



HAL
open science

FIDELIS landscape survey analysis

Thomas Jouneau, Maaïke Verburg, Laurence Horton, Geert van Geest, Sanni Tujunen, Joel Kallio, Tobias Paulsen, Severine Duvaud, Jonas Recker, Åse Jorun Holthe-Tveit, et al.

► **To cite this version:**

Thomas Jouneau, Maaïke Verburg, Laurence Horton, Geert van Geest, Sanni Tujunen, et al.. FIDELIS landscape survey analysis. FIDELIS. 2025, 98 p. <hal-05467646>

HAL Id: hal-05467646

<https://hal.science/hal-05467646v1>

Submitted on 20 Jan 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Project Title	FIDELIS: Establishing A European Network of Trustworthy Digital Repositories
Project Acronym	FIDELIS
Grant Agreement No.	101188078
Start Date of Project	2025-01-01
Duration of Project	36 months
Project Website	https://eden-fidelis.eu/

FIDELIS landscape survey analysis

Work Packages	WP5 - Building common understanding of TDRs and a harmonised matrix of repository capabilities and characteristics WP8 - Repository training and support: Initiating Network cohesion and expansion
Lead Authors (Org)	Thomas Jouneau (INRAE, Université de Lorraine, 0000-0001-5986-8128), Maaïke Verburg (KNAW-DANS, 0000-0001-9408-3190)
Contributing Author(s) (Org)	Laurence Horton (JEDIN-DCC, 0000-0003-2742-6434), Geert van Geest (SIB, 0000-0002-1561-078X), Sanni Tujunen (TAU-FSD), Joel Kallio (TAU-FSD, 0009-0003-5076-9985), Tobias Paulsen (Sikt, 0009-0004-5123-6782), Severine Duvaud (SIB, 0000-0001-7892-9678), Jonas Recker (GESIS, 0000-0001-9562-3339), Åse Jorun Holthe-Tveit (Sikt, 0009-0005-8185-8854), Deborah Thorpe (KNAW-DANS, 0000-0002-2307-8770), Philipp Conzett (UiT, 0000-0002-6754-7911), Mari Kleemola (TAU-FSD, 0000-0001-8855-5075), Vasso Kalaitzi (KNAW-DANS, 0000-0001-8337-120X), Robert Huber (UBremen, 0009-0007-3843-2385), Clement Jonquet (INRAE, 0000-0002-2404-1582), Fernando Aguilar (CSIC, 0000-0002-5345-1716), Ann-Kristin Forshaug (Sikt, 0009-0004-1823-8604), Tuomas Alaterä (TAU-FSD, 0000-0002-3448-3448), Eduardo Esteves (ELIXIR, 0000-0001-5458-4978)
Due Date	2025-06-30
Date	2025-06-27
Version	V1.0
DOI	https://doi.org/10.5281/zenodo.15744996

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author(s)	Notes
0.1	2025.03.11	Maaike Verburg (DANS), Philipp Konzett (UiT), Mari Kleemola (TAU-FSD), Åse Holthe-Tveit (Sikt), Ann Kristin Forshaug (Sikt), Tobias Paulsen (Sikt), Joel Kallio (TAU-FSD), Sanni Tujunen (TAU-FSD), Tuomas Alaterä (TAU-FSD)	Survey design and launch
0.2	2025.03.25	Maaike Verburg (DANS)	Set up ToC of the report
0.4	2025.06.03	Maaike Verburg (DANS), Laurence Horton (UEDIN-DCC), Geert van Geest (SIB)	Drafted chapter 2 and repository characteristics results
0.5	2025.06.06	Maaike Verburg (DANS), Laurence Horton (UEDIN-DCC), Geert van Geest (SIB)	Draft ready for internal review WP8
0.6	2025.06.16	Thomas Jouneau (INRAE, Université de Lorraine), Sanni Tujunen (TAU-FSD), Joel Kallio (TAU-FSD)	Draft ready for internal review WP5
0.7	2025.06.23	Tobias Paulsen (Sikt)	Draft ready for internal review WP2
0.8	2025.06.25	Deborah Thorpe (DANS), Fernando Aguilar (CSIC), Severine Duvaud (SIB), Philipp Konzett (UiT), Laurence Horton (UEDIN-DCC), Mari Kleemola (TAU-FSD), Jonas Recker (GESIS), Vasso Kalaitzi (DANS), Eduardo Esteves, Robert Huber (UBremen), Clement Jonquet (INRAE), Åse Jorun Holthe-Tveit (Sikt)	Internal review provided
0.9	2025.06.25	Maaike Verburg (DANS), Laurence Horton (UEDIN-DCC), Geert van Geest (SIB), Thomas Jouneau (INRAE), Sanni Tujunen (TAU-FSD), Joel Kallio (TAU-FSD), Tobias Paulsen (Sikt)	All internal review comments incorporated
1.0	2025.06.27	Maaike Verburg (DANS), Thomas Jouneau (INRAE)	Version ready for publication

