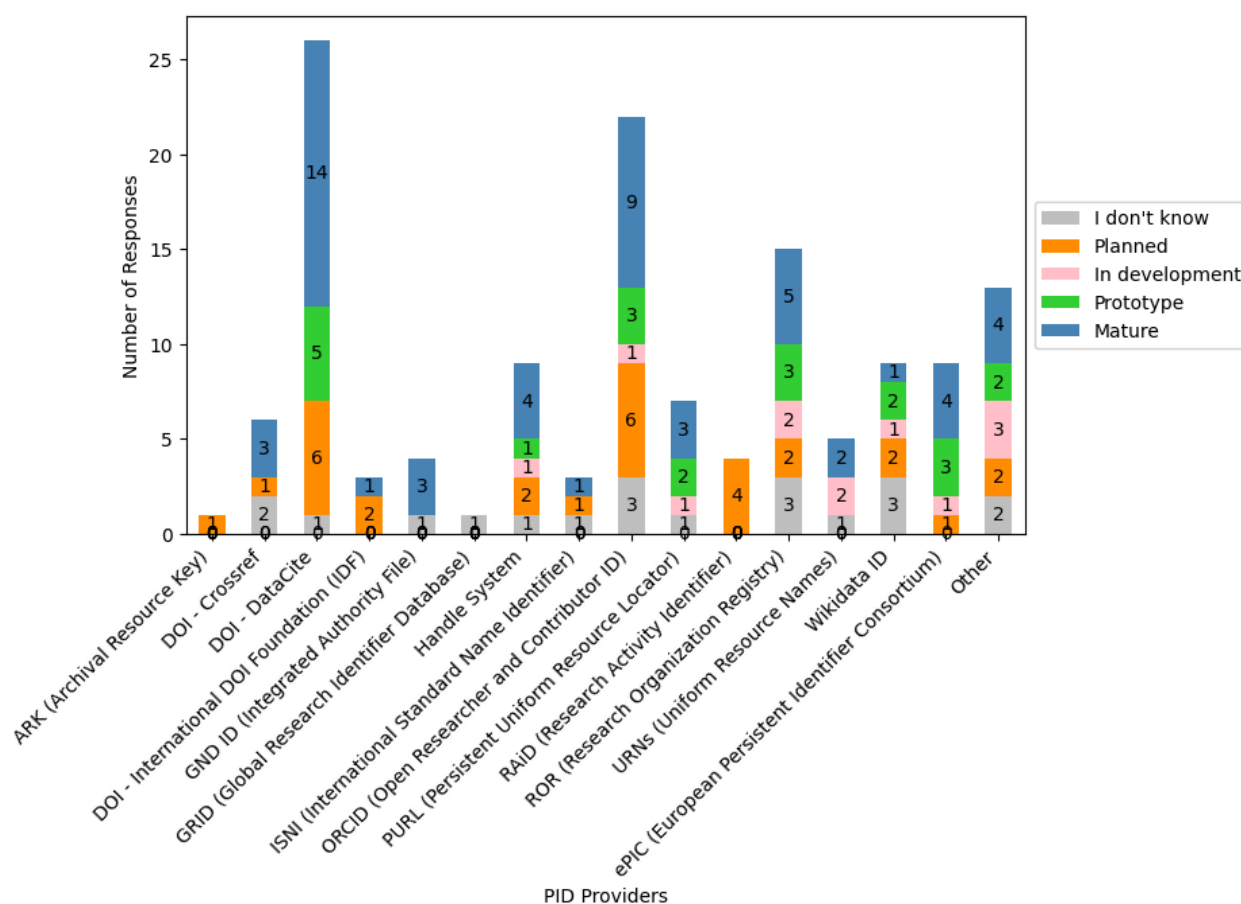


Figure S7 a) Organizations (planning to) work with PID providers. b) Current phase of PID integration for each service provider.

For those respondents who indicated that their organization works with or plans to work with a specific PID provider, **Figure S7b** shows the current phase of integration with the same PID provider. The figure reveals that several organizations have plans to integrate with providers such as DataCite, ORCID, and RAiD. Additionally, multiple services are still in development or in prototype stages, while the majority are already in mature operation.

The figure below displays the satisfaction levels of respondents towards PID providers. The averages displayed on top of the bars range from -2 (very dissatisfied) to 2 (very satisfied). The PID providers receiving the highest satisfaction scores are Wikidata ID, with a score of 1.2, followed by Handle System, ePIC, and GND, each with a score of 1.0.

