

# RDA Professionalising Data Stewardship – Current Models of Data Stewardship: Survey Report



RDA Supporting Output

**DOI:** [10.15497/RDA00075](https://doi.org/10.15497/RDA00075)

**Authors:** Bill Ayres, Liise Lehtsalu, Graham Parton, Ádám Száldobágyi, Eleanor Warren, Angus Whyte, Niklas Zimmer, [Professionalising Data Stewardship IG](#)

**Published:** 14 December 2022

**Version:** 1.0

**Abstract:** This report presents the first analysis of the Models of Data Stewardship survey completed by the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) in October-November 2021. The purpose of the survey, based around a conceptual model of data stewardship services, was to engage with the global research data management (RDM) community to capture information about data stewardship approaches taken by organisations and the wider landscape of data stewardship and services. In particular, this initial report provides insights on a) roles involved in providing data stewardship services, the functions these provide, to which customer groups, and at which stages of the data lifecycle; b) stakeholder involved in the governance of data stewardship; c) indicative ranges of service capacity of data stewardship services; d) benefits of data stewardship promoted by services; and e) main lessons learned by responding data stewardship services.

**Impact:** The ‘RDA Professionalising Data Stewardship – Models of Data Stewardship Initial Report’ output will be of value to stakeholders seeking to develop support services and competence centres serving research producers and users of FAIR data and Open Science. The survey findings are particularly relevant to individual data stewards as well as coordinated data stewardship service provision at all levels, including institutions, national, regional, and domain-based services. The output offers insights to current data stewardship service definition and provision across the globe, but in Europe in particular, and facilitates learning from the benefits promoted, the barriers experienced, and the main lessons learned by existing data stewardship services. Alongside the publicly available survey dataset this report therefore provides a useful evidence base to further aid the professionalisation and delivery of data stewardship across the research data management community.

**Language:** English

**License:** [Attribution 4.0 International \(CC BY 4.0\)](#)

**RDA webpage:** <https://www.rd-alliance.org/group/professionalising-data-stewardship-ig/outcomes/rda-professionalising-data-stewardship-models>

**Citation and Download:** Ayres, B., Lehtsalu, L., Parton, G., Ádám Száldobágyi, Warren, E., Whyte, A., & Zimmer, N. (2022). *RDA Professionalising Data Stewardship - Models of Data Stewardship Survey Initial Report*. Research Data Alliance. <https://doi.org/10.15497/RDA00075>

## Foreword

The *Models of Data Stewardship* survey was carried out by the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) in October-November 2021. The purpose of the survey was to engage with the global research data management (RDM) community to capture information about the approaches that organisations are taking to data stewardship, and produce a first synthesis of these approaches. This survey report presents the first analysis of the *Models of Data Stewardship* survey data.

In keeping with the process for finalising outputs of RDA Interest Groups, the final draft report was open for community review. The comments received included helpful feedback about the structure of the report. As a result we have moved the findings to the beginning of this report, immediately after this foreword.

In parallel, the authors invited feedback from a number of individuals they knew to have a keen interest in the professionalisation of data stewardship. We include their comments below.

*"Setting out the ecosystem precisely, the report quantifies emerging data stewardship services devolved from global research infrastructures to institutional and domain-based services. Despite limited geographic participation in the study, the findings provide valuable indicators to nascent services in developing countries on successful approaches to engage with their organisations and user communities. Sharing the lessons learned, this report is a tactical roadmap for professionalizing data stewardship services even in less well-resourced research environments."*

**Dr Dale Peters**

*"I read the report with great interest. I was particularly surprised to read that the 'Lack of incentives/rewards for data stewardship' is still considered as the most significant barrier to the adoption of good data stewardship practices. This is an important and urgent message to take into account when addressing changes in recognition and rewards structures in academia."*

**Marta Teperek**, TU Delft and 4TU.ResearchData, the Netherlands

*"The Models of Data Stewardship Report is urgently needed. Key gaps and areas to support in the developing open research data space are well described and in line with our own State of Open Data survey, particularly with regards to funder guidance and financial support for open data. Whilst the space is progressing, this report further confirms the need for more action from funders. It is shocking to see responses of volunteer labour to support this critical part of the research lifecycle. Whilst European skewed, it is encouraging to see responses from 6 continents, echoing our findings that this is a global research priority."*

**Mark Hahnel**, figshare

*"This valuable analysis of data stewardship services shows that the majority is still domain-agnostic and delivered at institutional level. In the Netherlands, three*

*national Thematic (i.e. domain-specific) Digital Competence Centers (TDCCs) are currently being established. The report's insights will help prioritise our TDCC activities to strengthen data stewardship capacity, roles and services. In the European context, the report will also be of great value for the EOSC Association data stewardship task force that is currently collecting implementation examples of data stewardship."*

**Fieke Schoots, Mijke Jetten, Celia van Gelder**, Health-RI and Dutch Techcentre for Life Sciences (DTL), the Netherlands

*"The Models for Data Stewardship Survey report is very relevant for us in Skills4EOSC. The three-year project Skills4EOSC (Skills for the European Open Science commons: creating a training ecosystem for Open and FAIR science; <https://www.skills4eosc.eu/>) started in September 2021 and is coordinated by GARR. It will provide solutions to fill most of the gaps identified by the RDA Report, by designing the Minimum Viable Skill sets (MVS) for each profile involved in Open Science - including data stewards, researchers and policy-makers - to ensure alignment, uniformity, and quality and recognition of the competences across Europe and beyond. Moreover, it will set up a methodology for creating learning materials that are FAIRby-design and a quality assurance and certification framework, that will consider the lifecycle of the materials so that training has clear learning outcomes that match the appropriate MVS for each professional profile. Skills4EOSC will embrace a co-creation process to set up, refine and disseminate its results and it will collaborate strictly with the RDA Stewardship Interest Group to define and align activities on job profiles and training for data stewards, career tracks, certification and exchange networking and knowledge."*

**Sara Di Giorgio**, Horizon Europe Skills4EOSC project and GARR, Italy

The *Models of Data Stewardship* survey findings offers a snapshot of current data stewardship service provision, providing insights on

- Roles involved in providing services, staffing levels, the functions these provide, to which customer groups, and at which stages of the data lifecycle.
- Stakeholders involved in the governance of data.
- Indicative ranges of capacity provided in terms of numbers of DMPs, projects, and users supported.
- Benefits promoted, and factors that respondents believe helpful towards realising these benefits. These include factors likely to influence engagement with research data producers, factors leading to the adoption of a data stewardship service, as well as overall success factors.
- Lessons learned that respondents would like to share with others seeking to establish or develop data stewardship further in their organisation.

## Key Findings of the Survey

The *Models of Data Stewardship* survey was completed in October-November 2021. The survey received 136 full or partially complete responses. Our analysis, which is summarised in the [Results](#) section of the report, produced the following key findings:

- The majority of responses (76%) to the survey came from European services, with North-American services submitting 14% of the responses; Africa, South America and Asia-Pacific responses contributed just 10% of the total response to the survey.
- In terms of the institutional type of respondents, the majority of responses came from institutional level services that are multi-domain or generic/domain agnostic.
- Most respondents indicated that their service included 1-5 FTE staff.
- The responding data stewardship services were based on various funding models. While half of the respondents indicated that their service is funded from one funding source, others indicated that they combine various funding sources. The most frequently mentioned funding sources include 1) recurring dedicated budget, 2) direct costs to research grants or projects, and 3) fixed-term budget for a commissioned service.
- The respondents identified numerous roles and job titles they associate with data stewardship. In addition to research and 'core' data stewardship roles (e.g. 'research data manager', 'data librarian/archivist', and 'research software engineer'), they also identified roles offering professional advice on specific aspects of data management (e.g. GDPR, ethics, metadata) and roles involved in coordinating data stewardship (e.g. project manager, service manager, lab coordinator). The large variability of roles and job titles mentioned suggests that organisational context and structures have a significant impact on how data stewardship services are organised and implemented and that data stewardship services are provided in a variety of contexts.
- The three most frequently identified factors that supported the successful initiation of data stewardship services included: 1) availability of training services or networks for building key skills and knowledge; 2) availability of institutional policies related to data management; and 3) senior management recognising the value in data stewardship.
- Respondents identified the host organisation and an advisory/steering group as the stakeholders most frequently involved in the governance of the respondents' data stewardship service.
- The most frequently named user groups for data stewardship services included 1) funded producers of data, 2) affiliated researchers, 3) unfunded producers of data, and 4) students.
- The data stewardship actions the responding services most frequently provide to their communities include 1) guidance for selecting data, services and tools for data management, 2) supporting data management planning, and 3) ensuring good practice in versioning, curation and archiving. The least frequently provided data stewardship function was developing a sustainable business model for a trustworthy service.
- Data stewardship services do not engage with users equally throughout the research data life cycle. Engagement with users occurs most frequently in the 1) 'share and publish', 2) 'manage, store and preserve', and 3) 'discover, reuse and cite' stages of the research data life cycle.
- The respondents promote various benefits from data stewardship services to their users. The benefits promoted by more than half of the respondents include: 1) resource use efficiency, 2) research visibility/ citation, 3) sustainable long-term digital preservation, 4) research integrity, transparency and reproducibility, 5) contribution to data and metadata quality, and 6) compliance with funder policy. The least frequently promoted benefits included enhanced value from data linking and efficient risk management.
- Just over half of the respondents described the level of engagement their service has with data stewardship practices, policies, and standards relevant to the communities they serve as a collaboration. Collaboration was defined such that the data stewardship

service also engages with the design, development, and review of data stewardship practice in the communities it serves and consults and collaborates widely, potentially also taking a community coordination and leadership role.

- When asked to identify factors influencing engagement with research data producers, respondents pointed to three main aspects of service provision; policy compliance, resourcing, and the scope of the support offered. Coupled with these were several aspects relating to service delivery and response. These were visibility, outreach, trust, and embedding in research practice.
- Respondents rated the services they offered to be mostly at earlier stages of maturity (rated as at an 'initial' or 'managed' level). The service area that received the assessment of 'defined' from most respondents was 'supporting data management planning, e.g. DMP creation and review, data curation at project end point'. In national to higher-level services, some respondents also considered the service areas 'ensuring good practice in versioning, curation and archiving', 'guidance about selecting data, services and tools for data management', and 'defining the policy and research data governance environment' as 'defined'.
- Respondents identified the following factors most frequently as factors that might increase the adoption of data stewardship services by users:
  - Exemplars from communities with well-developed data stewardship practices;
  - Uptake of tools/services that help produce outputs based on good data stewardship practice;
  - Awareness of and ready access to skills and knowledge for data stewardship;
  - Training to support the development of key skills and knowledge for data stewardship;
  - Senior management see enough value in data stewardship to promote and resource it;
  - Requirements from publishers and funders for the availability of data outputs.'Demand from funders' was the factor least frequently identified by respondents using a given list of factors.
- Respondents offered several approaches that their services have used to engage with research data producers. The methods and approaches can be categorised as support to data production, skills development, recognition of good practice, data tools provision and effective communication.
- 'Lack of incentives/rewards for data stewardship compared with journal publications', 'difficulty changing stakeholders' attitudes or practices' and 'lack of buy-in from senior managers or researchers' were indicated by respondents as the three most frequent barriers to data stewardship service.
- The main lessons respondents have learned from setting up their data stewardship service can be summarised in the following six points:
  - Build on concrete benefits to service users and stakeholders
  - Embed data stewardship in teaching and training
  - Engage with external stakeholders
  - Coordinate across the organisation
  - Communicate using effective methods
  - Drive and monitor improvements

## Intended Audience and Usage

We report survey findings that will offer substantial input to stakeholders seeking to develop support through professional networks and competence centres serving research producers and users of FAIR data and other products of Open Science. The survey findings are relevant to individual data stewards, and to coordinated service providers at all levels, including institutional, regional, national and domain-based services. They will be especially relevant to capacity-building projects and task forces developing Open Science Clouds, e.g. the Global Open Science Cloud, European Open Science Cloud, the proposed African Open Science Cloud, Australian Research Data Commons and related initiatives.

## Interconnections with RDA PDS Interest Group

This report forms part of a wider corpus of work being undertaken by the Research Data Alliance's (RDA) Professionalising Data Stewardship (PDS) Interest Group (IG), in short, the RDA PDS IG. It supports the wider IG efforts in the following specific ways:

- Business Case Task Group - providing an evidence base and conceptual model for the work of this task group as well as enabling specific use-cases to be identified for later exploration. Levels of service maturity may also assist in the development of the group's work.
- Terminology Task Group - the range of data stewardship role titles and their mapping to associated functions (Section 2) as well as the base concepts within the conceptual model provide content and base framework for activities seeking to establish a common set of terms to use within this sphere. Additionally, issues around the survey reach may be informative to these discussions about terminology.
- Career Paths Task Group - the role job titles and associated functions covered in Section 2 may assist the Careers Paths Task Group in distilling responses from their career paths survey due to close in October 2022.
- Training Task Group - The range of role types depicted within Section 2 may be useful in helping to identify the range of roles and therefore skills required for successful service operation. These should therefore indicate the nature of training curricula that need to be established for a professionalised approach moving to such provision.

## Data Availability Statement

The data that supports the findings of this report are openly available in Zenodo at <https://doi.org/10.5281/zenodo.6665306>. The survey text used to collect the data is also openly available in Zenodo at <https://doi.org/10.5281/zenodo.6665146>.

## Acknowledgements

Iryna Kuchma (<https://orcid.org/0000-0002-2064-3439>) from EIFL contributed to the initial definition of the survey questions. In-kind support to the survey was contributed from the project FAIRsFAIR “Fostering FAIR Data Practices In Europe”, which received funding from the European Union’s Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. Other contributions were made with in-kind support from individual contributors’ home organisations.

## Introduction

This report provides a brief outline of the *Models of Data Stewardship* survey undertaken by the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) in October-November 2021. The purpose of this survey was to engage with the global research data management (RDM) community to capture information about the approaches that organisations are taking to data stewardship, and produce the first synthesis of these approaches.

## Background

### Aims

The survey was produced with the aim of gaining insight into how services deploy staff roles and resources to achieve the goal(s) of data stewardship in a way that meets the needs of stakeholders and the targeted research communities' requirements. This was in response to the aims of the Model's Task group within the RDA-PDS Interest Group, which states:

*"We aim to model the **Data Stewardship Roles** within particular **service provisions of Data Stewardship**, the larger **Organisational Contexts** within which such provisions exist, and how those roles and services interact with their **Target Communities**."*

## Methodology

The PDS-IG Models Task Group approached this task by first considering what we understood by the term data stewardship, drawing on previous definitions such as:

'The responsible planning and executing of all actions on (digital) data before, during and after a research project, with the aim of optimising the usability, reusability and reproducibility of the resulting data.' \*

\* *About Research Data Management* - Dutch Techcentre for Life Sciences (dtls.nl)

To scope the survey questions we initially applied a 'Business Model Canvas' (BMC) perspective, aiming to capture basic elements of the models used by organisations that provide any form of support service for data stewardship. An early draft of the survey was based on a version of the BMC adapted for research data management contexts by the



Digital Curation Centre.<sup>1</sup> This provided a basic set of concepts applicable at the level of an individual organisation, as follows:

- *Beneficiaries*, e.g. service users
- *Proposed benefits*, e.g. more trustworthy stewardship
- *Key capabilities*, e.g. data discovery
- *Key partnerships*, e.g. research office, research funders
- *Key resistance*, e.g. senior management buy-in
- *Key resources*, e.g. staff knowledge and skills
- *Cost structure*, e.g. activity-based costing
- *Funding streams*, e.g. core funding, project funding, subscription
- *Metrics*, e.g. user metrics, quantitative impact figures
- *Delivery channels*, e.g. research grant support
- *Beneficiary relationships*, e.g. researcher affiliation, membership

These basic concepts were then reformulated to propose a broader generic model intended to express the key variables we expected to find across the operational models/approaches to be surveyed at an organisational level.

This broader model, given in the figure below, has the following key principles (**bold** items denote classes of information whilst *italics* capture relationships between classes of information):

- **Service** - this details the central aspect of this model. It may be a standalone service or (as indicated by the '*has Context*' relationship) it may be found within another service (e.g. a domain data centre that is federated into a wider network of domain repositories).
- The **Service** will be *servicing* the requirements of one or more **Stakeholders**.
- This **Service** undertakes or *performs* one or more **Functions** to deliver the service, each of which will *require* one or more **Resources**.
- Underpinning the delivery of the functions of the service one or more **Data Stewards** *undertake* one or more **Roles** to perform the **Functions**.
- The Service itself will also have a set of **Drivers** and ought to *follow* one or more **Standards**.