Table of contents

1 Executive summary	8
2 Rationale and method	10
2.1 Rationale of the survey	10
2.2 Creation of the survey	11
2.3 Dissemination of the survey	12
2.4 Analysis methodology	13
3 Results	15
3.1 Respondent characteristics	15
Retention, curation, and preservation	18
Certification	20
3.2 Repository capabilities and standards	22
Digital Object Management	22
Repository platforms and software	31
Persistent Identifiers (PIDs)	33
Metadata schemas	34
Semantic artefacts	37
Shared services	39
Organisational Infrastructure	41
Licensing	49
Continuity	54
Funding	54
Technology and Security	56
Storage and integrity	61
Other standards, best practices, and solutions for technology and security management	62
3.3 Legal challenges	65
Additional legal questions	67
3.4 Training and support needs	69
3.5 Expectations of the FIDELIS Network	76
4 Conclusions and next steps	80
5 Appendices	84
Appendix A - Survey glossary	84
Appendix B - Additional tables and lists relevant to repository characteristics	91

Disclaimer

FIDELIS has received funding from the European Commission's Horizon Europe funding programme for research and innovation programme under the Grant Agreement no. 101188078. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

List of tables

Table 1. TTRAM Activities and Functions (A F): Deposit & Appraisal	24
Table 2. TTRAM Activities and Functions (A F) : Curation, Quality & Compliance	26
Table 3. TTRAM Activities and Functions (A F) : Discovery & Identification	27
Table 4. TTRAM Activities and Functions (A F) : Access	28
Table 5. TTRAM Activities and Functions (A F) : Interoperability	28
Table 6. TTRAM Activities and Functions (A F) : Preservation	29
Table 7. TTRAM Activities and Functions (A F) : Provenance & Authenticity	29
Table 8. TTRAM A F: User Support	30
Table 9. Other ideas (not explicitly related to digital objects but mentioned)	30
Table 12. TTRAM Activities and Functions (A F) : Mission & Scope	44
Table 13. TTRAM Activities and Functions (A F) : Governance	44
Table 14. TTRAM Activities and Functions (A F) : Resources	45
Table 15. TTRAM Activities and Functions (A F) : Continuity of service	45
Table 16. TTRAM Activities and Functions (A F): People & expertise	46
Table 17. TTRAM Activities and Functions (A F): External engagement	46
Table 17. TTRAM Activities and Functions (A F): Legal & Ethical	47
Table 18. TTRAM Activities and Functions (A F) : Third-party dependencies	48
Table 19. Additional ideas expressed by respondents linked to TTRAM A F	49
Table 20. Licences for metadata	51
Table 21. Licences for data	51
Table 22. Licences for software	51
Table 23. Licences for other digital objects	52
Table 24. Licences for named “other” digital objects.	53
Table 25. TTRAM Activities and Functions (A F): Security	58
Table 26. Sensitive data management	59
Table 27. Tech watch	59
Table 28. TTRAM Activities and Functions (A F) : Technical infrastructure	60
Table 29. Other technology & security challenges	60
Table 10. Purposes mentioned and metadata schemas employed for them	92
Table 11. Purposes of specific catalogues, registries, resources and services mentioned by respondents.	94

List of figures

Figure 1. An overview of the different Pillars of the FIDELIS project.	10
Figure 2. The geographical spread of the survey responses.	16
Figure 3. Overview of the different types of repository in the sample.	16
Figure 4. The domains served by the repositories in the sample.	17
Figure 5. The levels of retention, curation, and preservation provided by the repositories in the sample.	19