## Landscape Analysis of PID Practices in NFDI

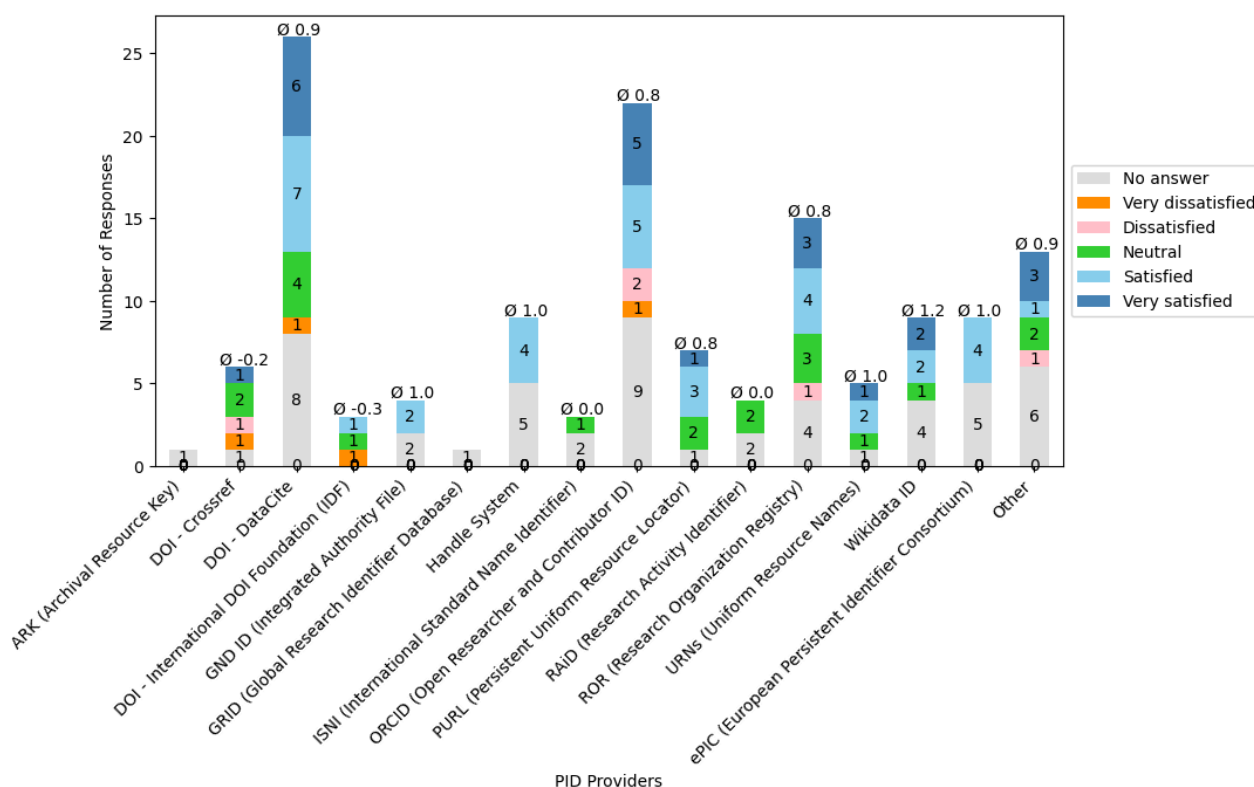


Figure S8 Satisfaction level with each PID Service. Averages are based on weighted categorical responses: “Very dissatisfied” = -2, “Dissatisfied” = -1, “Neutral” = 0, “Satisfied” = 1, “Very satisfied” = 2.

Respondents were also asked to rank factors influencing PID membership decisions based on their importance. **Figure S9** shows that the most important factors are "Alignment with organizational needs" and "Technical compatibility," both scoring 2.4. "Ease of integration" (score: 2.1) and "Cost factors" (score: 2.0) were also deemed important. In contrast, "Community recommendations/Peer influence" and "Familiarity and awareness of PID and related services" had lower influence scores (1.7).



## Landscape Analysis of PID Practices in NFDI

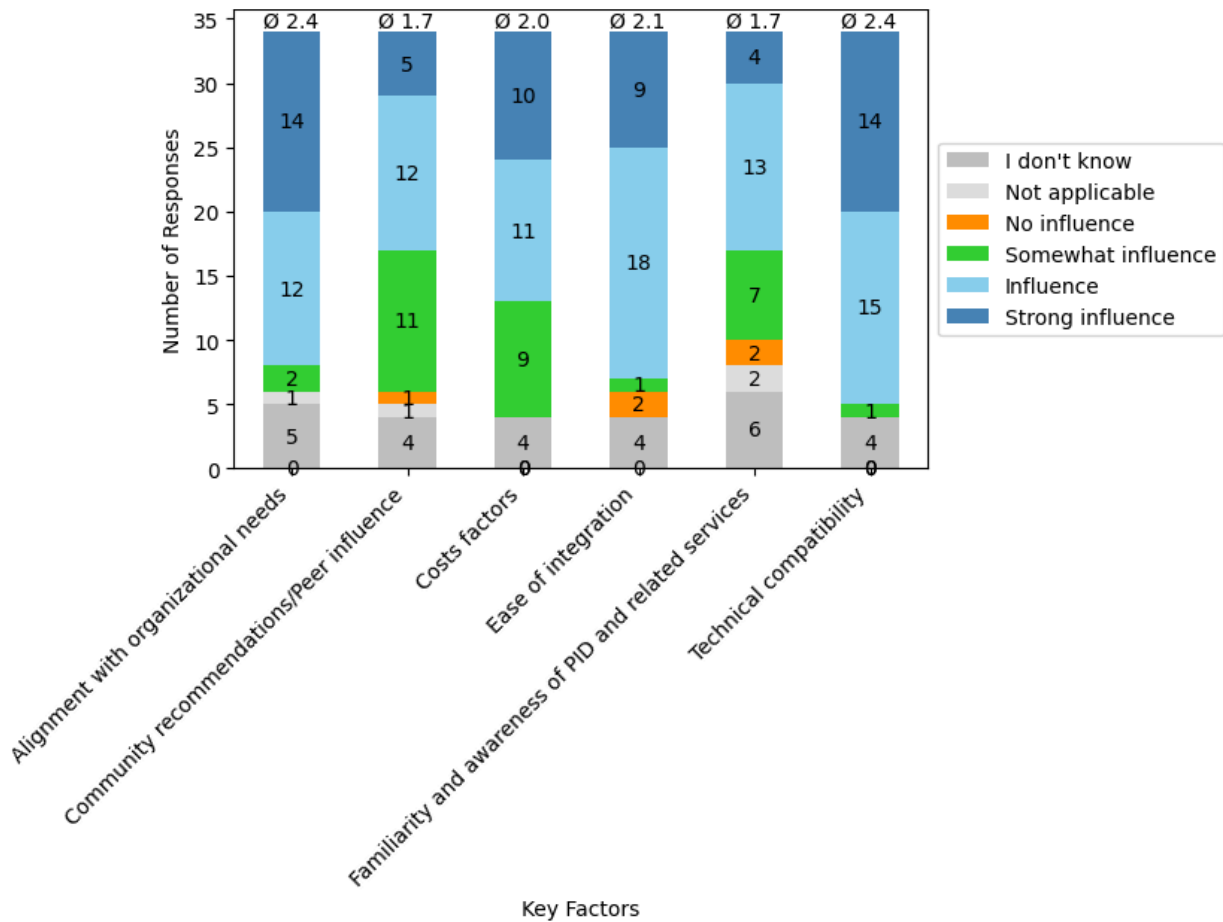


Figure S9 Key factors influencing PID membership Decisions. Averages are based on weighted categorical responses: “No influence”=0, “Somewhat influence”=1, “Influence”=2, “Strong influence”=3.

Other key factors mentioned by respondents include:

- **Metadata Requirements and Flexibility:** Ensuring the design or scheme of identifiers fits the specific use case (whether for human or machine use).
- **Accessibility and Searchability of Data:** Making data easy to locate and use.
- **Community Recommendations:** Discussions with the research community and, ideally, endorsements by NFDI.
- **API Availability:** The presence of well-documented APIs for integration.

## 7.4. Resource Types and Trends

**Figure S10** shows the diverse range of resource types being adopted within the NFDI community. Nearly all respondents (94%) indicated that they register PIDs for datasets, followed by software, code, and tools (62%), metadata (53%), and multimedia files (47%). Newer resource types include instruments (35%), physical objects (32%), variables in datasets (29%), and Data Management Plans (DMPs) (27%).