---

<sup>1</sup> Whyte, A. 'Business Model Canvas for RDM Services' (2021) Digital Curation Centre available at: <https://docs.google.com/document/d/1TEpx7oLNh-ErAtG6Kp8cWRgk6kSz2wJo/edit>

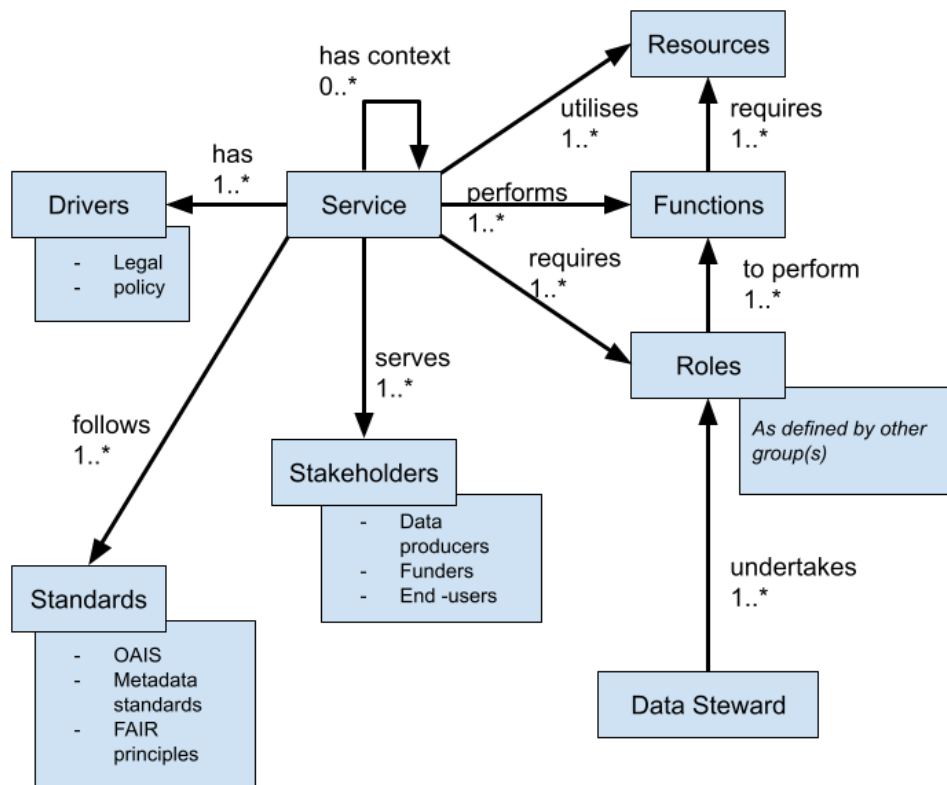


Figure 1. RDA PDS conceptual diagram for models of data stewardship.

This model informed further iterations of the survey, of which an online copy is available to view at: <https://doi.org/10.5281/zenodo.6665146><sup>2</sup>.

The survey did not cover the standards and drivers components of the conceptual model as these were understood from the experience of the survey team members to either be too specific, and thus not transferable to other cases, or already well understood within the general RDM landscape and thus not needing to be captured.

Additional elements beyond the conceptual diagram were also explored allowing the experiences of services to be captured as well as further profiling questions covering levels of engagement, service maturity and size.

This survey had three main sections:

1. Section 1 - basic info about the 'service'/data stewardship provision. This primarily sought to establish the geographical context of the service being described as well as the nature of the service in terms of two principal axes: the level of domain-specificity (from very subdomain-specific to generic) and organisation level (project level through to global/pan-national service).
2. Section 2 - data stewardship roles/functions within the respondents' own context. Capturing the working parts of the service to identify models of operation.

<sup>2</sup> Ayres, Bill, Lehtsalu, Liise, Kuchma, Iryna, Parton, Graham, Száldobágyi, Ádám, Warren, Eleanor, Whyte, Angus, & Zimmer, Niklas. (2022). RDA Data Stewardship Organisational Models Survey Final Offline Version. Zenodo. <https://doi.org/10.5281/zenodo.6665146>

- Section 3 - capturing the wider context, to allow the content of the service to be understood (e.g. key drivers).

A mapping between the questions and areas of the conceptual model is given in Figure 2

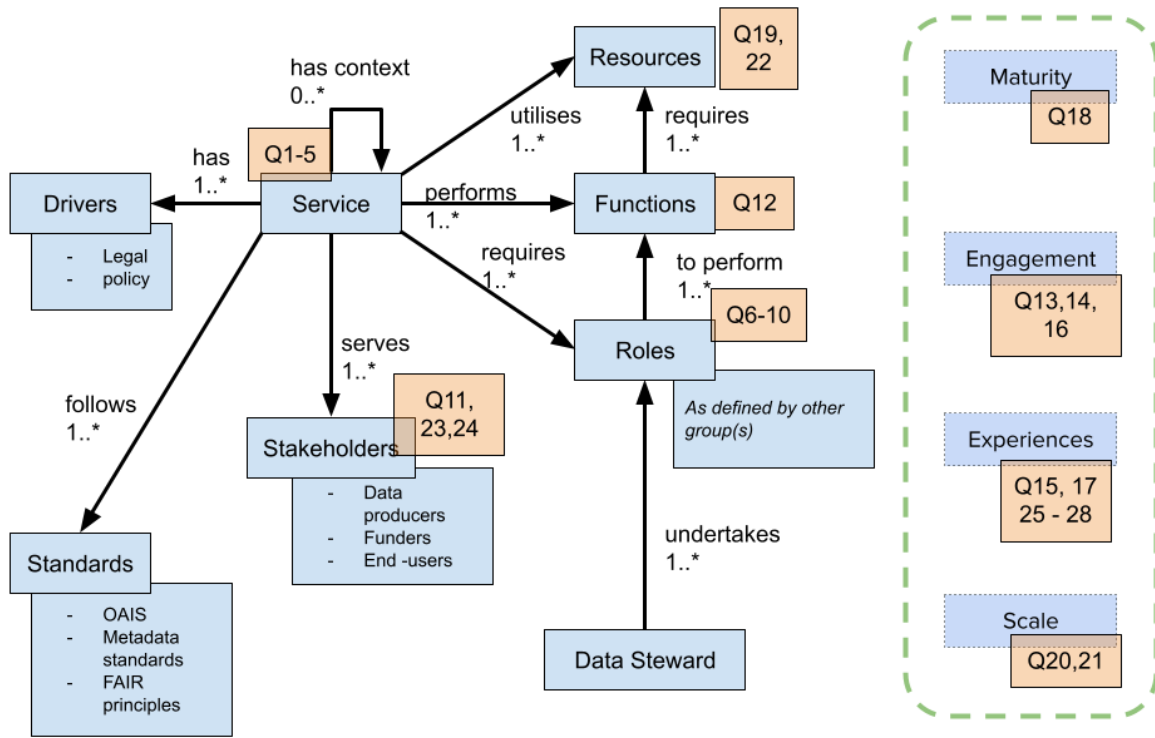


Figure 2. RDA PDS conceptual diagram for models of data stewardship and additional areas covered by the survey (green dotted area) with survey question mapping (orange boxes).

Through a combination of all these factors, it was hoped to then establish if there were noteworthy commonalities or differences between instances of data stewardship across the research data stewardship landscape. To help identify groupings of such instances following the survey, the team drew on two principal axes which were used in earlier discussions exploring the data stewardship landscape: domain specificity (i.e. how 'targeted' a service is to a given domain or if more generic), as well as the organisational level at which the service was constituted. Earlier work by the team had identified sample cases and heard from a number of instances at the 17th RDA plenary, as shown in Figure 2. Examples included those from well-known global services such as Zenodo and Figshare, through to institutional, domain-specific and project-focused repository services.

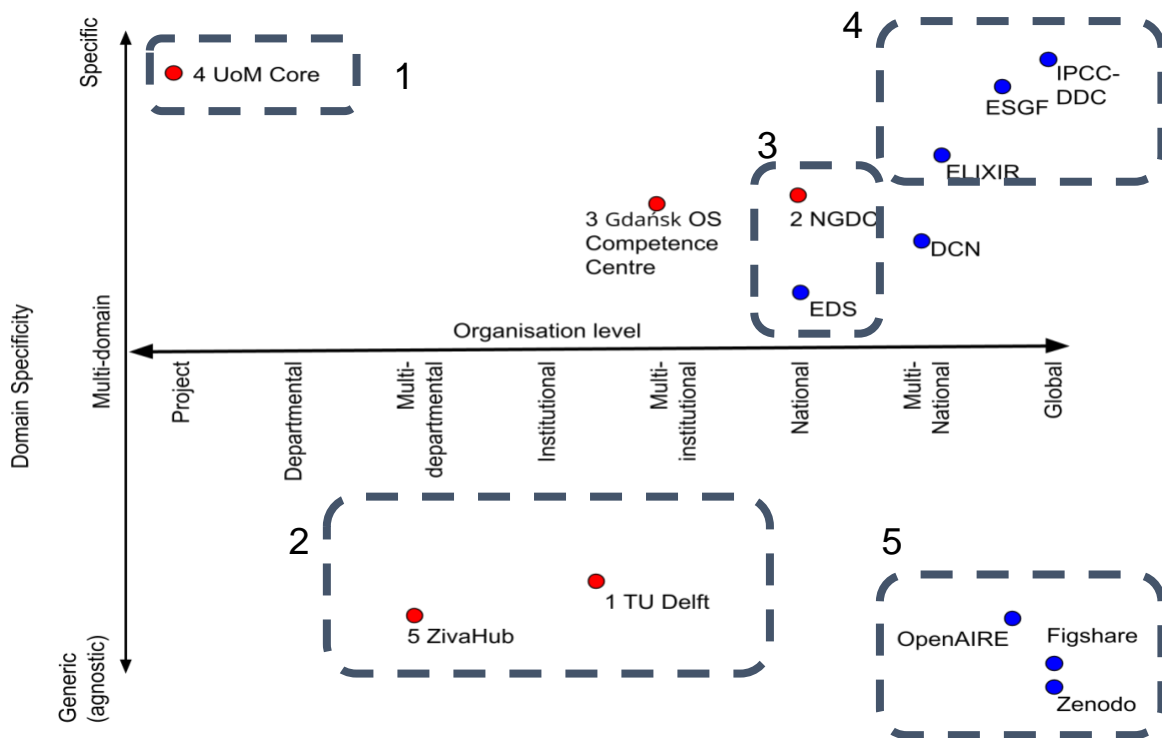


Figure 3. Domain specificity vs organisation level with examples of services presented at the 17th RDA plenary. Red indicates services invited talks at the plenary as exemplars of the landscape. Blue indicates previously identified examples to establish the veracity of using these axes to profile the data stewardship service landscape.

Having these example cases to hand aided the survey team to ensure that the survey remained broadly relevant across the landscape. Furthermore, this initial landscape scan, alongside the prior experiences of the task group members, helped to identify anticipated groupings of services and data stewardship roles - as labelled in Figure 2 - namely:

1. Small-scale, bespoke services - data stewardship has the opportunity to be closely engaged with data producers and reusers but is ad-hoc in terms of integration with organisational processes or wider community standards.
2. Institution-wide (generic) services - data stewardship role is broad (perhaps to the point of being overstretched) and with varying levels of stakeholder interaction and community engagement.
3. National domain repositories - with a defined disciplinary focus and varying degrees of sustained investment by stakeholders in the discipline.
4. Pan-national research data infrastructures - with a defined disciplinary scope and service portfolio, and varying degrees of investment by international stakeholders in the discipline; technically heavy and with a narrow scope.
5. Pan-national generic services - with a trans-disciplinary scope and service offering, varying degrees of investment by international stakeholders; data stewardship is largely left to the data provider to self-manage.

## Target population

The survey responses were from a self-selected sample, as a link to the online questionnaire was widely distributed via the PDS-IG group and lists that members were familiar with. The target population was loosely defined to include anyone in the research data community involved in 'providing data stewardship services'.

With further development of data stewardship as a profession, it may in time, be possible to accurately target a sample. The group took the view that it is not currently possible, as these roles are still emerging, not formally professionally recognised, and the services provided still rapidly evolving. We did, however, attempt to gauge some of the parameters involved in this.

We assumed that research data stewardship *mostly* happens in research universities, institutes, and repository services. It is relatively straightforward to quantify these from published sources, e.g. (respectively) from Leiden University Rankings,<sup>3</sup> Wikipedia,<sup>4</sup> and the Re3data directory.<sup>5</sup> However, it is guesswork to estimate the numbers of organisations in each category that currently offer any data stewardship capabilities or the number of staff directly involved in providing these. Our rough and probably conservative estimate was below 2,000 individuals (globally).<sup>6</sup>

The purpose of the survey was partly to shed further light on the kinds of roles involved in data stewardship in the view of our self-selected sample. The responses should help more effective targeting in future, and we return to this point in our analysis.

## Analysis

The Qualtrics XM survey platform, provided by the University of Manchester, was used to collect responses. The survey was accessed 261 times and received 136 full or partially complete responses that also provided consent to be included in the study. The partially completed responses varied in their level of progress made through the survey. Of the 136 responses, 82 provided answers beyond the first section, while 53 answered the first section only.

For the analysis, we retained all 136 responses. These responses were anonymised by stripping them of directly and indirectly identifiable information (e.g. respondent names, organisation names, revealing titles, references to specific individuals, events and outputs). The anonymised dataset was shared with the RDA-PDS IG models group survey team (i.e. the authors of this report) for analysis. For quantitative analysis, the team made use of a range of tools. Initial data preparation was undertaken within Google Sheets to further clean data to code clearly missing values and standardise text in qualitative responses. Quantitative analysis was completed using IBM SPSS statistics (version 28.0.1.0) and

---

<sup>3</sup> Leiden University CWTS (2021) <https://www.leidenranking.com/>

<sup>4</sup> [https://en.wikipedia.org/wiki/Category:Research\\_institutes\\_by\\_country](https://en.wikipedia.org/wiki/Category:Research_institutes_by_country)

<sup>5</sup> <https://www.re3data.org>

<sup>6</sup>

<https://docs.google.com/spreadsheets/d/1iXGjS8RcWlyExOq6w2ZZMYuHvEbt3zpc/edit#gid=959198800>

Microsoft Excel (version 16.61.1), as well as a combination of a Python Notebook using the Python Pandas library and Google Sheets. A common colour palette (Tableau 10 from [https://jrnold.github.io/ggthemes/reference/tableau\\_color\\_pal.html](https://jrnold.github.io/ggthemes/reference/tableau_color_pal.html)) was adopted for the resulting graphs to aid consistency in presentation and accessibility. For qualitative analysis, inductive thematic analysis was performed.

In the rest of this report, the term ‘**all responses**’ is used to indicate where answers from the full sample of 136 responses were used as fully as possible, whilst ‘**complete responses**’ indicates the use of the 82 responses that provided content in all sections of the survey (though some answers may have been left blank).

Moreover, where specific cohorts of answers are given in the following analysis, these will be indicated clearly. To aid this, the responses to the question about the organisational level at which the service operates were analysed to determine the ‘highest’ and ‘lowest’ organisational levels that were selected from the multiple choice responses to accommodate where respondents selected a range of organisational levels above which their indicated services operated.

Section 1 responses allowed profiling of the services in terms of their degree of domain specificity (i.e. were they specialised or generic services), organisation level(s) at which they operated and the types of function that the service provided. Additional details regarding geographic coverage also allowed the cohort of responses to be further profiled in terms of how representative (or not) the sample would be compared to the global population of services. **These responses, therefore, provide several caveats users of the dataset should bear in mind when seeking to draw any insights or conclusions either from this report or the associated dataset.**

The survey team undertook quantitative and qualitative analysis to give some initial insights from the responses. It is hoped that by providing open access to the cleaned dataset from the survey and this report, we may offer a valuable resource to be further analysed by others.

The qualitative analysis consisted of two group members independently carrying out open coding of textual responses to open-ended questions. Each member then reviewed the other’s coding, and a consensus agreed on how to group these in a simple hierarchical schema. Selected results are included in this report, and the anonymised coded responses are available.

## Dataset

The raw survey responses were held securely by members of the core survey team at the University of Manchester following university regulations. The data were cleaned and anonymised by core team members. The cleaned, anonymised data retain the original ‘response IDs’ to allow the team to track responses throughout the cleaning and anonymisation process and to allow checking back to the raw responses subsequently if required.

The resulting data have been made available alongside the output of a thematic textual analysis file produced by two members of the team (see details elsewhere on this). They are available as a community resource under a Creative Commons By Attribution 4.0 International licence from Zenodo:

Ayres, Bill, Lehtsalu, Liise, Kuchma, Iryna, Parton, Graham, Száldobágyi, Ádám, Warren, Eleanor, Whyte, Angus, & Zimmer, Niklas. (2022). RDA Data Stewardship Organisational Models Survey Final Offline Version. Zenodo. <https://doi.org/10.5281/zenodo.6665146>

## Results and Discussion

### Geographical coverage of the responses

**Section 1** of the survey asked two questions regarding the service’s location (Q4, Q5). The first (Table 1) allowed the respondent to select all global regions they felt their service was serving. The second question (Figures 3 and 4 below) then asked for a specific country where the service was based (i.e. where the service itself is based - as opposed to the community it serves).