Figure 6. The levels of retention, curation, and preservation provided by discipline-specific versus generalist repositories.	20
Figure 7. Certification status and opinion on some formal certification schemas.	21
Figure 8. Challenges and needs that repositories face in terms of digital object management activities and functions.	22
Figure 9. Challenges and needs that institutional repositories face in terms of digital object management activities and functions.	23
Figure 10. Challenges and needs that national repositories face in terms of digital object management activities and functions.	23
Figure 11. Number of respondents having expressed challenges in each topic.	24
Figure 12. Top 10 platforms or software environments used by repositories.	31
Figure 13. Technologies and services used to deploy repository software or platforms.	32
Figure 14. Use of persistent identifiers in repositories for digital objects and metadata elements.	34
Figure 15. The 15 most commonly used or currently being implemented persistent identifiers.	34
Figure 16. Repartition of respondents by number of metadata schemas cited (n=159)	35
Figure 17. Ranking of metadata schemas cited.	36
Figure 18. Most commonly cited purposes for metadata schemas.	36
Figure 19. Implementation of the metadata schemas in the repository.	37
Figure 20. Use of semantic artefacts in repositories.	37
Figure 21. Implementation of the semantic artefacts in the repository.	38
Figure 22. Ranking of catalogues, registries, resources and services most cited (n>1).	40
Figure 23. Purposes most cited (n>1) for catalogues, registries, resources, services, or federated systems.	41
Figure 24. Challenges and needs that repositories face in terms of organisational infrastructure.	42
Figure 25. Challenges and needs that institutional repositories face in terms of organisational infrastructure.	43
Figure 26. Challenges and needs that national repositories face in terms of organisational infrastructure.	43
Figure 27. Repartition of types of licences by category of digital object.	50
Figure 27. State of continuity agreements.	54
Figure 28. Stakeholders of continuity agreements.	54
Figure 29. Duration of sustainable funding for repositories.	55
Figure 30. Main funding sources of repositories.	56
Figure 31. Challenges and needs that repositories face in terms of technology and security.	57
Figure 32. Number of respondents having expressed challenges in each topic.	58
Figure 33. Measures used to ensure the storage and integrity of digital objects and metadata.	61
Figure 34. Main other standards, best practices and solutions for technology and security management.	62
Figure 35. What is the legal status of your repository?	65
Figure 36. What would you say are the biggest legal challenges your repository faces when sharing data between repositories or with scientists from other research organisations?	66
Figure 37. Would your repository be interested in collaborating on frameworks and guidelines regarding GDPR and/or Intellectual Property Rights with others in the Network?	66
Figure 38. Does your repository have clear legal frameworks, policies and guidelines for handling personal data in accordance with GDPR and other relevant data protection laws?	67

Figure 39. Does your repository have clear frameworks and guidelines for sharing data within Europe?	67
Figure 40. Does your repository follow one or several specific legal metadata standards for data governance and compliance?	68
Figure 41. Would your repository be open to change to/adopt/add another core metadata schema if that would improve legal interoperability?	68
Figure 42. Where do you currently (or have you previously) find/receive training on topics related to Trustworthy Digital Repositories?	69
Figure 43. Received (blue) and desired (orange) training and support topics.	71
Figure 44. Received (blue) and desired (orange) training and support formats.	72
Figure 45. Desirability judgement of some training and support characteristics.	73
Figure 46. Judgement of importance of some training and support qualities.	74
Figure 47. How likely do you currently think it is that you would join or apply for each activity?	75
Figure 48. What do you see as the most important benefit(s) this Network should provide to you or your organisation?	76
Figure 49. What format of collaborative mechanisms would you be most likely to join?	77
Figure 50. Expected barriers to joining the FIDELIS Network.	78

TERMINOLOGY

Terminology/Acronym	Description
AI	Artificial Intelligence
API	Application Programming Interface
CESSDA ERIC	Consortium of European Social Science Data Archives
CLARIN ERIC	Common Language Resources and Technology Infrastructure
COAR	Confederation of Open Access Repositories
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DMP	Data Management Plan
ELIXIR	Distributed infrastructure for life-science data
EOSC	European Open Science Cloud
EOSC EDEN	EU Project for Enhancing Digital preservation strategies at European and National level
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
PID	Persistent Identifier
R&D	Research and Development
RDA	Research Data Association
RDM	Research Data Management
TDR	Trustworthy Digital Repository
TTRAM	Transparent Trustworthy Repository Attributes Matrix
TTRAM A F	Activities and Functions, the elements in which the TTRAM is structured. These Activities and Functions are derived from a range of existing standards, criteria, and requirements.

Note #1: Many standards, schemas, and softwares are mentioned in the results. The clarifications and links to these will not be included in full in this report, but rather accompany the data when it is published.

Note #2: Appendix A presents the glossary that accompanied the survey, which presents more information about the terminologies and terms used.

1 Executive summary

The FIDELIS project aims to establish a European Network of Trustworthy Digital Repositories (TDRs) that will support their development and growth, foster harmonisation and interoperability, and strengthen relevant skills through training and support. For this project to successfully achieve its goals, it was deemed essential to gather input from the community on their current practices, challenges, and needs. The outputs of this survey will be used to shape the activities of the project in the next years, and subsequently help shape the FIDELIS Network as it is designed and implemented.

This activity was a collaborative effort across the project, with each part of the project focusing on their own sections of interest. This one report presents the full analysis of the survey results, covering the two associated Milestones (M11 *survey analysis ready* and M19 *Analysis of survey results with regards to support and training*) and additional analyses on other relevant topics.