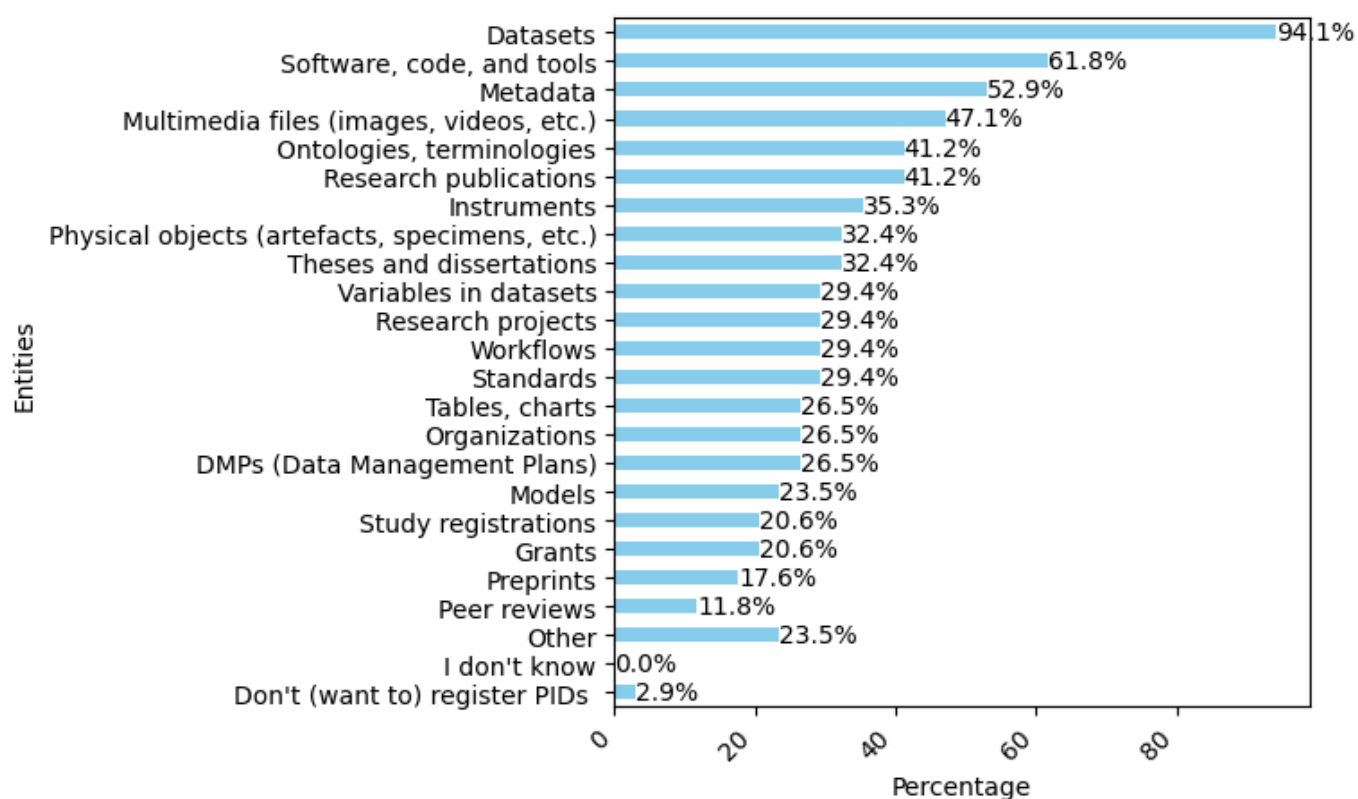


Figure S10 Entities with plans for registration.

**Figure S11** provides an overview of resource types registered by each consortium. The figure illustrates a clear correlation between resource types and the research disciplines represented by the various consortia, demonstrating that the types of resources being registered are often discipline-specific.