Africa	Asia	Australasia	Europe	North America	South America	Responses per region selection
✓						7
	✓					1
		✓				5
			✓			96
				✓		14
					✓	1
✓	✓	✓	✓	✓	✓	6
✓			✓	✓	✓	1
	✓		✓	✓		1

✓			✓			1
				✓	✓	2
15	8	11	105	24	10	<b>Responses covering region</b>

Table 1. Global regions selected by respondents that their service is aimed towards serving. (n=135). Right-hand column shows numbers for each region-selection combination. The bottom row shows number of responses selecting each given region.

124 responses indicated that they only had a single geographic area as their main region served, compared to 6 indicating global service provision and the remaining 5 a range of regions. Europe was the most served region by far, with a greater number of responses (105) selecting Europe than all the other regions considered together (68). As will be seen below, this is in line with a large proportion of the responses to this survey coming from European-based institutional repositories.

Shifting the focus to look at where the country the service itself was based shows that whilst all continents had a representation it was by no means evenly spread:

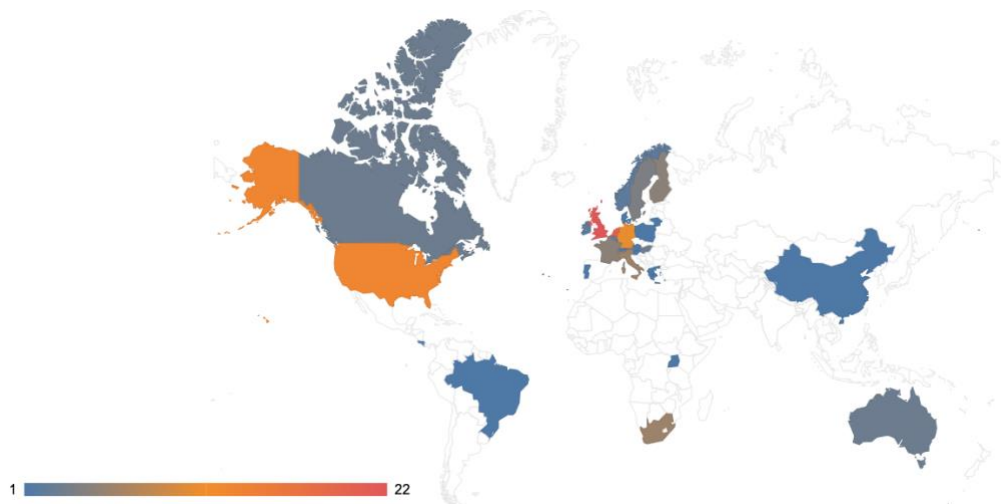


Figure 4. Distribution of responses by country selected for service location (n=120)

Examining European responses (Figure 4) the unevenness of the spread is seen with the UK (22), Netherlands (20) and Germany (11) accounting for the majority of the 91 responses from Europe. These 3 countries were also within the top 4 countries by the number of responses, USA was 3rd with 13 responses, further highlighting the sample's unevenness.



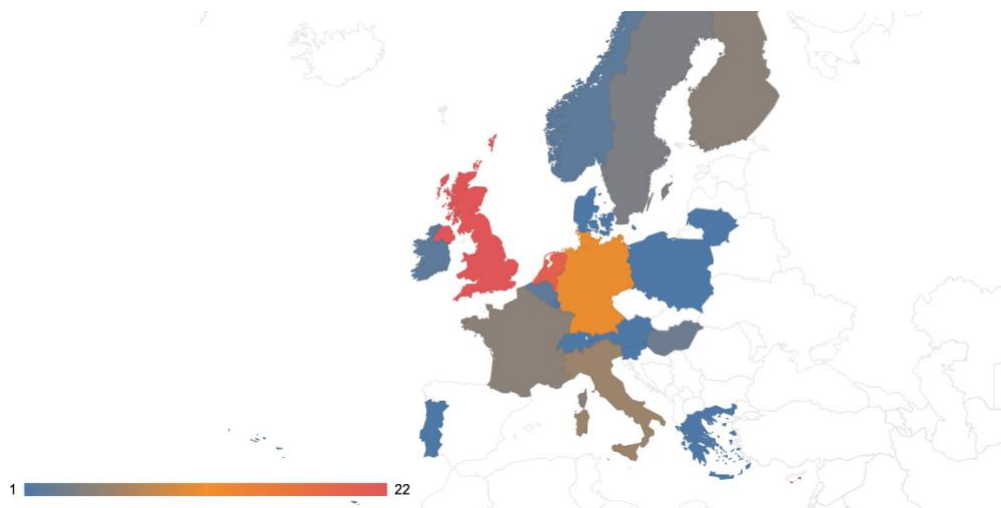


Figure 5. Distribution of responses by country selected for service location - European countries only (n=91).

(It ought to be noted that there is probably a correlation between those countries with a higher representation in the survey being similar to the nationalities of the members of the RDA Professionalising Data Stewardship Interest Group who had an active interest in the survey and engaged in promoting it - a factor that highlights the difficulties in reaching a wider audience to participate in such surveys).

Continent	Responses	Countries responded
Africa	7	2
Asia	1	1
Australasia	3	1
Europe	91	20
North America	17	3
South America	1	1
Totals	120	27

Table 2. Counts of responses of countries selected by continent and respondents by continent.

#### A note on language

During analysis, the survey team also considered if there were any noticeable language barriers to responses to the survey. In particular, if the language of the survey (English), but also the terminology used, raised barriers to (potential) respondents.

We note that 43 of the 136 responses were from just 5 English-speaking countries, while the other 77 responses came from 23 non-English speaking countries. Given that 22 of the 43 'native English' responses are from the UK alone, and that both the Netherlands and Germany were strongly present (Figure 4), we might say that the use of English for the survey may not have been too high a barrier in (at least Northern) Europe, but should remain a factor to be considered in future study design.

After opening the survey, and as the first responses were arriving, the survey team started to see an emerging dominance of European responses, despite prior attempts to reach as globally as possible. In response, the survey team discussed how to encourage greater participation from non-European respondents. This led to feedback from a number of contacts that the terminology around 'data stewardship' was likely to resonate more with particular parts of the research data management community than others - e.g. it was noted in particular by a colleague from the US that 'data steward' was not in common use in North America, in contrast with European colleagues where it is a very common term. However, the survey was not designed to explore this and thus is unable to show this empirically; yet it is a factor that ought to be explored further and, alongside language, might need to be taken into consideration in the design and communication of future studies of data stewardship.

### Domain specificity vs organisational level of data stewardship services

Section 1 of the survey aimed to capture essential information about the respondents' data stewardship services. This included information about the organisational level of the data stewardship service (Q1) and the domain specificity of the data stewardship service (Q2). The survey team established these as two key axes for service profiling during our study design (Figure 2). Figures 5 and 6 give the spread of responses given in terms of the highest and lowest organisational levels selected vs domain specificity:

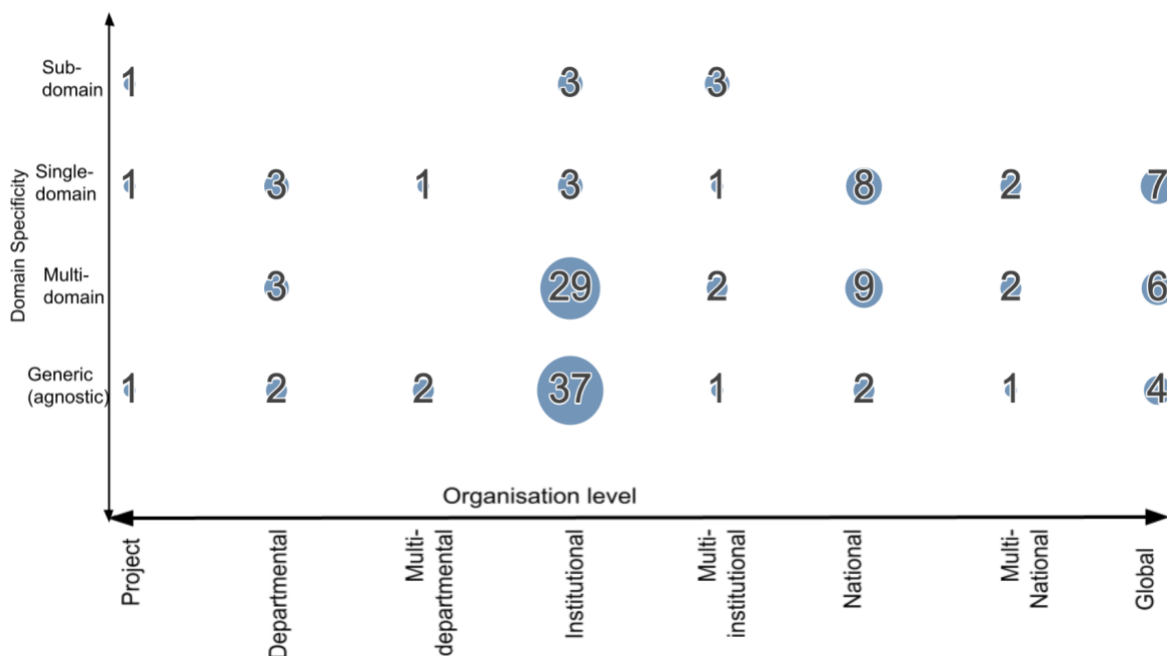


Figure 6. Highest Organisational Level vs Domain Specificity (n=134).

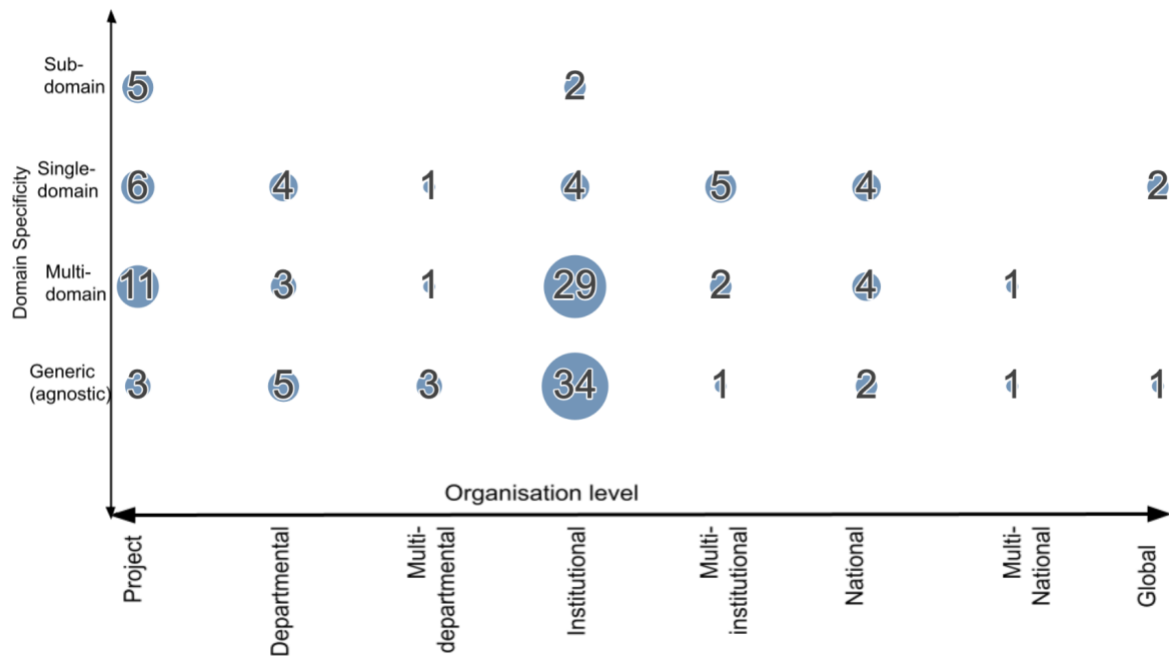


Figure 7. Lowest Organisational Level vs Domain Specificity (n=134).

From this a number of conclusions can be drawn:

- The bulk of survey responses came from institutional level services that were multi-domain or generic/domain agnostic. This is to be expected given that the number of institutes setting up their own research data management services is far larger than national/international efforts. This concurs with the anticipated target population for the survey (see above).
- Responses from national to international scale services indicated these were single or multi-domain services, with a smaller number of global, generic services.
- The low number of multi-national services may also reflect respondents selecting 'global' instead of 'multi-national' to describe their organisational level, or that such multi-national services are hard to coordinate.
- Multi-institute services responding to the survey appear to be more domain-targeted than generic.

It should be noted that whilst most responses were confined to the departmental or institutional level there were also a proportion of respondents who selected a wide spread of organisational levels. This suggests that these respondents felt their services were being used by users across broader parts of the organisational spectrum, which made interpretation of the results less clear-cut than had originally been anticipated based on the pre-survey work that led to Figure 2, i.e. modelling the broad landscape of data stewardship is hard, and boundaries are often blurred.

Overall, however, the survey received responses from across the landscape anticipated by the survey team (Figure 2). The range of responses, even acknowledging that this is not a comprehensive survey, indicates that data stewardship is a complex landscape with much variance in service provision.

## Description of service: service types

Section 1 of the survey also asked respondents to describe the scope of their data stewardship service (Q3). Respondents were asked to select from a range of service-type keywords that they would associate with their service.

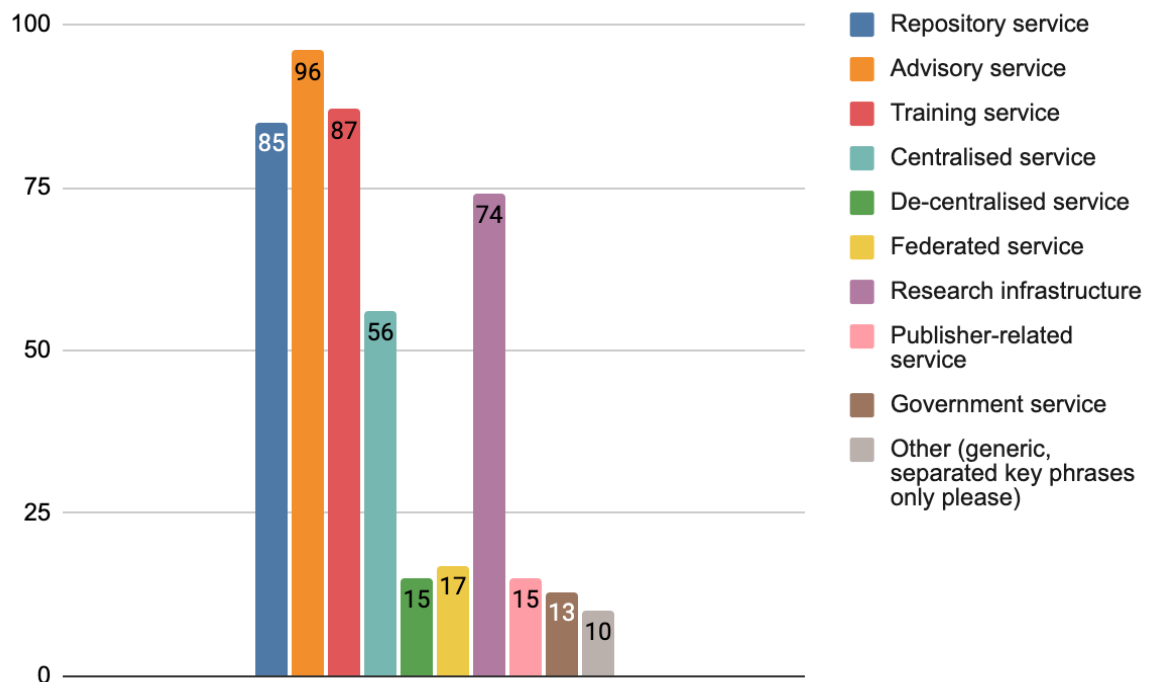


Figure 8. Service keyword selections for all responses (n=136).