## Landscape Analysis of PID Practices in NFDI

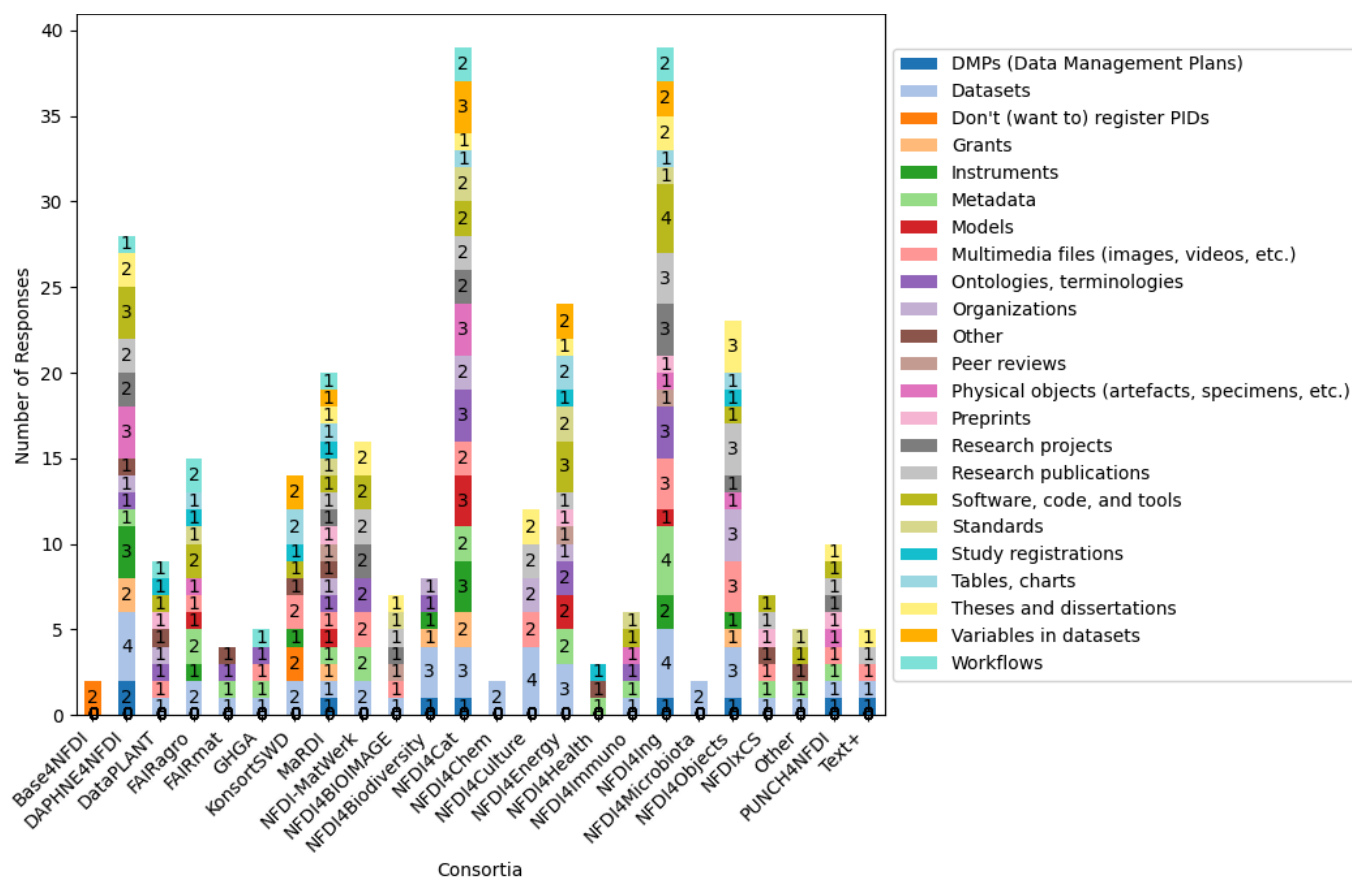


Figure S11 Resource Types registration stratified by consortia stratified.

In addition to the trends illustrated in Figure 35, respondents listed the following perceived trends:

- Semantic Web Technologies:** There is a growing trend toward using semantic web technologies as an alternative approach for defining and interconnecting concepts, resources, and entities. These technologies often incorporate PIDs to some extent, particularly for linking with PID providers.
- Granular PID Usage in Social Sciences:** In the social sciences, the intention to use PIDs at a more granular level (e.g., for variables, code, instruments) has significantly increased in recent years. While the implementation of such practices is still pending in many cases, there is a clear rise in awareness of the importance of PIDs beyond just study or dataset levels.

## 7.5. Metadata

Survey respondents were also asked about their metadata practices, including the schemas and standards used in their repositories. In addition to listing metadata schemas and standards (**Figure 18**), respondents provided descriptions of extensions and mappings to these schemas. 44% of respondents described specific mappings or extensions. An overview of applied metadata standards, mappings, and extensions is available in our deliverable D2.2 “Catalog of Relevant Metadata Standards”<sup>11</sup>.

79% of survey respondents answered that they are involved in groups or forums that focus on metadata issues, including discussions on standards, quality and enhancements (Q38). 70% of those metadata groups focus on specific scientific fields or disciplines (Q39). An overview of those metadata groups can be found in D2.2 as well.

## 7.6. Interoperability and API Usage

**Figure S12** shows the various standards and protocols implemented within NFDI repositories to ensure interoperability with different PID systems and services. The most commonly used standard is RESTful APIs, which are utilized by 74% of respondents (25 out of 34), followed by OAI-PMH (53%) and Linked Data and RDF (44%).

---

<sup>11</sup> Böhm, Jana. ‘D2.2. Catalog of Metadata Standards Relevant to NFDI’. Zenodo, 23 December 2024. <https://doi.org/10.5281/zenodo.14512768>.

## Landscape Analysis of PID Practices in NFDI

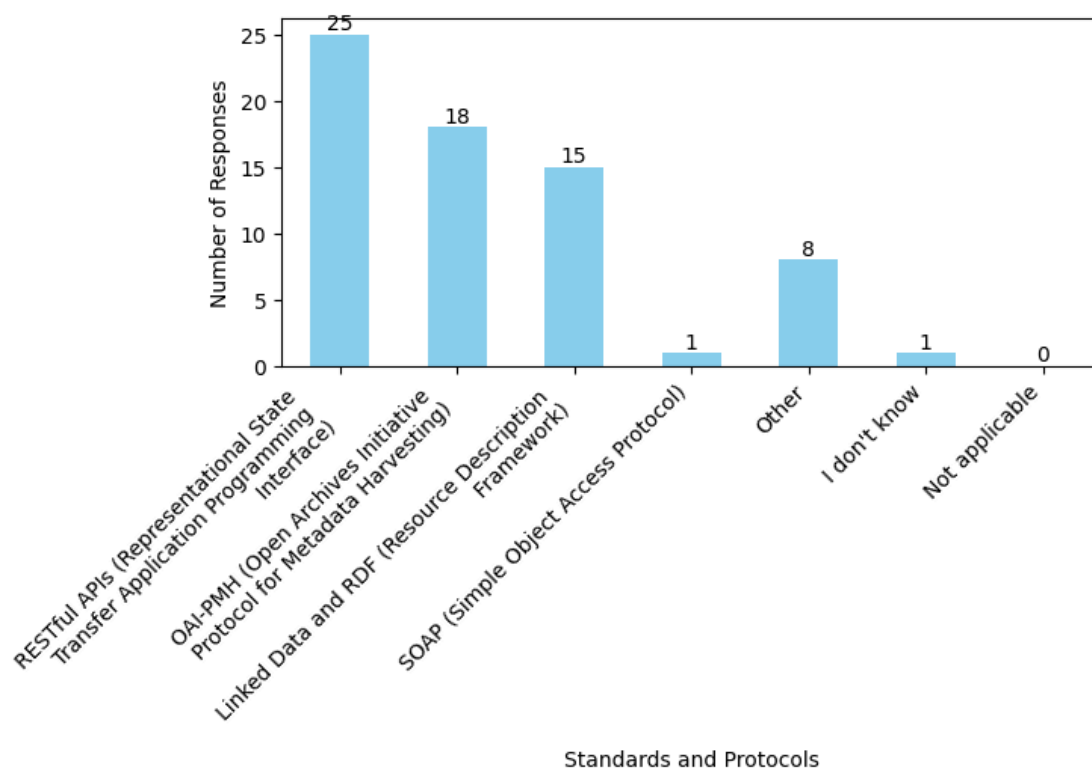


Figure S12 Standards and Protocols Implemented for Interoperability.

**Figure S13** provides an overview of the specific PID service APIs used within each consortium. The selection of APIs varies based on the particular PID services each consortium works with, illustrating the popularity of DataCite API across different NFDI consortia.

## Landscape Analysis of PID Practices in NFDI

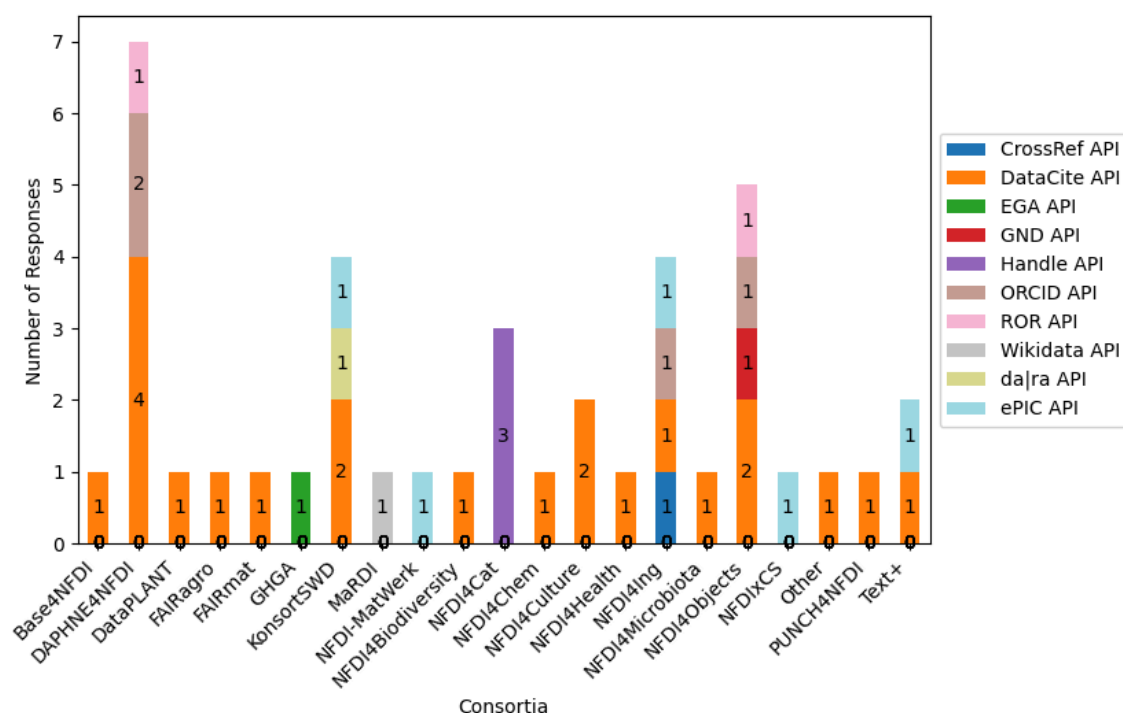


Figure S13 APIs used by each consortium.

Respondents also provided suggestions for improvements in the area of technical interoperability with PID providers, including:

- **Better Support for Intrinsic Identifiers:** Improved support for intrinsic identifiers, such as hashes and decentralized solutions for PIDs.
- **Improvements in ROR:** Enhanced granularity for resolving sub-units within the Research Organization Registry (ROR).
- **Automatic Metadata Extraction:** Enabling PID providers to automatically extract most of the required metadata from submitted datasets, reducing manual entry requirements.
- **Tools for Non-DOI PIDs:** Better-maintained tools for working with non-DOI PIDs.
- **DataCite Model:** A suggestion that DataCite should treat NFDI consortia as organizations and assign an identifier to them.

## 7.7. Training

**Figure S14** illustrates the frequency of respondents' attendance at training courses. Most participants reported attending training rarely (once a year or less) or occasionally (2–3 times a year), with DataCite training emerging as the most frequently attended option.

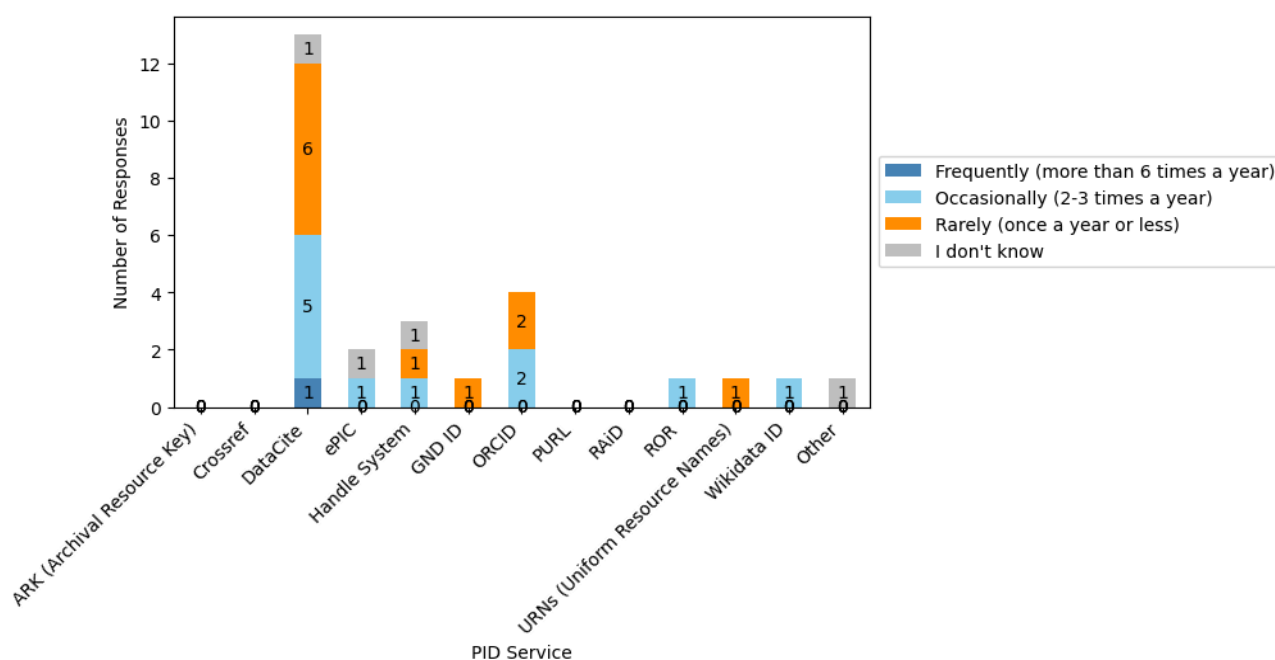
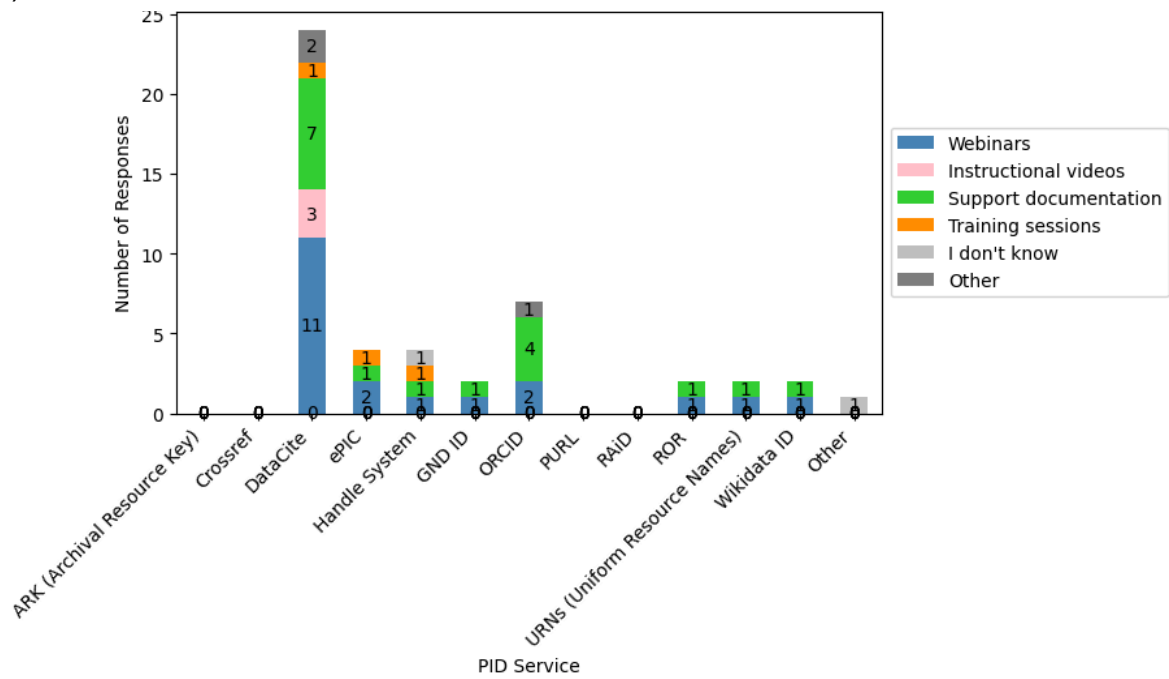