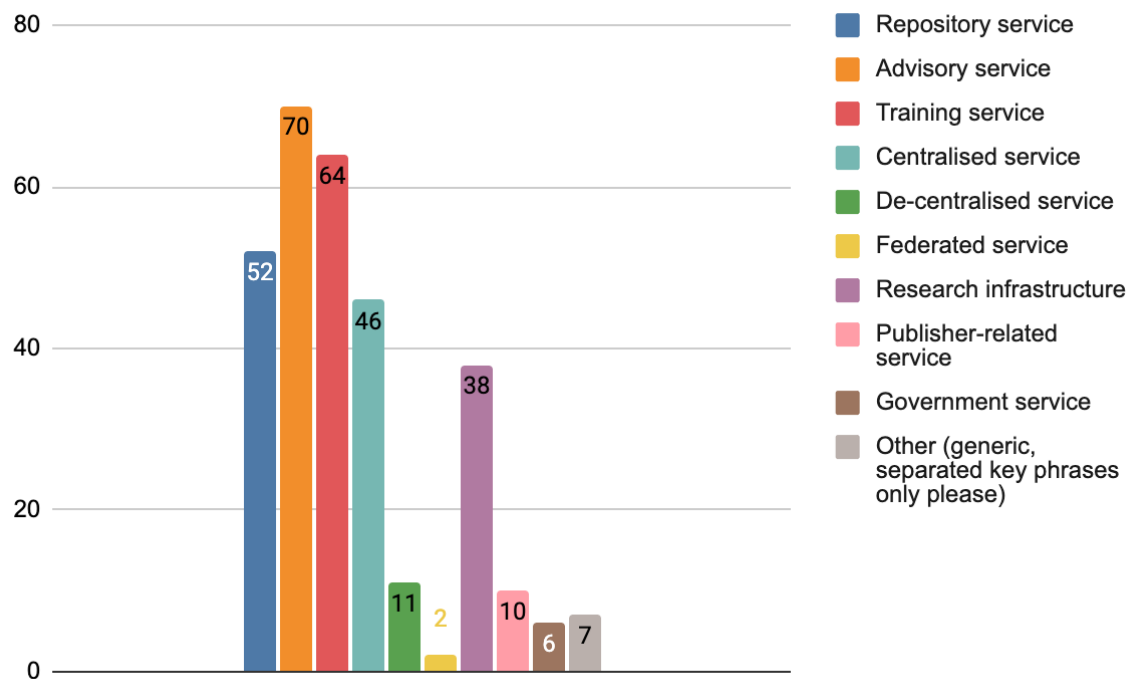


Figure 9. Service keyword selections for Institutional responses only (n=73).

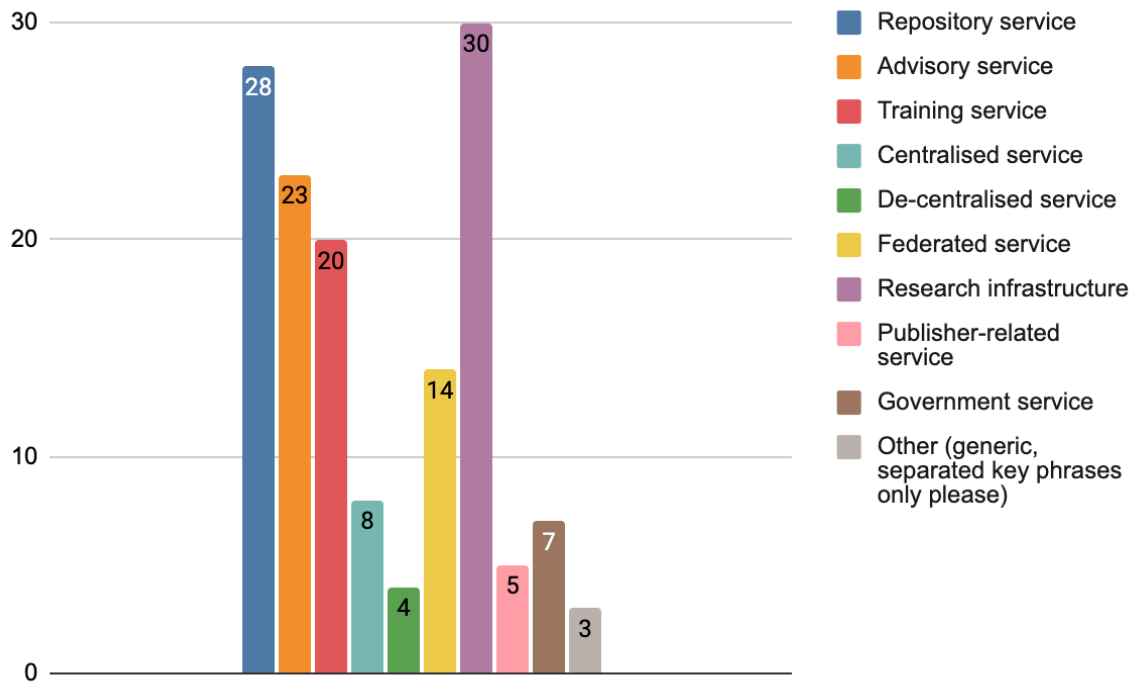


Figure 10. Service keyword selections for higher level organisation responses (n=42).

Figure 7 shows the range of keywords that were selected which were also cross-examined against domain-specificity and highest organisational level to see if there were any patterns noticeable:

- de-centralised service selection was associated with (in all but one case) generic/multi-domain services with a spread across all organisation levels
- except for 2 instances, the 17 'federated services' only occur at national to international level.
- publisher-related services were seen at all organisation levels, mainly multi-domain.
- government services were multi-domain/generic services implemented at various organisational levels

Of the 10 'Other' responses provided nothing of particular note was recorded.

## Data Stewardship Roles

**Section 2** of the survey addressed data stewardship roles, asked about the target groups of respondents' services and the services' level of engagement with their end users, the data stewardship actions performed by the respondents, the values and benefits promoted by the data stewardship services, the services' level of engagement with various aspects of data stewardship, the maturity of service provision, and resourcing of respondents' data stewardship services.

The first questions in section 2 (Q6-Q10) explored different types of roles associated with data stewardship. The survey drew on the roles identified by initiatives in the Netherlands<sup>7</sup> and Denmark<sup>8</sup>. These are closely aligned as illustrated in Figure 8.

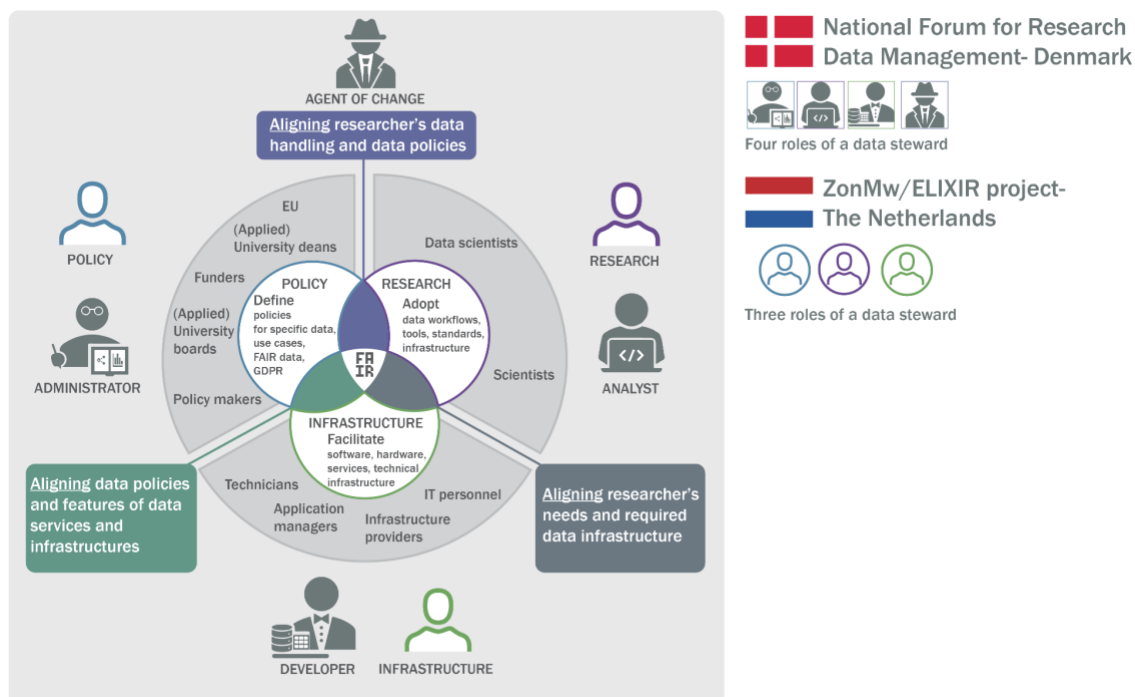


Figure 11. Emergent data steward roles from studies in the Netherlands and Denmark Image credit Frederike Schmitz. (2020) via Zenodo.

Expanding on the broad areas defined in these two national initiatives, further work by an EOSC working group on Digital skills for FAIR and Open Science also identified Data Stewards as one of a number of key emerging roles, also including Research Software Engineers, Data Curators, and Data Scientists among others. To gain further insight into the relevant role descriptions currently being used by organisations, our survey asked for job titles associated with data stewardship. Responses to Q6-Q9 are presented as word clouds to give the reader a sense of both the breadth of job titles used for each role type and also where there were recurring terms. These are from the ‘complete responses’ cohort (n=82) of responses. For each section, the first word cloud is for complete responses followed by a second that limits these responses to where the highest organisational level given was national, multi-national or global to reflect the split between institutional level (and below) services and those at these wider scales to accommodate the balance responses being dominated by institutional services.

Q6 inquired about job titles associated with “traditional” RDM support and service roles. Traditional RDM support and service activities were defined as

<sup>7</sup> Scholtens, S., Jetten, M., Böhmer, J., Staiger, Ch., Slouwerhof, I., Van der Geest, M. & Van Gelder, C.W.G.. (2019, October 3). Final report: Towards FAIR data steward as profession for the lifesciences. Report of a ZonMw funded collaborative approach built on existing expertise. Zenodo. <http://doi.org/10.5281/zenodo.3474789>

<sup>8</sup> Wildgaard, L., Vlachos, E., Nondal, L., Larsen, A. V., & Svendsen, M. (2020, January 31). National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum) (Version 1). Zenodo. <http://doi.org/10.5281/zenodo.3609516>











Figure 18. Word cloud for roles around liaising between policy stakeholders and data centres, repositories or other research data infrastructures to deliver compliant services.



Figure 19. Word cloud for roles around liaising between policy stakeholders and data centres, repositories or other research data infrastructures to deliver compliant services for national and higher organisational level responses.

A final question about roles (Q10) asked respondents to write in any other relevant roles “beyond those defined above” e. As the responses to this question were open, they were analysed qualitatively through an open coding approach. This led us to 4 main categories of job titles:

- data stewardship core roles: roles frequently mentioned in responses to this question and Q6-9,
- support coordination roles; generic managerial roles that respondents mentioned as supporting or coordinating,
- research roles: labelled as ‘research’ by respondents,
- related professional advisory roles: areas of specialist expertise that respondents identified as relevant to data stewardship.

The mapping of these categories to the types of associated job roles indicated by the respondents is shown in Figure 17.

It seems likely that survey respondents’ organisations may employ a ‘data steward’ in any one of the ‘core’ roles, to liaise with the other roles identified in Figure 17. Our analysis does not reveal the nature of relationships between roles. However, the variety of job titles suggests the organisational structures will depend on factors identified in the Dutch and Danish studies illustrated previously in Figure 11, such as the extent of coordination required

across support services, and the extent to which an organisation needs to specialise in its provision of data policy, support for analytics, and bespoke infrastructure.

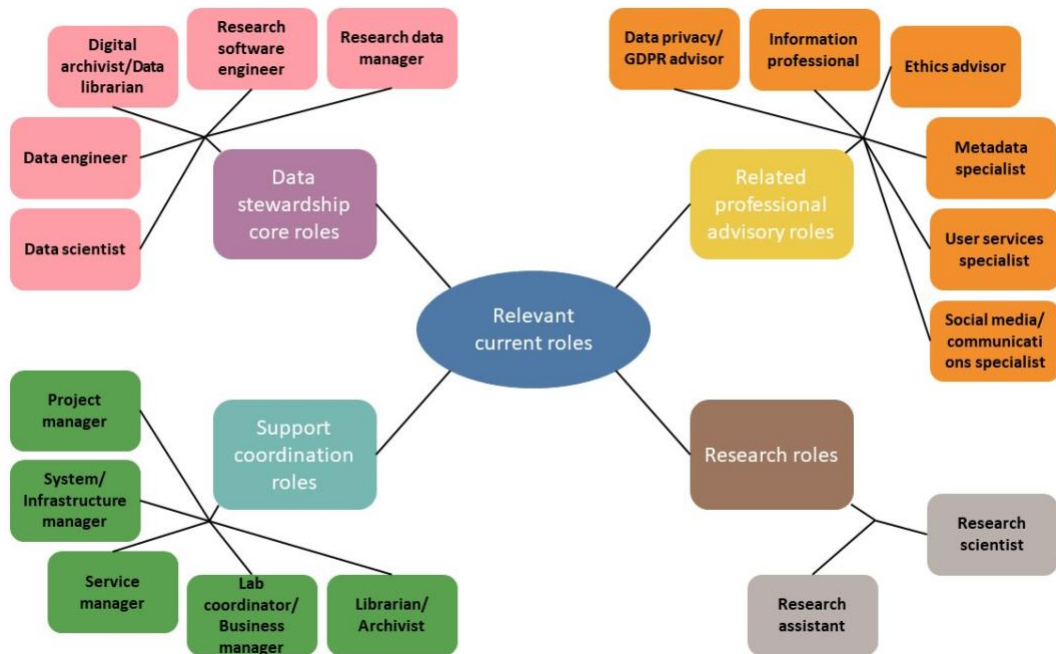


Figure 20. Graphic representation of the thematic analysis performed on responses to Q10 (n=29).

### Customer groups or service users for data stewardship services

Q11 inquired about the main user groups of the respondents' services. Respondents could choose between 9 defined user groups and had to indicate for each user group the frequency with which service was provided. The charts below (Figure 17) show the frequencies indicated for each user group.

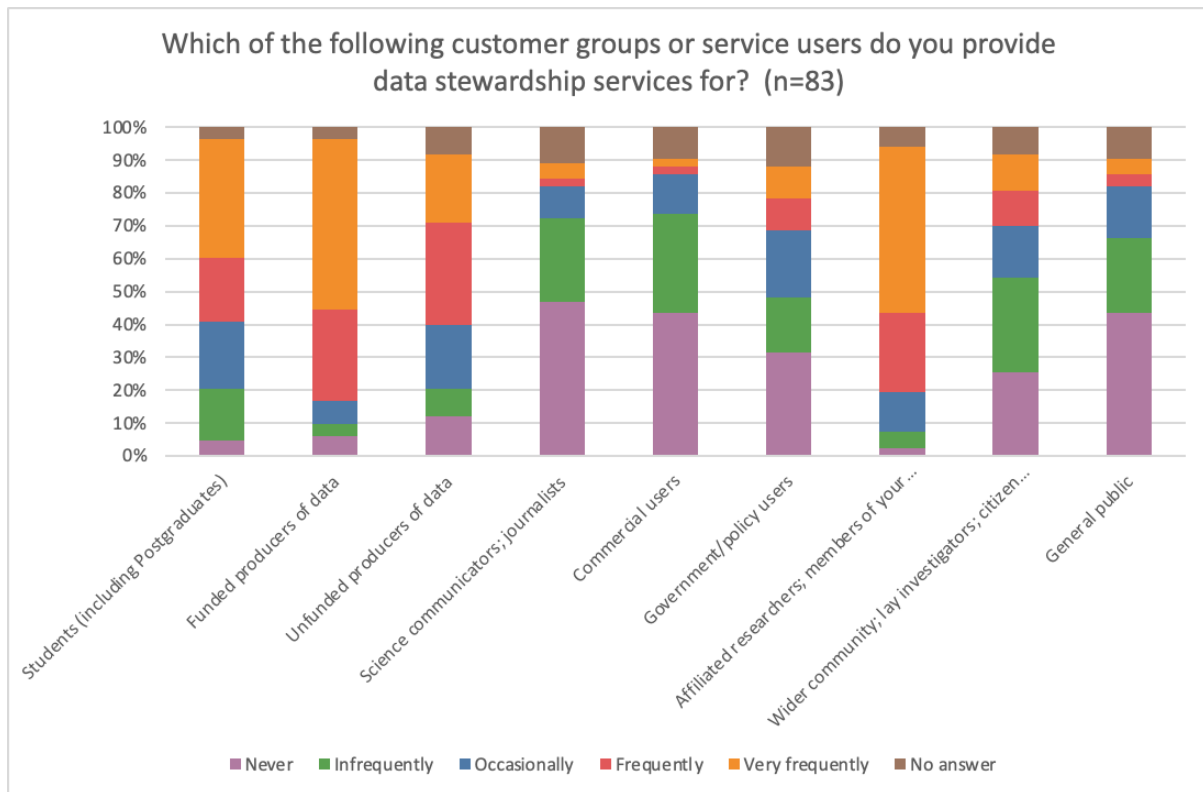


Figure 21. Frequency of options selected for the different target users that respondents' data stewardship services were provided for.

## Data stewardship support provided to the community

In Q12, we asked respondents to indicate which data stewardship actions their service provides to the community. Respondents were able to choose between multiple options. Among the respondents, the most frequent data stewardship actions provided to the community included (i) guidance for selecting data, services and tools for data management, (ii) supporting data management planning and (iii) ensuring good practice in versioning, curation and archiving. The data stewardship function that the respondents provide least frequently was developing a sustainable business model for a trustworthy service.

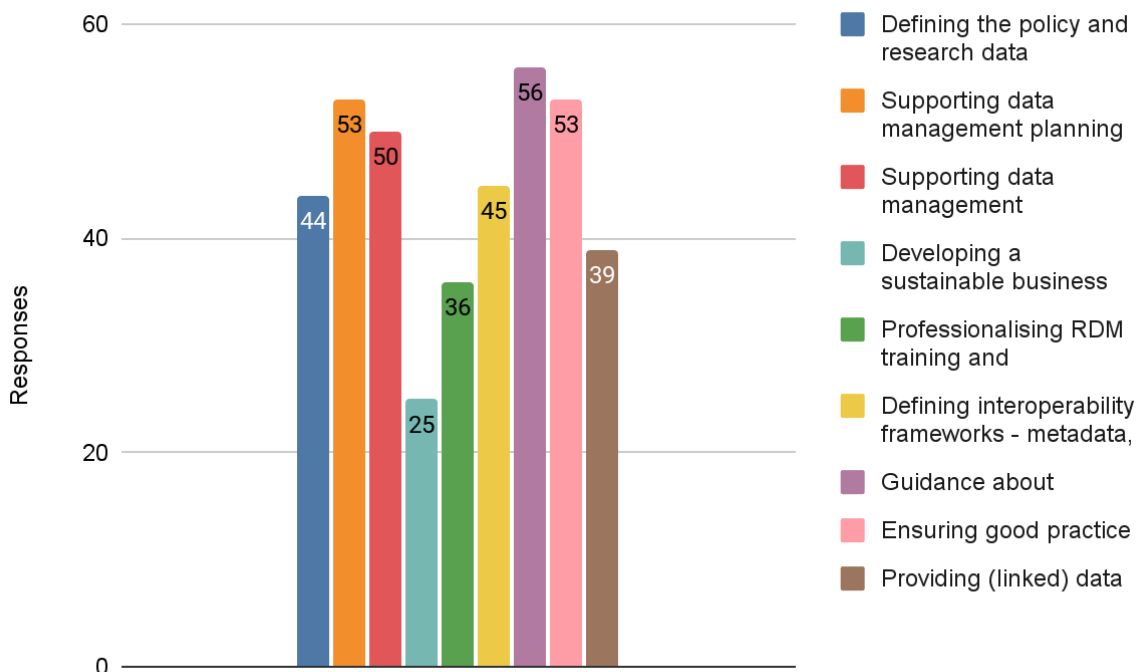


Figure 22. Frequency of option selection for which data stewardship functions the respondents' service provides to the community.

## Engaging with users throughout the research project lifecycle

In Q13 we asked respondents to rate their service's level of engagement with end-users in the various stages of the research project lifecycle, as drawn from the JISC "Research Data Management Toolkit"<sup>9</sup>. The scale used 0 for "little or no engagement" and 4 for "actively engaged".

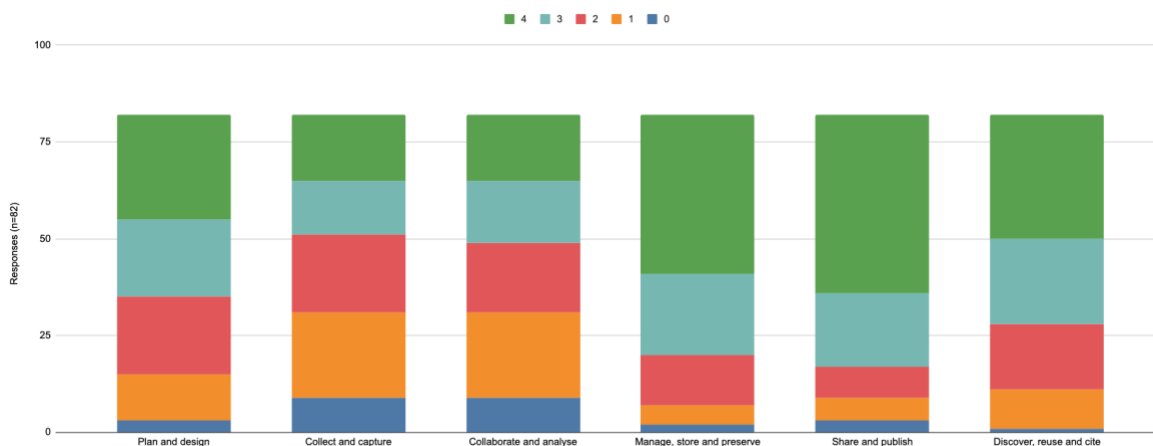


Figure 23. Responses for each part of the JISC "Research Data Management Toolkit" research project lifecycle. Column order from left to right as per the project lifecycle stages of the toolkit.

<sup>9</sup> <https://www.jisc.ac.uk/guides/rdm-toolkit>

## The benefits of data stewardship promoted to users

In Q14, we asked respondents about the added value or benefit of data stewardship that their organisation promotes to the users of data stewardship services. Respondents were able to choose between multiple options. Among the respondents, the most frequent benefit of data stewardship promoted to users included (i) support to meet FAIR principles, (ii) research visibility/citation and (iii) sustainable long-term digital preservation. The least frequently promoted benefits included enhanced value from data linking and risks being effectively managed. Respondents were also able to write in 'Other' values or benefits of data stewardship that their organisation promotes to users. In the 'Other' field, respondents indicated data reuse and interoperability most frequently.

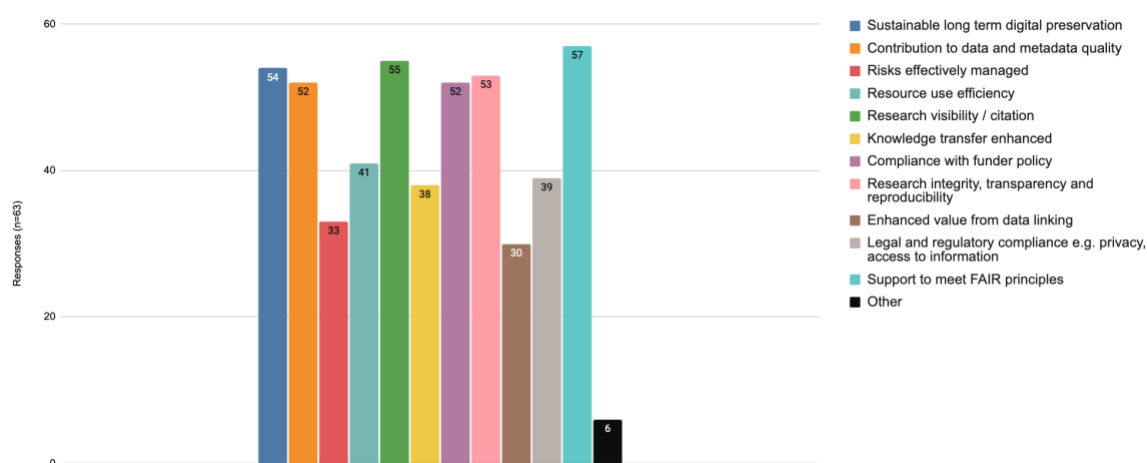


Figure 24. Frequency of options selected for the added value or benefit of data stewardship that respondents' organisation promotes to the users of data stewardship services.

## Service initiation and engagement

The next set of questions in the survey focused on service initiation and engagement.

### Key factors for initiating data stewardship support

Q15 explored which factors led to the successful initiation of data stewardship support in respondents' organisations. This was a multiple choice question, and respondents could select several factors. Respondents indicated the following key factors most frequently: (i) training services or networks are available for building the key skills and knowledge, (ii) the community or organisation has policies relating to data management and (iii) senior management sees the value in data stewardship.

The respondents also had the possibility to indicate 'Other' and write in key factors for initiating data stewardship support in their organisation. In the 'Other' field, respondents indicated (i) funder (and publisher) mandates, (ii) research culture and community uptake of data stewardship services and (iii) collaboration between various research data service providers within an organisation as additional key factors.

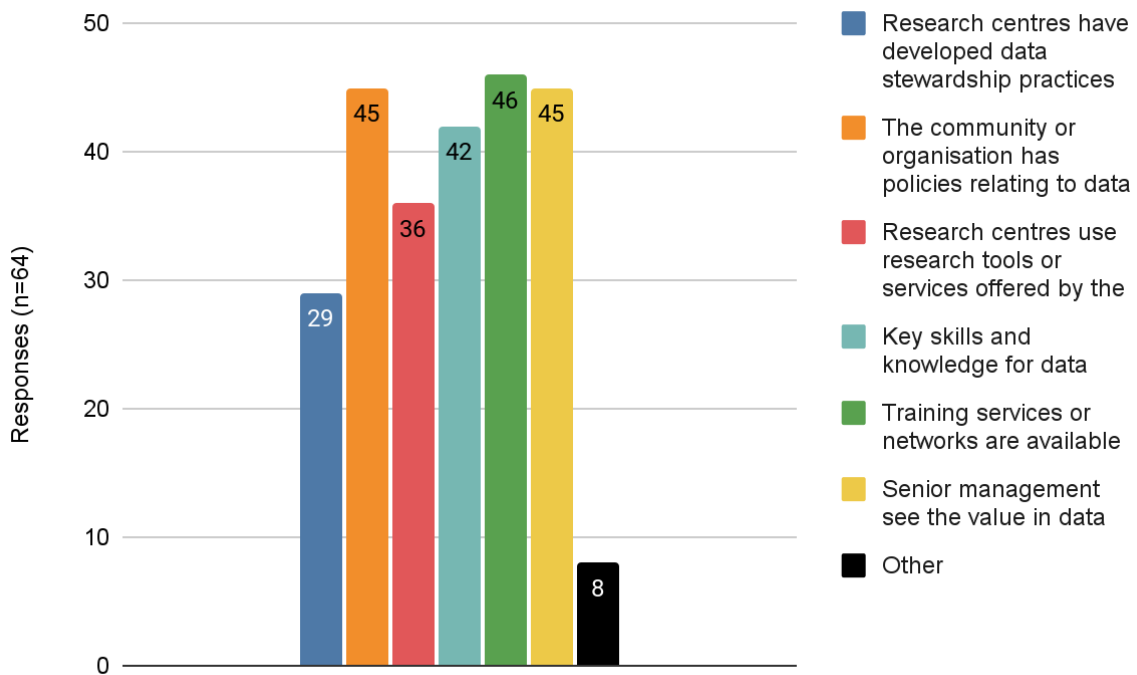


Figure 25. Frequency of options selected for factors led to the successful initiation of data stewardship support in respondents' organisations.

## Level of engagement with data stewardship practices, policies and standards

The respondents were also asked about the level of engagement their service has with data stewardship practices, policies and standards relevant to the communities it serves (Q16). The majority of respondents selected 'Collaboration' as their level of engagement.

Collaboration: the service also engages with the design, development, and review of data stewardship practice in the communities served. Consults and collaborates widely, potentially also taking a community coordination and leadership role.	42
Awareness: the service monitors data stewardship practice in the community or communities it serves and makes local practitioners aware of it.	23
Adoption: the service or its host organisation also supports data stewardship practitioners to embed community practice locally.	15
	n=80

Table 3. Responses for levels of engagement.

## Factors influencing engagement with research data producers

In Q17, respondents were then asked to discuss the main factors that influence their service's ability to engage with research data producers. This was an open-ended question,

and inductive thematic analysis was performed to study the qualitative responses received (n=54). The table below shows the main themes and the related codes that resulted from the thematic analysis. The first three themes cover aspects of engagement relating to the provision of a data stewardship service, i.e. the ‘offering’. The fourth theme is expressed by the acronym VOTER and covers aspects of service delivery and response from user communities.

Themes	Codes
Policy compliance	Funder data policy Journal data policy Institutional data policy Ease of implementation of recommendations/policies
Resourcing for support	Funding Capacity/ competence availability Senior management support
Scope of support offered	Data value/quality assurance Open licensing metadata/standards/ interoperability Integration with related support functions, e.g. HR Cross-lifecycle support Information literacy training Domain-specific support
Visibility, Outreach, Trust, and Embedding in Research practice (VOTER)	Service visibility Centralisation of support Researcher engagement with RDM/ FAIR Embedding/ experience in research Proactive user engagement Trust in service provided Sector leadership User review

*Table 4. Themes and codes for question 17 (n=59).*

## Service Maturity

Q18 asked respondents to assess the maturity of their service in the following areas of data stewardship (numbering corresponds to bars in Figures 32 and 33):

- A1. Defining the policy and research data governance environment
- A2. Supporting data management planning, e.g. DMP creation and review, data curation at project endpoint
- A3. Supporting data management documentation and traceability
- A4. Developing a sustainable business model for a trustworthy service
- A5. Professionalising RDM training and engagement
- A6. Defining interoperability frameworks - metadata, versioning, standards & identifiers
- A7. Guidance about selecting data, services and tools for data management
- A8. Ensuring good practice in versioning, curation and archiving



## A9. Providing (linked) data catalogues for accessing the data

Respondents were asked to rate the level of their service's maturity in each of the areas listed above using the following rating options:

0. no coverage or service
1. initial: new or immature service, but an intention to develop and to address performance issues
2. managed: complete coverage of this area but service may be limited, performance objectives are actively identified and monitored
3. defined: comprehensive coverage via a fully developed service aligned with overall organisational standards and practice, and with performance objectives that reflect this

Figure 32 shows the ratings selected by respondents for all areas of data stewardship as defined in the question. The responses in Figure 32 are from the 'complete' cohort (n=82), though not all respondents replied to all questions. Figure 33 provides the subset of responses relating to those selecting their highest organisational level to be national or higher.

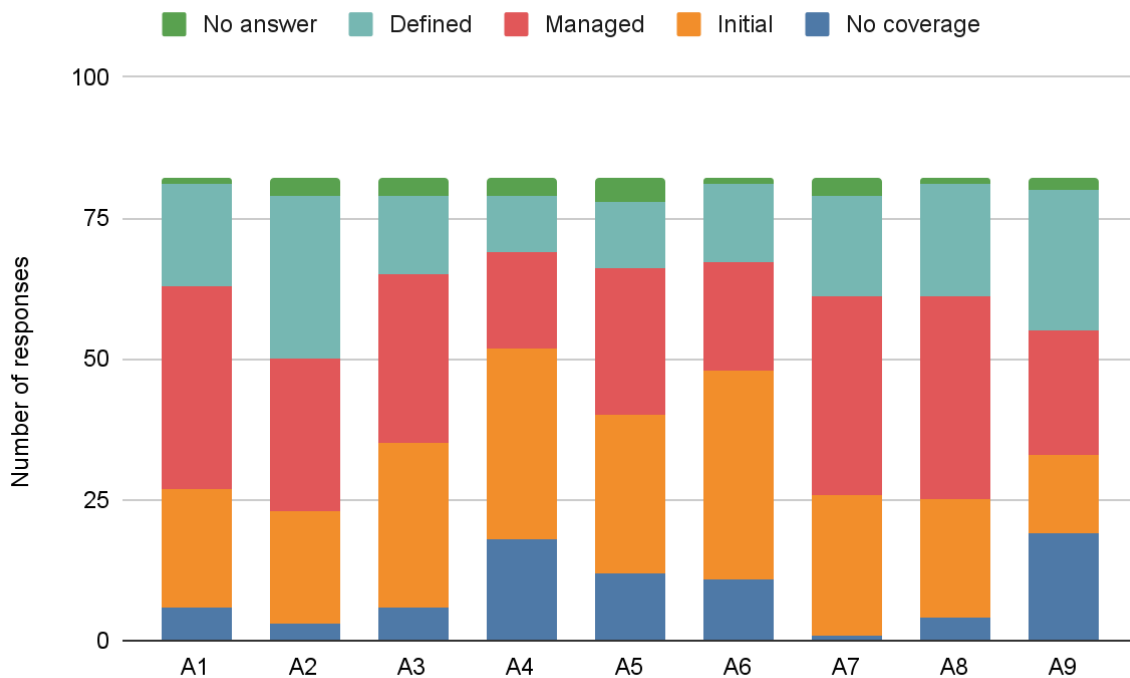


Figure 26. Service maturity responses.