Figure S14 Frequency of participation in training events and webinars.

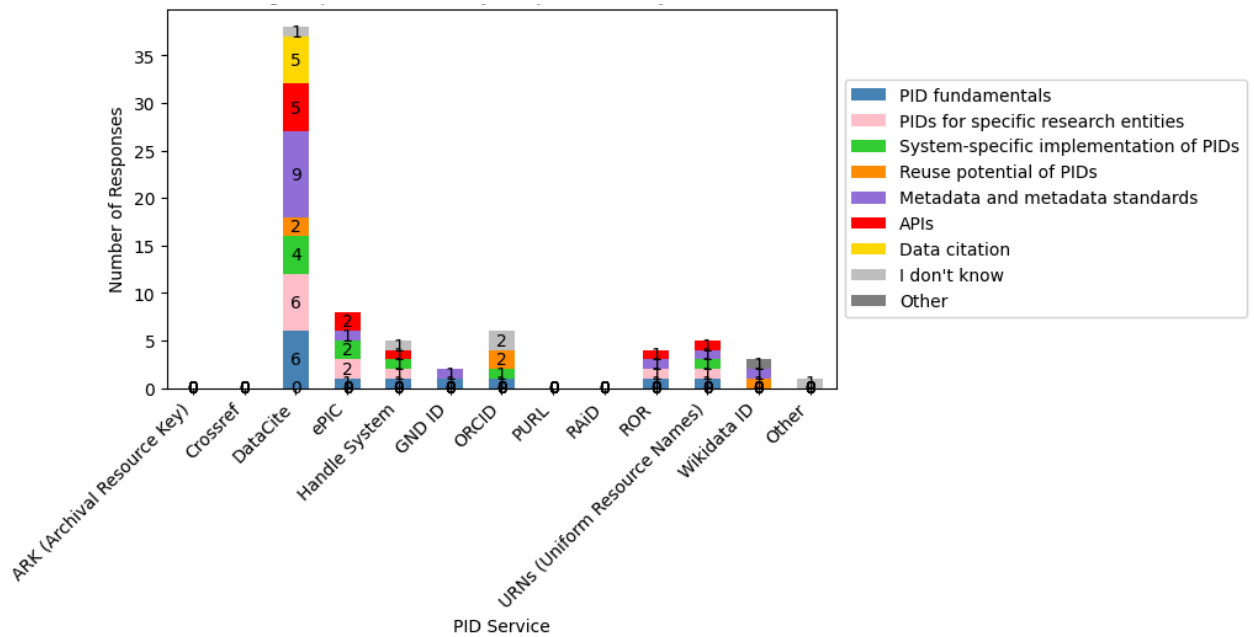
**Figure S15** offers a closer look at the types of training sessions attended, the topics covered, and the specific needs for training from each PID provider. Interestingly, DataCite stands out as a key provider of training on metadata and standards, making it well-positioned to improve metadata quality and promote harmonization across the NFDI. This aligns with the widespread use of the DataCite Metadata Schema within the NFDI community as seen in the Interoperability and Metadata Alignment Needs section.

## Landscape Analysis of PID Practices in NFDI

a)



b)





c)