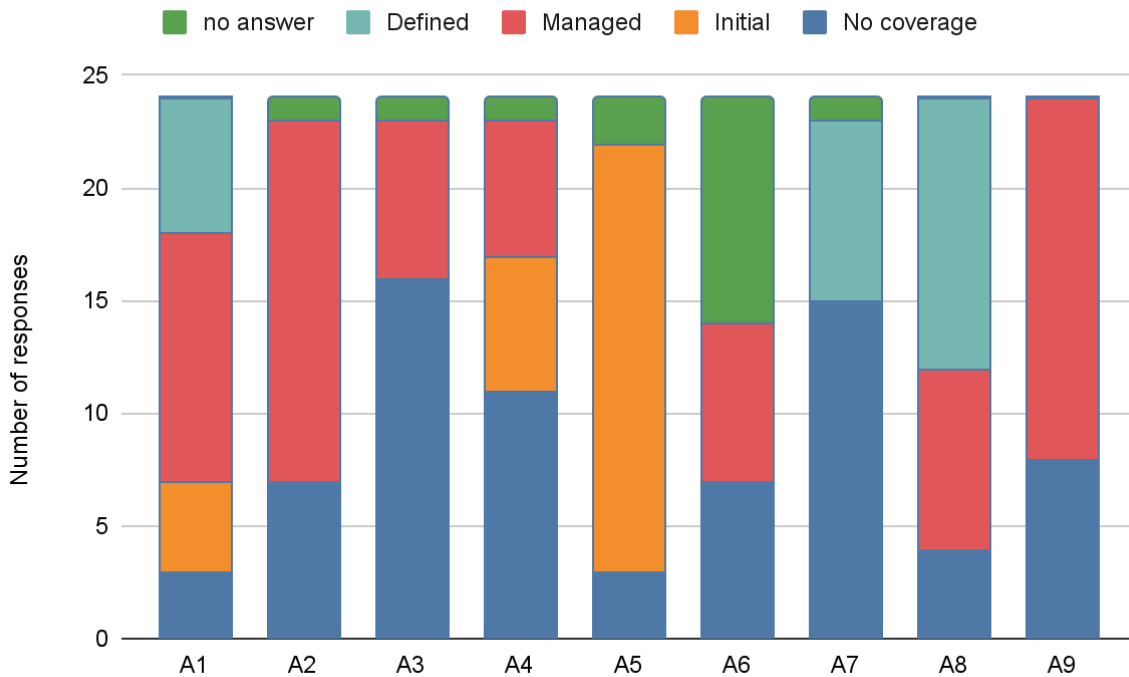


Figure 27. Service maturity responses for national to higher organisational level responses only.

There is a notable contrast between the overall set of responses and those from the sub-set given in Figure 33. For most questions, the latter cohort can appear to fall into two very distinct sets of responses with highly contrasting 'No coverage' and 'Managed'/'Defined' being selected.

### Resourcing the service: people and costs

The next block of questions (Q19-22) focused on resources. More specifically, we asked about the staffing of data stewardship services, user numbers as well as the budgets of these services.

### Staffing of data stewardship services

In Q19, respondents were asked to indicate the number of Full-Time Equivalent (FTE) posts in their service.

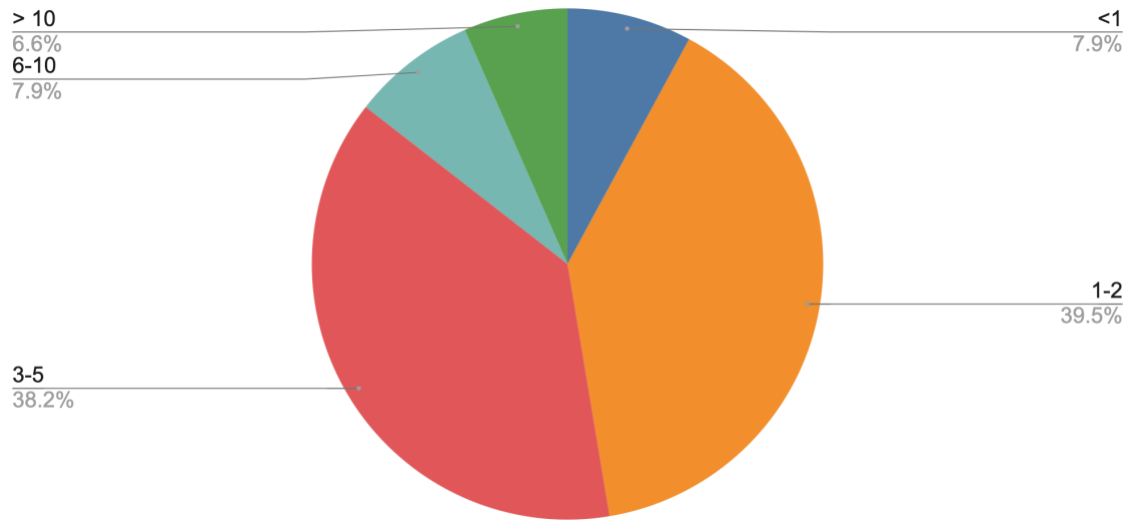


Figure 28. FTE staffing levels for all 'complete' responses (n=76).

Most respondents indicated that their service included 1-5 FTE staff. This response was from the "complete responses" cohort, which was dominated by the institutional level responses as indicated above. When considering only those responses whose highest organisation level was national to global, we can note a slight shift to higher staffing levels:

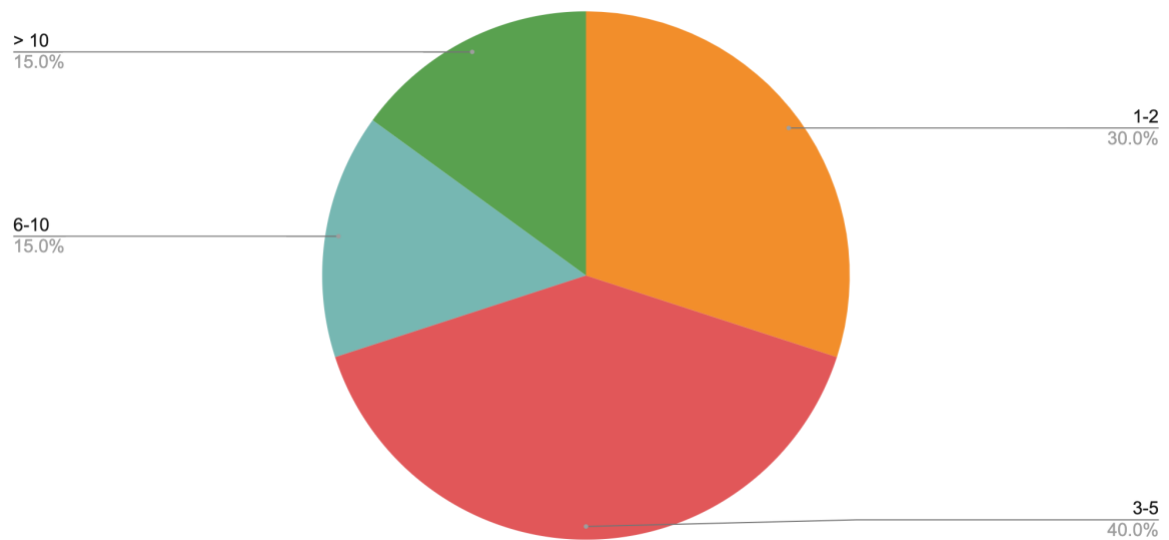


Figure 29. FTE staffing levels for 'complete' level responses where the highest organisation level was national to global (n=20).

Those services that included less than 1 FTE were all institutional, departmental or research group-based services:

Institutional	3
---------------	---

Departmental	2
Research group	1

*Table 5. Highest organisation level spread for responses with less than one FTE.*

Furthermore, those services which had 6-10 or more than 10 FTEs were all multi-institutional or higher organisation level services:

Global	3
Multi-national	1
National	2
Multi-institutional	1
Institutional	2
Inter-departmental	1
Research group	1

*Table 6. Highest organisational level for responses with 6 or greater FTE staff.*

Among our respondents, the data suggests that the higher the organisational level of the service, the more likely it is to have more staff.

### Size of the supported community

In Q20 we asked respondents to indicate the size of the user community that their service supports. We asked the respondents to indicate the size of the community in terms of 3 specific parameters: Data Management Plans (DMPs), projects, and researchers. These three parameters were selected in recognition that different services would use different metrics to measure their communities.

These plots show the spread of responses in terms of a) number of DMPs, b) number of projects; c) number of users on the three plots' y-axis. The X-axis shows the highest organisational level selected and the colours show the level of 'domain specificity' for the service.

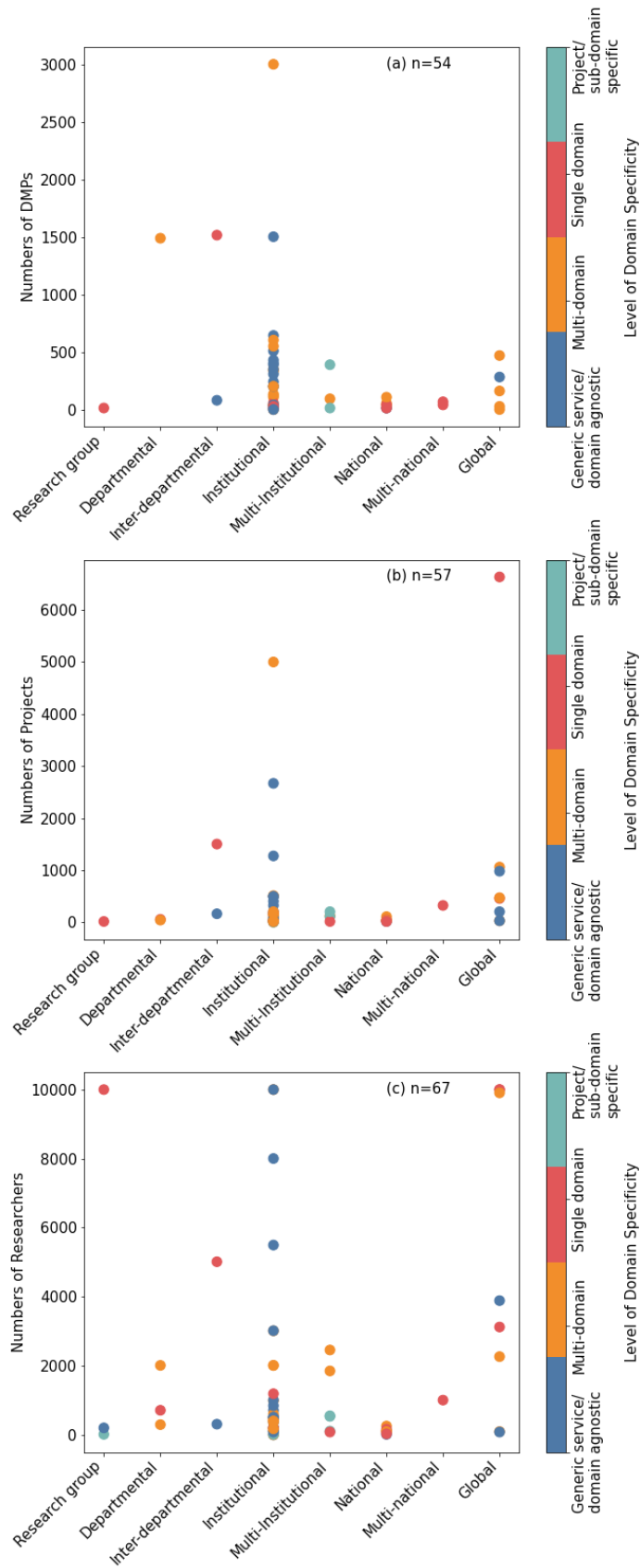


Figure 30. Plots showing scale of services in regards to a) number of data management plans (DMPs) created each year; b) number of projects supported each year; and, c) number of researchers supported each year by the service. Results are plotted against the highest organisational level selected with colours depicting the level of domain specificity of the service.

## Access to service products

In Q21, we asked respondents to estimate how many unique users access their service's datasets annually.

	All responses (n=34)	National and higher responses (n=10)
< 100	12	1
100 - 1000	14	6
1001 - 10,000	8	3
No response	102	32

*Table 7. Numbers of unique users accessing the respondent's service.*

## Service funding.

In Q22 we asked how data stewardship support is resourced in respondents' organisations. This was a multiple choice question, and respondents could choose between 8 following statements:

1. Recurring dedicated budget is made available for data stewardship
2. Fixed-term budget for a commissioned service
3. Direct costs to research grants or projects
4. Allocated overheads from research grants or projects
5. Fixed-term development grant/ seed funding
6. Volunteer effort
7. Don't know
8. Other

The question also specified that it referred to financial costs.

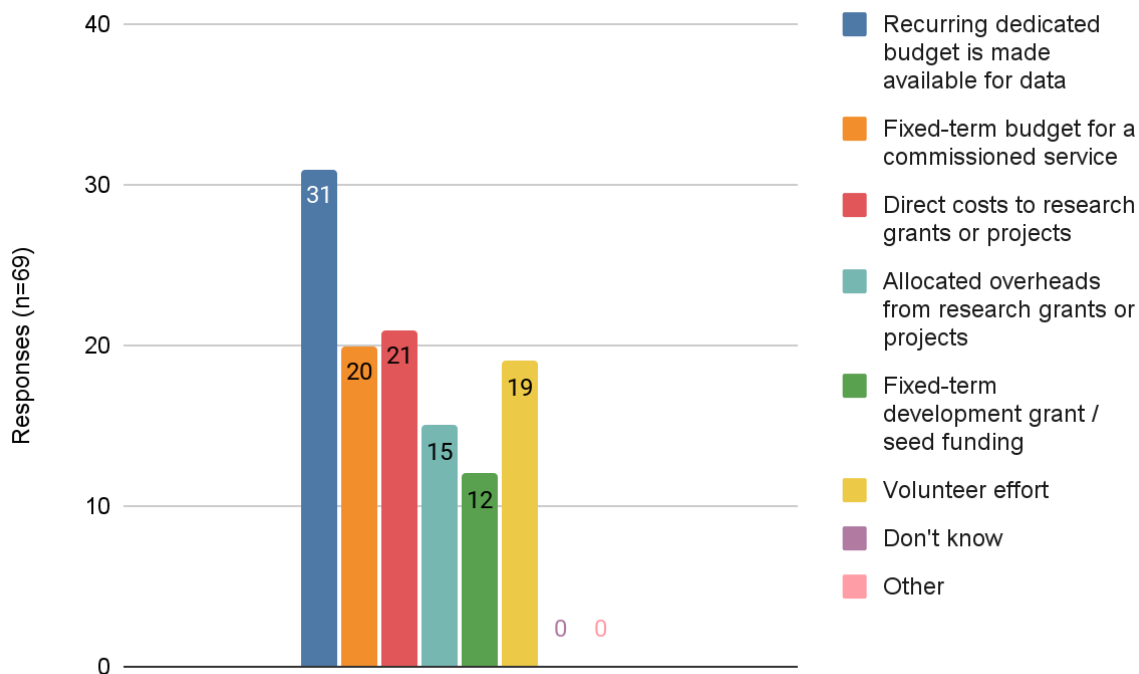


Figure 31. Selected resource options for research data management in respondents' organisations (n=69).

As table 6 shows, only one funding option was present for half of the respondents' services whilst the other half showed varying levels of diversity in funding models. Among the funding options included in the Other field project-related budgets (n=3) and fixed-term grants (n=2) were most frequently mentioned. It is also interesting to note that volunteer effort is an important component for around 1/3 of respondents' services. 7 of the 19 services selecting this option were entirely reliant on this as their source of resource.

no. options	responses (n=69)	responses (n=21)
1	35	8
2	15	4
3	16	8
4	3	1

Table 8. Spread of the number of responses selected for 'complete' responses (n=69) and those from national to higher organisational level selections (n=21).

When examining these figures only for those services at a national or higher organisational level (n=21) funding models appear to have a greater diversity of components (Figure 35). In both cohorts, half indicated that recurring dedicated funding for the stewardship activities is present, which indicates that there is a degree of stability in funding available to data stewardship services, though not for all services. A comparison between responses from the service types

and service maturity questions would enable further insight but has not been performed for this report.

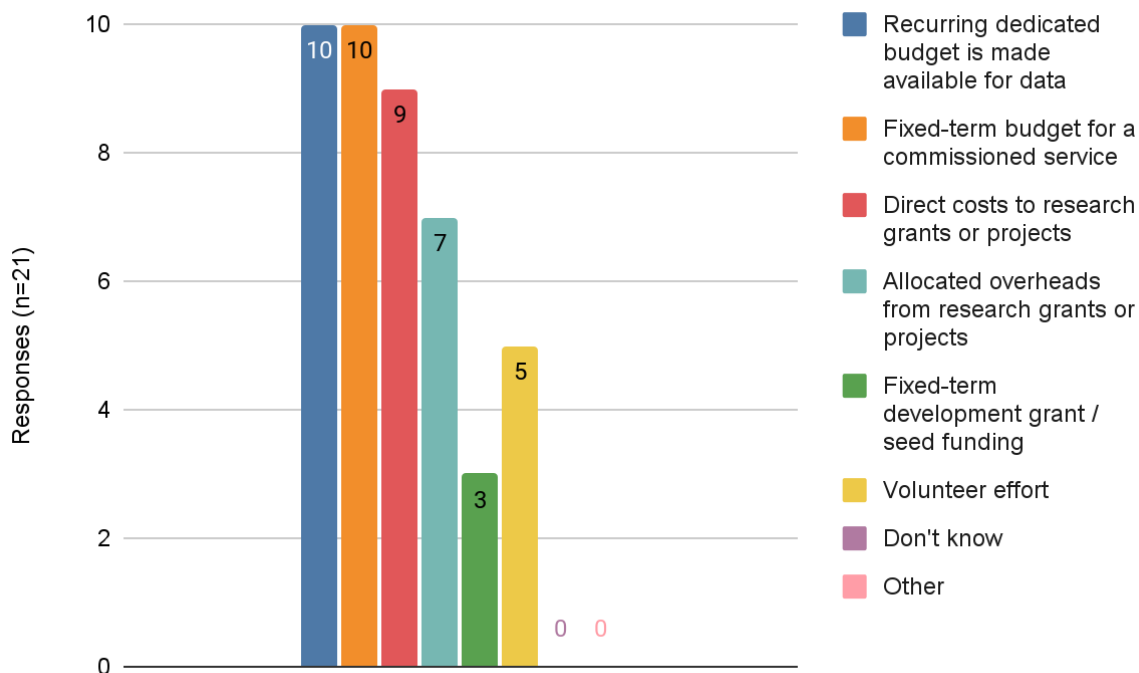


Figure 32. Selected resource options for research data management in respondents' organisations for national level and higher level organisational responses (n=21).

## Context for the Data Stewardship Service

The final block of questions aimed to surface a better understanding of the broader context of the respondents' data stewardship services. The questions asked focused on the data stewardship stakeholders in the respondents' organisations, the perceived factors that could increase data stewardship adoption, the barriers to developing the support as well as the key factors for successful data management adoption/engagement.

## Governance of data stewardship services

In Q23, we asked about stakeholders who are involved in the governance of the respondents' data stewardship service. This was a multiple-choice question.



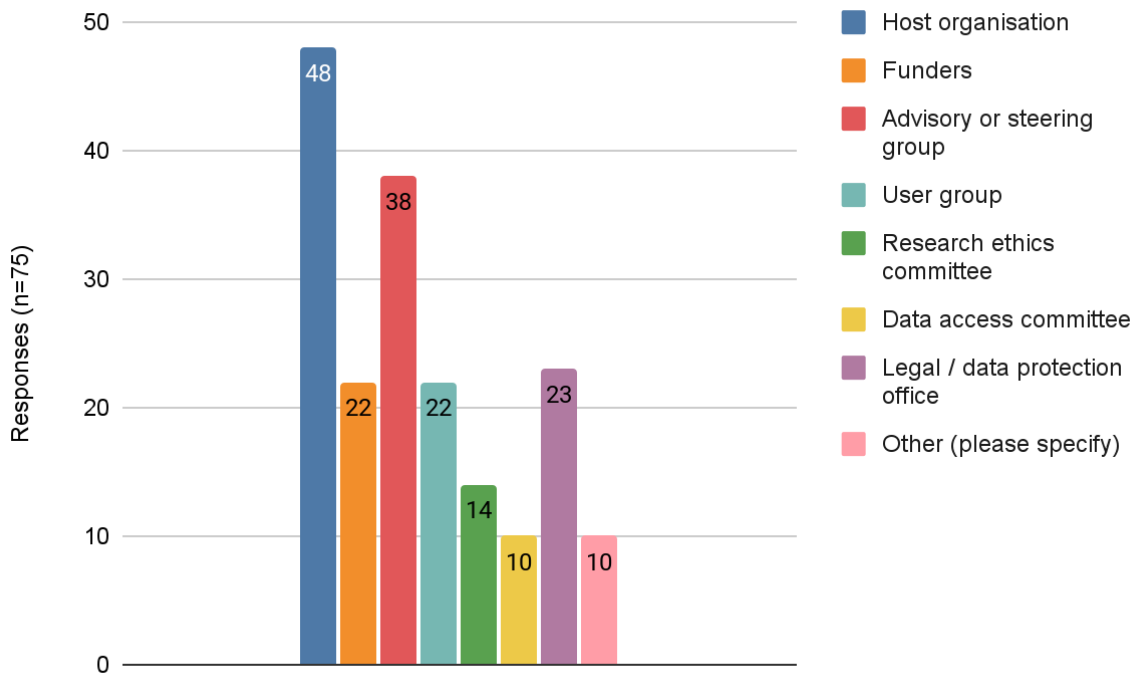


Figure 33. Selected stakeholder options for all responses (n=75).

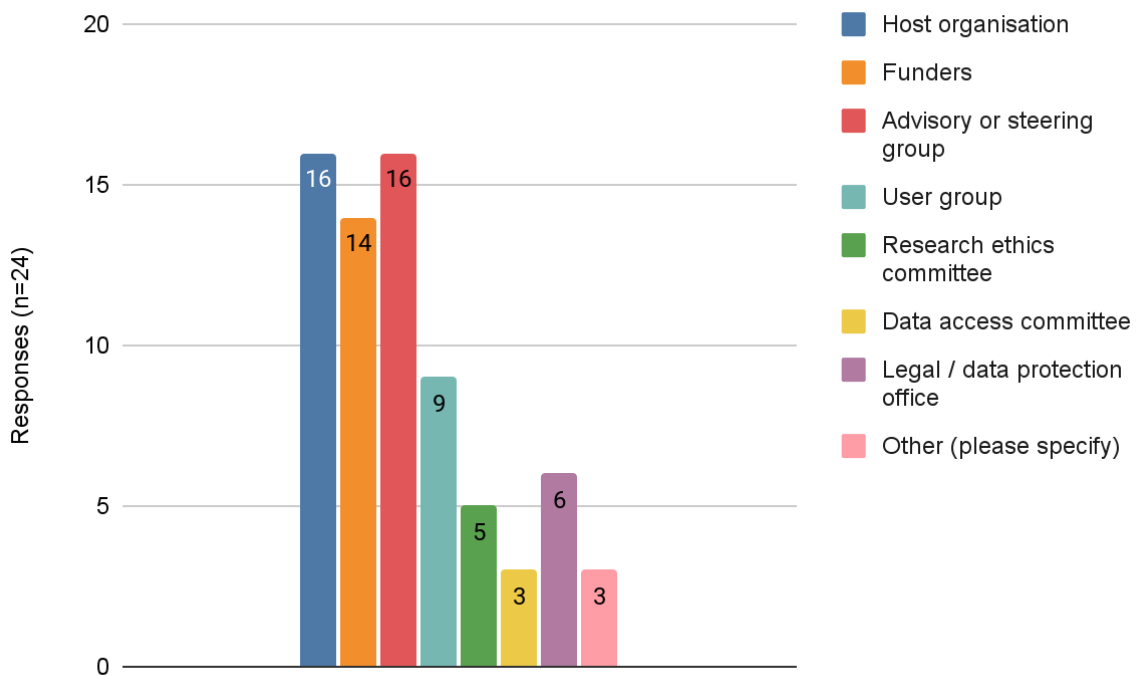


Figure 34. Selected stakeholder options for national and higher organisation level selections (n=24).

no. of stakeholders selected	responses (n=75)	responses (n=24)
1	22	3

2	15	5
3	18	7
4	11	3
5	4	1
6	2	0
7	3	2
8	0	0

Table 9. Spread of the number of stakeholders selected for all responses (n=75) and national and higher organisation level responses (n=24).

'Other' responses included a number related to IT support/infrastructure (n=3), library services (n=2), project boards (n=2) as well as publishers in one instance.

### Key communities related to providing data stewardship service

In Q24 we asked respondents to select the key communities involved in providing their data stewardship service. This was a multiple-choice question.

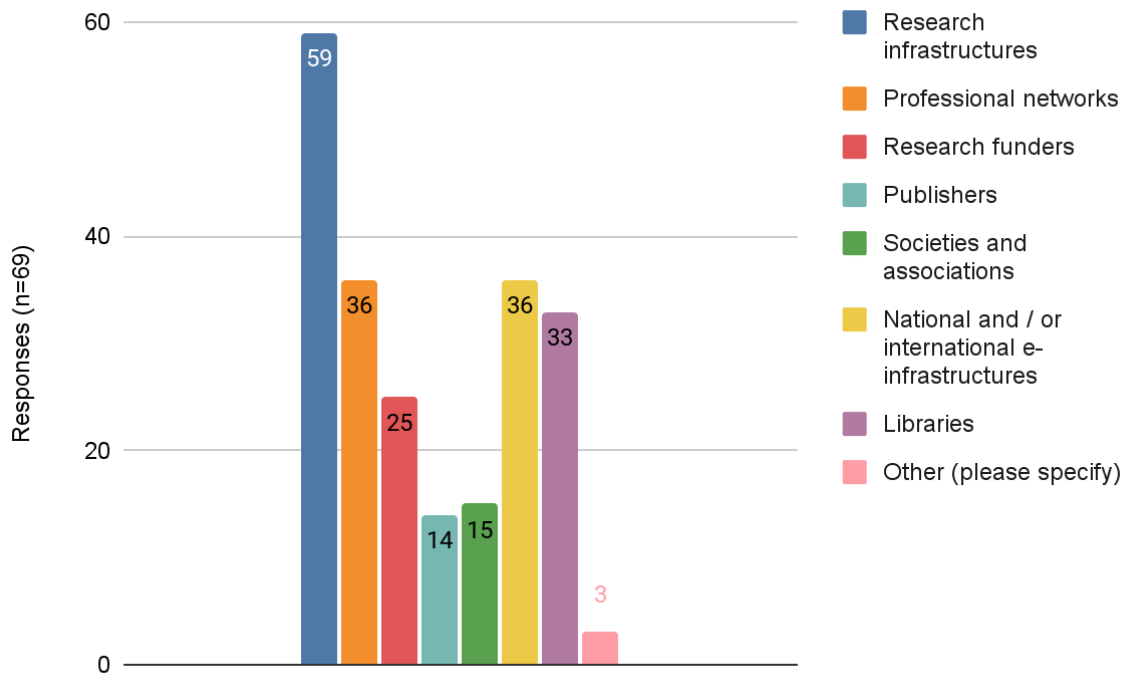


Figure 38. Selected communities supported options (n=69).

The "Other" response field included references to students, the global disciplinary community, and the research project board.

We also disaggregated the responses based on the institutional level.

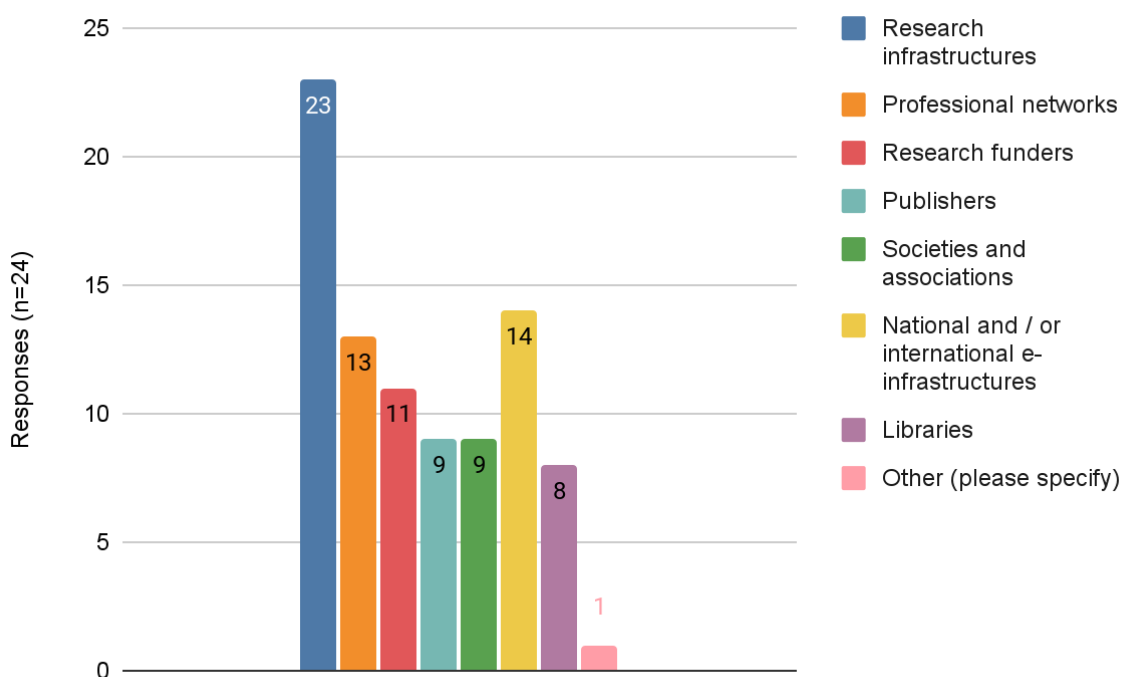


Figure 35. Selected communities supported options for national and higher organisation level selections (n=24).

no. options selected	All responses (n=69)	National and higher responses (n=24)
1	4	2
2	3	3
3	7	8
4	2	5
5	2	1
6	1	0
7	2	2
8	0	0

Table 10. Spread of options selected for 'complete' responses (n=69) and those from national and higher organisation level selections (n=24).

## Factors to increase adoption of data stewardship service by users

In Q25 we asked about factors that might increase the adoption of data stewardship service by users. This was a multiple choice question, and respondents could choose between the following factors:

1. Exemplars from communities with well-developed data stewardship practices
2. The community has developed policies for data management
3. Uptake of tools/services that help produce outputs based on good data stewardship practice
4. Awareness of and ready access to skills and knowledge for data stewardship
5. Training to support the development of key skills and knowledge for data stewardship
6. Senior management sees enough value in data stewardship to promote and resource it
7. Requirements from publishers and funders for the availability of data outputs
8. Demands from funders or commissioning organisations to deliver impact from data
9. Other

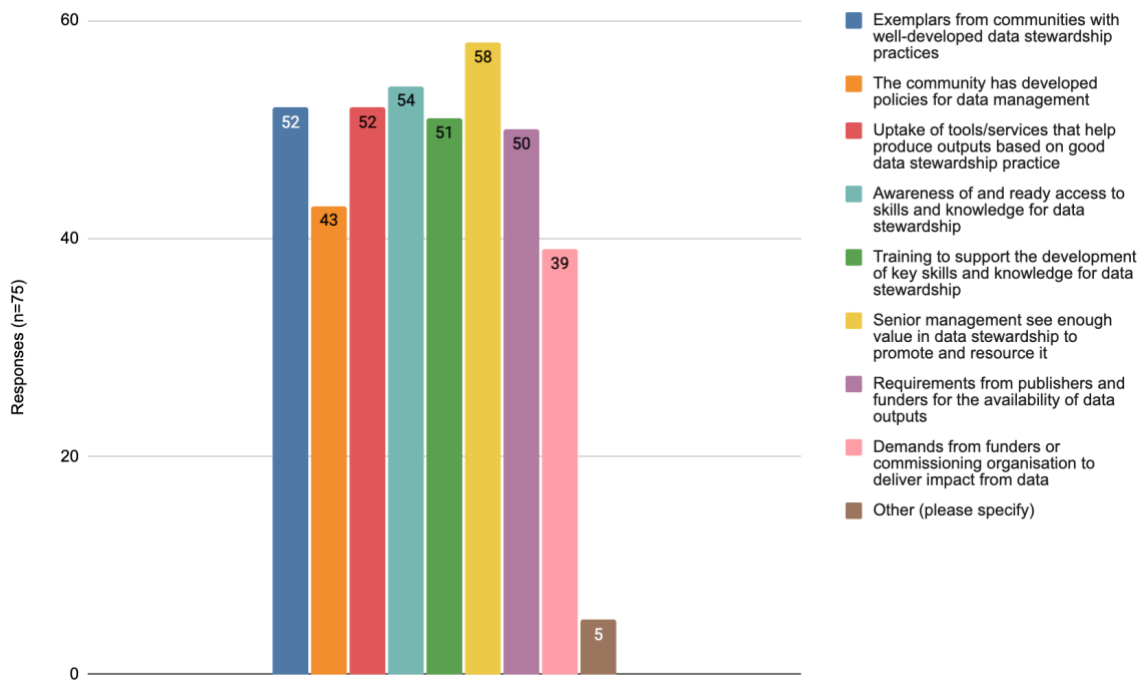


Figure 36. Frequency of selection covering the factors increasing data stewardship service adoption, all responses.