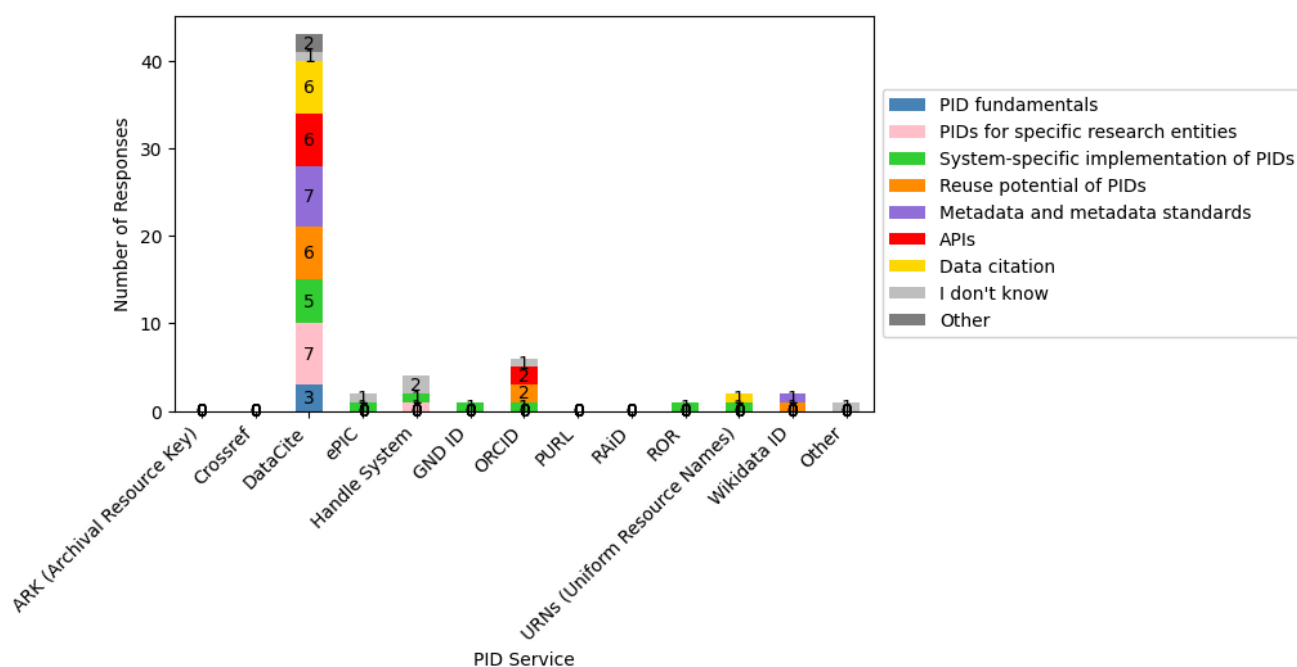


Figure S15 a) Training formats for each service provider. b) Training topics attended stratified by service providers. c) Desired training topics stratified by service providers. For the aggregated images, see also **Figures 25-27**.

## 7.8. PID Management Support

The following figures provide insights into how respondents sought support for managing or implementing PIDs. A majority of respondents (59%, or 20 out of 34) did not seek external support (**Figure S16**). This trend, when considered alongside responses to challenges (**Figure 6**), appears to stem more from limited resources or budgeting constraints rather than a lack of interest. Additional factors could include insufficient knowledge about where to seek support for specific issues or questions.

## Landscape Analysis of PID Practices in NFDI

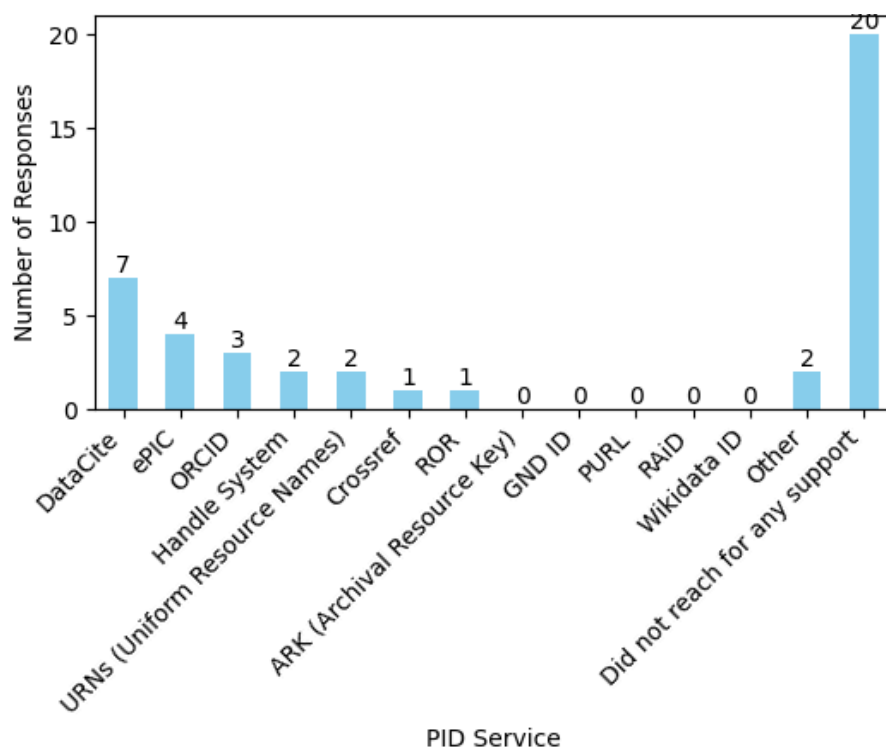


Figure S16 Responses to whether organizations have sought external support for managing or implementing PIDs, including technical help, consulting, training, or expert resources.

Among those who did seek support (41%, or 14 out of 34), most reached out occasionally, as needed, rather than on a regular basis (**Figure S17**). These 14 respondents provided 22 responses detailing the specific PID providers they consulted, as shown in Figure S16. The frequency of reaching out for support is summarized in **Figure S17**.