The responses in the “Other” field focused included 3 broad areas of factors: (i) awareness of funders of the value of data stewardship; (ii) visibility of the contribution of data stewards and data stewards' engagement with researchers, also those researchers not subject to data

sharing mandates; and (iii) existence of guidelines, standards and tools for data management.

We also disaggregated the responses based on the institutional level.

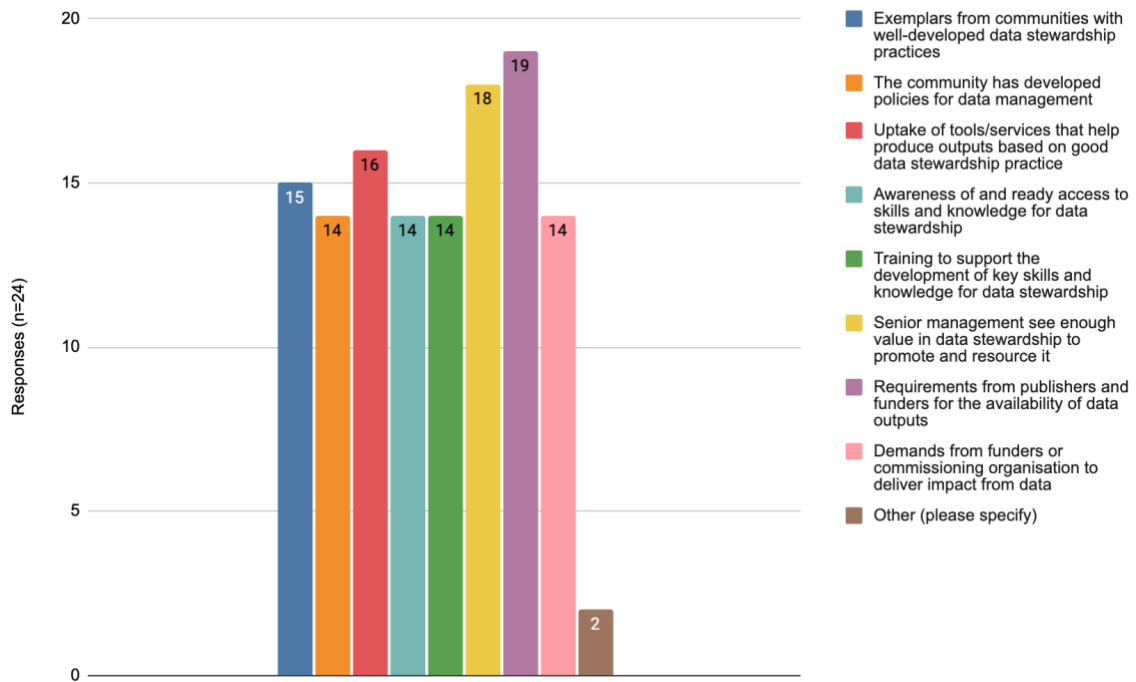


Figure 37. Frequency of selection covering the factors increasing data stewardship service adoption, national and higher organisation level selections (n=24).

no. options selected	responses (n=75)	responses (n=23)
1	4	1
2	5	1
3	3	2
4	13	7
5	14	2
6	12	7
7	10	3
8	14	0

Table 11. Range of factors selected for increasing data stewardship services for all responses (n=75) and for national and higher organisation level selections (n=23).

## Contributors to success

In Q26, we asked respondents to elaborate on successful methods/approaches their service has used to engage with research data producers. Thematic analysis was performed on these narrative responses; the themes and codes that emerged are presented in the table below.

Themes	Codes
Support for data production	Whole lifecycle support/ 1:1 support Participation in research project governance Grant application support DMP support/ policy implementation Support for standards development Support for data publishing, archiving, curation
Skills development	Doctoral training Information literacy Awareness training Training fellowships
Recognition of good practice	Offer badges for skills acquisition Recruit researchers as champions/ ambassadors Build reputation for data/service utility, quality, trust
Data tools provision	Database provision Data standards implementation Co-design of tools
Effective communication	Community networking/workshops Using institutional channels Surveys- open, targeted Participation in research, scholarly communication Senior management advocacy Make service visible, promoting benefits and guidance

*Table 12. Thematic analysis of textual responses where respondents elaborated on successful methods/approaches their service has used to engage with research data producers (n=46).*

## Barriers to data stewardship services

Finally, in Q27-Q28, we asked about barriers to data stewardship services.

In Q27, we asked respondents to indicate how relevant certain barriers were to the adoption of good data stewardship in their organisations. The respondents had to evaluate the relevance of the following barriers:

1. Duplication of effort with similar organisations or existing services
2. Legal and ethical regulatory environment
3. Lack of buy-in from senior managers or researchers
4. Difficulty changing stakeholders' attitudes or practices
5. Lack of incentives/rewards for data stewardship compared with journal publication
6. Lack of clarity from funders regarding allocations for data management costs

### Barriers to data stewardship services (n=82)

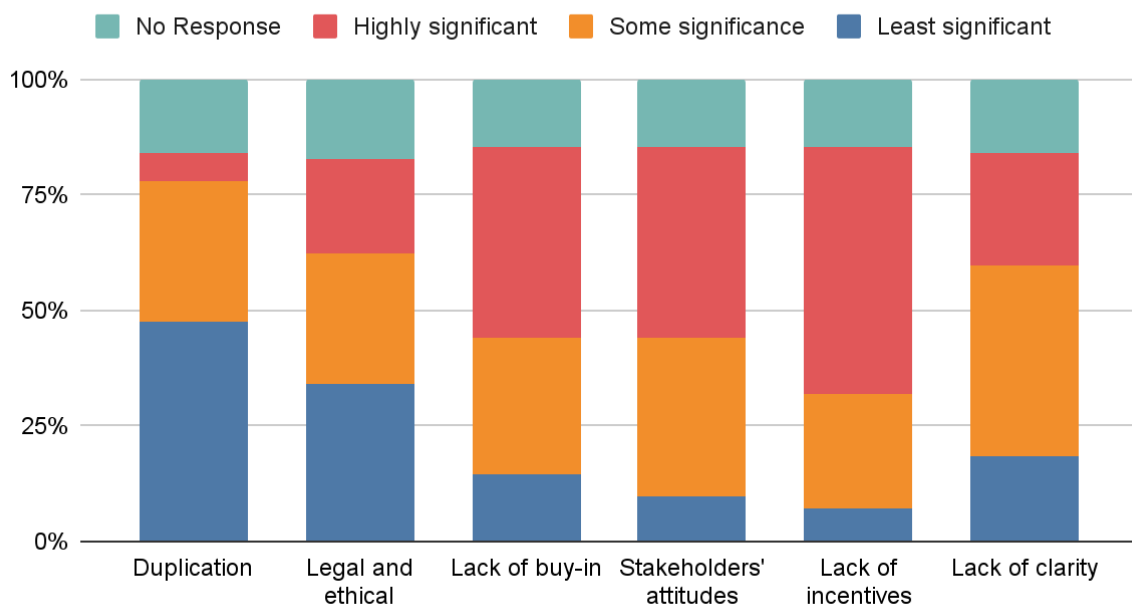


Figure 38. Barriers to data stewardship service, all responses (n=82).

### Barriers to data stewardship services - National level and above (n=24)

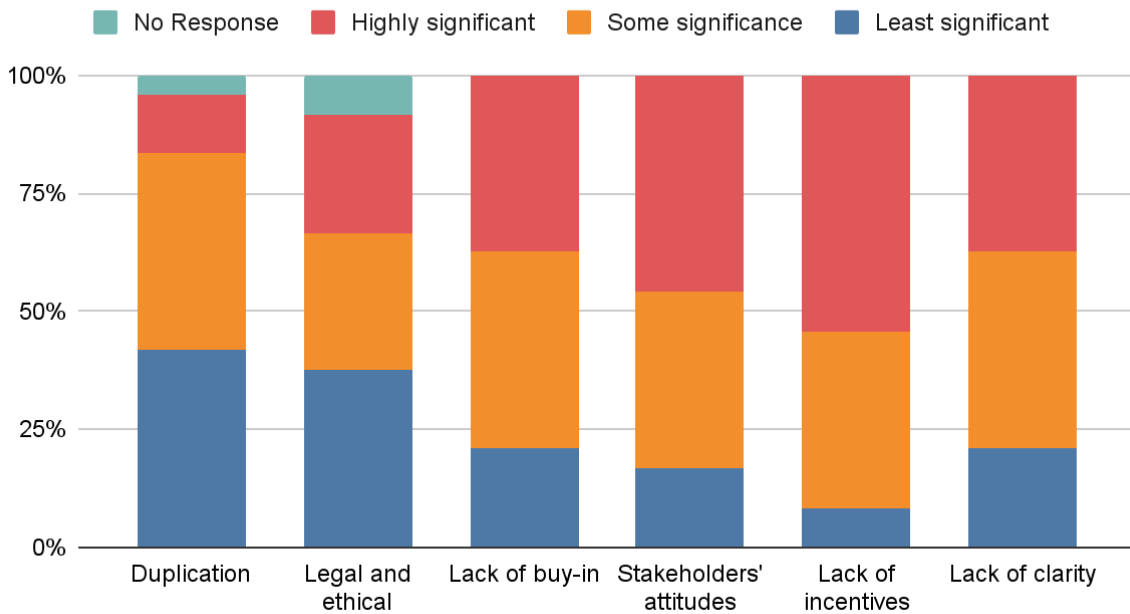


Figure 39. Barriers to data stewardship service, for national and higher organisation level selections (n=24).

### Barriers to data stewardship services - Institutional (n=38)

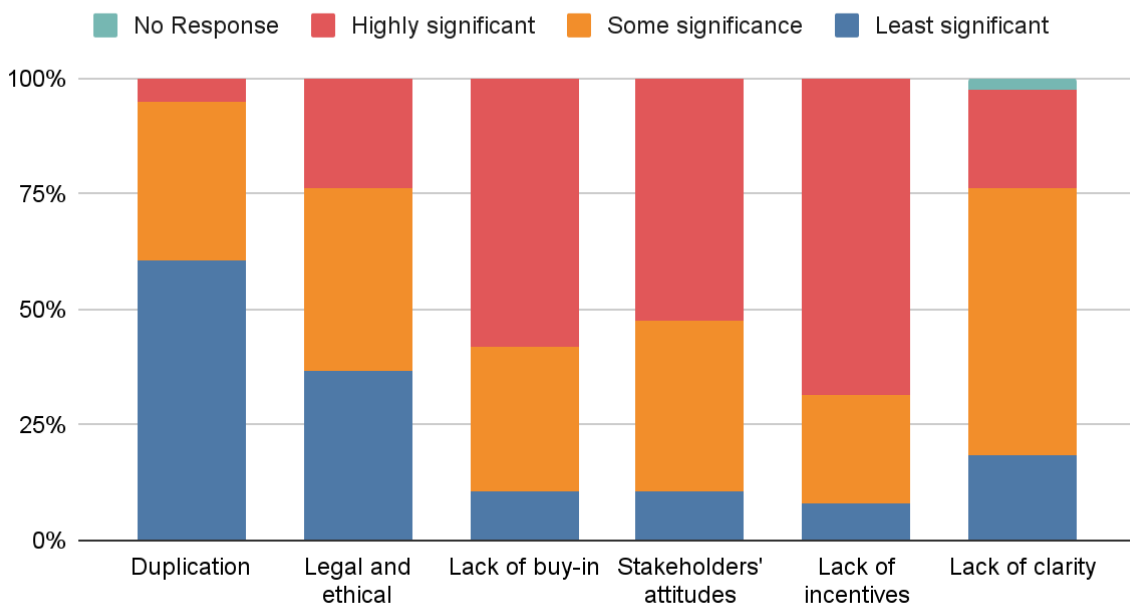


Figure 40. Barriers to data stewardship service, for institutional level selections (n=38).

The barriers that more respondents considered highly or somewhat significant included (1) lack of incentives/rewards for data stewardship compared with journal publications, (2) difficulty changing stakeholders' attitudes or practices, and (3) lack of buy-in from senior



managers or researchers. The barrier that was considered least significant by respondents was duplication of effort with similar organisations or existing services.

In Q28, respondents had the possibility to include responses about other barriers they see as impacting the adoption of good data stewardship practices. Thematic analysis was performed on these narrative responses; the themes and codes that emerged are presented in the table below.

Themes	Codes
Lack of incentive/credit and reward	Lack of accountability for poor practice Lack of credit or recognition for good practice Cultural inertia among academics Unwillingness to share data
Unmet research support needs	Lack of domain-specificity Lack of awareness training/ materials Lack of support for non-public data access Insufficient budget preparation/ support for costing Service complexity Lack of semantic competences in IT community Inadequate tool provision
Gaps in coordination of support	Insufficient institutional coordination to address the required scope Insufficient national/international coordination Lack of consensus on stewardship roles definition Lack of sustainable funding/ capacity Lack of senior management support in institutions

Table 13. Thematic analysis for text responses regarding barriers to data stewardship (n=36).

## Main lessons learned

In the last question of the survey, Q29, respondents were asked to share the main lessons they have learned from developing their data stewardship service that they would like to share with similar organisations. This was a free text question and thematic analysis was performed on the narrative responses; the themes and codes that emerged are presented in the table below.

Themes	Codes
--------	-------

<p>Build on concrete benefits to service users and stakeholders</p>	<p>Seek senior management support  Gain support from institutional governance  Take holistic approach to service provision  Use `DMPs to introduce RDM  Actively engage with community  Seek to understand the problem to be solved  Provide factual and relevant services that add value and tangible benefits, driven by science needs  Integrate interoperability  Research and document good practices  Build support incrementally  Advocate long-term stewardship &amp; research integrity to PIs/ community  Build trusted data repository</p>
<p>Embed data stewardship in teaching and training</p>	<p>Build discipline-specific training  Develop students' data literacy  Train early-career researchers  Develop Masters programmes in DS</p>
<p>Engage with external stakeholders</p>	<p>Engage with funders, adapt to their requirements  Work with journals for data quality control  Network with similar services  Participate in (inter)national expert groups  Scan the horizon  Respond to change in science and technology ecosystem</p>
<p>Coordinate across the organisation</p>	<p>Collaborate across institution  Build on existing networks  Involve Legal, IP and IT infrastructure services in RDM support  Seek both central and embedded roles  First centralise then grow through decentralisation  Develop 3 core role profiles - research, IT and information  Build permanent roles to sustain expertise</p>
<p>Communicate using effective methods</p>	<p>Avoid jargon  Define data stewardship  Promote understanding of FAIR  Use graphic communication  Connect people and problems  Develop and promote use cases with concrete results  Work with small groups  Provide and promote RDM benefits</p>
<p>Drive and monitor improvements</p>	<p>Gain insights into research work through impact studies</p>

	<ul style="list-style-type: none"> <li>Target simple things first</li> <li>Reduce administrative burdens for researchers</li> <li>Offer researchers incentives / reward</li> <li>Counter poor practice</li> <li>Improve service capabilities</li> <li>Seek efficiency gains</li> <li>Persist in building reputation</li> </ul>
--	--

*Table 14. Thematic analysis for text responses regarding the main lessons services/respondents have learned from developing their data stewardship service that they would like to share with similar organisations (n=46).*

## Avenues for Future Work

This report is placed within a wider context as indicated in the opening sections regarding wider connections within the Professionalising Data Stewardship Interest Group and further afield via the community comments. It sought to gain insights into data stewardship services globally, providing a resource that may assist the wider community in furthering its aims on professionalisation of data stewardship. Whilst these aims have been largely met in this report, limits highlighted and insights gained also give clarity on the direction of future work that may be undertaken by the community.

Whilst the wider Interest Group will reflect on this work and incorporate its outputs, where appropriate, into other areas of its work (e.g. feeding into the Business Case task group) other communities (e.g. EOSC) may also seek to engage with these outputs within their own contexts.

With regards to the report's main limit around its largely Eurocentric coverage and the under-representation of other global regions, a clear area of work would be around engaging those under-represented communities. Furthermore, besides developing specific engagement globally, exploring issues around terminology would also be advantageous to aid engagement and communication of findings.

To complement the wider landscape view that this survey and report has explored, studies about service evolution/life cycles would also benefit the wider community. Indeed, during the course of this work a number of side-discussions, not reported here, indicate there exist ready case-studies to be explored. Such work would be beneficial especially to those services evolving from initial stages of implementation through to more sustainable funding and operational models, whilst matters around end-of-service procedures and issues arising could underpin planning for such scenarios.

Finally, there are parts of the conceptual model (see [figure 2](#)) that were not explored, such as those around the policy and standards context for services and how they shape the services themselves, which would help to complete the picture of services and their contexts.

All these avenues, though, require further engagement with the community and appropriate further effort, but the authors hope that this output will serve to aid such future work.