## Landscape Analysis of PID Practices in NFDI

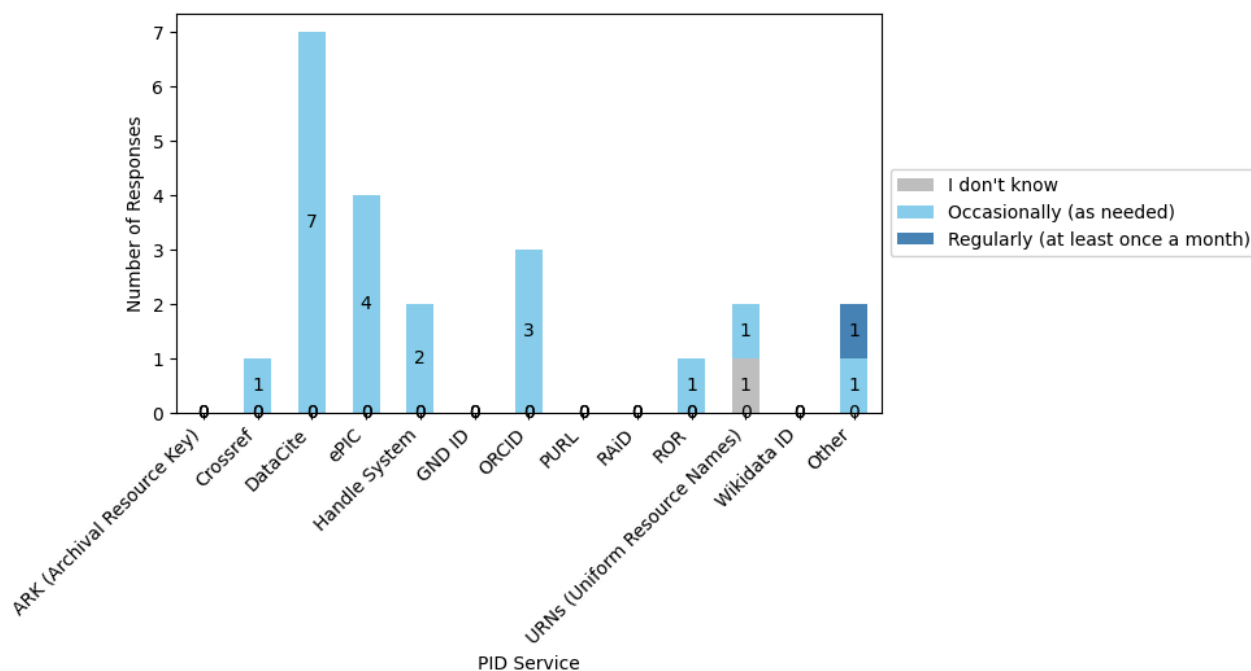


Figure S17 Frequency of reaching out for external support for implementing or managing PIDs.

The next two figures summarize the types of support services respondents required most frequently. **Figure S18** provides a general overview, while **Figure S19** delves into support requests for specific PID providers.

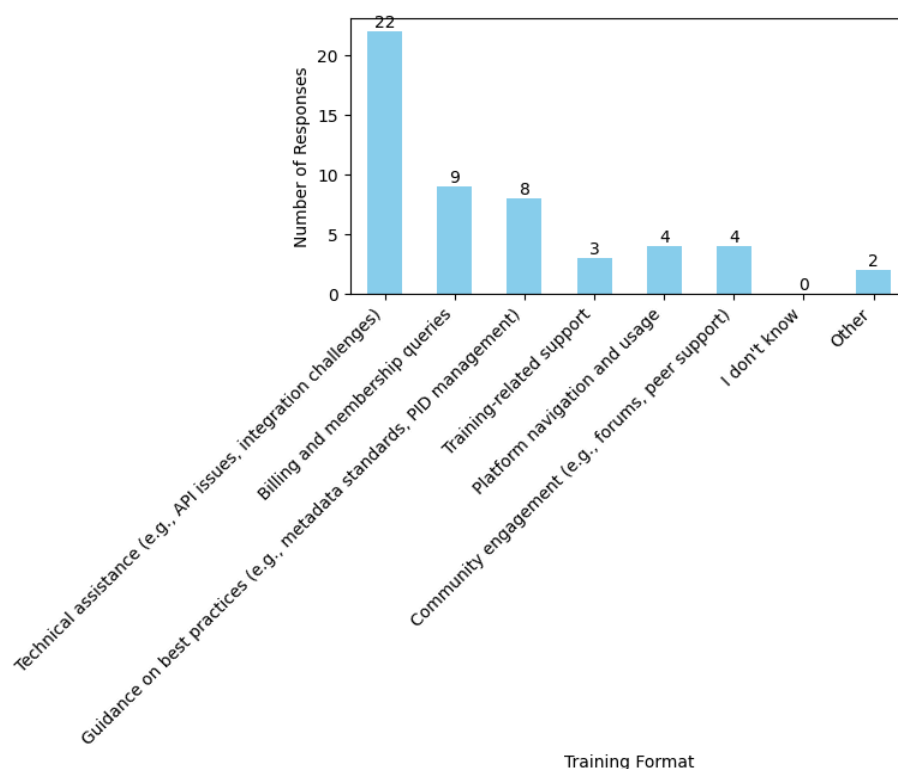


Figure S18 Types of support services frequently required by respondents for PID management.

## Landscape Analysis of PID Practices in NFDI

Technical assistance was the most frequently requested type of support, followed by billing and membership inquiries, as well as guidance on best practices (**Figure S18**). This trend was consistent across all PID providers (**Figure S19**).

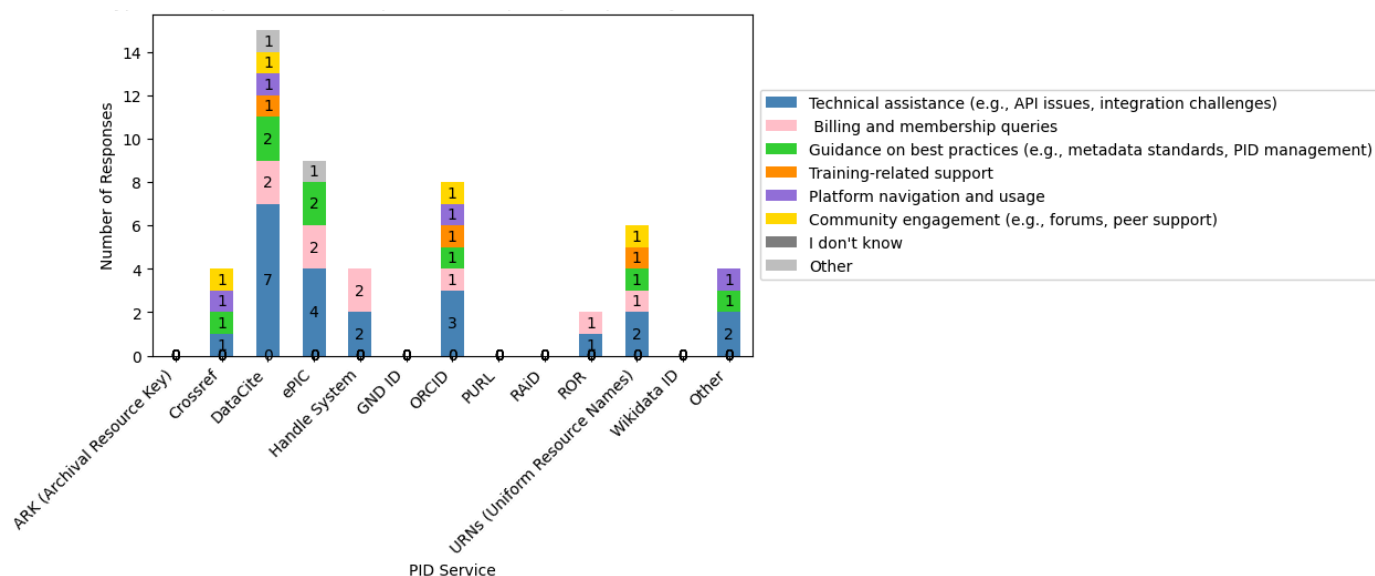


Figure S19 Types of support services required from specific PID providers.