The survey covered the following topics:

- Repository Characteristics
- Activities & Functions
 - Digital Object Management
 - Organizational Infrastructure
 - Technology & Security
- Training & Support Needs
- Legal Challenges
- Network Interest & Expectations

Responses were obtained from 159 individuals representing 144 different repositories (counting unique names). This report covers all results and findings from the survey. The main findings are considered to be:

- Most repositories offer Deposit Compliance as their level of care to their objects. However, many repositories (especially discipline-specific ones) are currently working on implementing Active Preservation as a more extensive level of care.
- Most repositories show good adoption of common artifacts, already interfaced or interoperable, such as persistent identifiers, commonly adopted metadata schemas (while retaining more confidential, disciplinary or “niche” artifacts where necessary), or licenses. The result is an international repository landscape that is increasingly converging in terms of common standards and practices.
- Lack of funding (particularly in a context where project-based funding is increasingly the norm), and lack of human resources or people with the right expertise, are the most commonly cited challenges. These two big concerns can be linked with another one, the difficulty to keep a repository up to date with the new technologies or the new demands and challenges emerging almost every year.

- While only few respondents seem to rely on third party services for some core services of their repository, many of them expressed this reliance as a challenge rather than an opportunity.
- The survey reveals a fragmented legal landscape, with many repositories lacking formal frameworks for GDPR compliance or cross-border data sharing, responsibility for GDPR compliance is often placed on researchers. The FIDELIS Network can help by promoting shared templates, peer-to-peer learning, and working groups, but success depends on active member participation.
- Topics that are valued most to receive more training and support on are repository certification, long-term preservation, developing and sustaining a repository, access conditions and sensitive data, and stakeholder engagement and communication.
- Other preferences expressed about training and support activities are for them to be practical and specific, have a clear value and return on investment (time and effort), and for participants to have the opportunity to connect with, learn from, and network with fellow participants, organisers, (guest) speakers, and other experts during the activity.
- Respondents showed a general enthusiasm about the potential benefits of the FIDELIS Network. It was suggested that granular and practical work on tangible outcomes that can then be adopted throughout the Network and communicating in a combination of (online) meetings and a forum would be the most effective way to establish the Network. Financial constraints were the most envisioned barrier to participating in the Network.

The results of this survey have provided valuable insights into the current activities, challenges, and needs of the community when it comes to a wide variety of topics, as well as their initial opinions on the FIDELIS project and our aims. The findings will be used to shape the activities of the project, identifying where support and training is needed and in which ways we can bring people together to jointly tackle challenges and advance their developments.

2 Rationale and method

The FIDELIS project aims to establish a healthy, vibrant, and self-sustaining Network of Trustworthy Digital Repositories (TDRs) that will foster a supportive open science environment and guarantee FAIR data sharing also in the future. Within its three-year lifetime, the project will set up, develop, and operate a European Network of trustworthy repositories that will support the development and growth of TDRs within the EOSC ecosystem, foster harmonisation and interoperability across repositories to enable an EOSC federation of TDRs, and strengthen the upskilling of repositories and expansion of the Network through an active training and support programme.¹

2.1 Rationale of the survey

Due to the active and self-sustaining nature of the emerging Network, it was deemed essential that there are close and continuous interactions with the community to help shape the different aspects of the project. The FIDELIS project started in January 2025 and one of the first planned activities was to execute a landscape survey to gain insight into community knowledge, capabilities, and needs as early as possible. This activity was jointly executed by the different pillars in the project, each focusing on their own area(s) of interest that related to (part of) their work description: Pillar 2 (focus on legal challenges), Pillar 3 (focus on repository activities and functions), and Pillar 4 (focus on training and support needs). The survey also included questions of relevance for Pillar 5 (focus on communication needs) and those of relevance across all Pillars (see Figure 1).

	Year 1	Year 2	Year 3
Pillar 1 (CSC)	Project Management (WP1)		
Pillar 2 (Sikt)	Preparing and initiating the network (WP2)	Consolidating the network and value-add (WP3)	Transition the network and value-add (WP4)
Pillar 3 (TAU-FSD+ Sikt)	Building common understanding of TDRs and a harmonized matrix of repository capabilities and characteristics (WP5)	Enhancing maturity and enabling federation of TDRs (WP6)	Recommendations and guidance (WP7)
Pillar 4 (DANS)	Repository training and support: Initiating network cohesion and support (WP8)	Strengthening the network (WP9)	Consolidating the network (WP10)
Pillar 5 (Trust-IT)	Strategic alliance, cascading grants, communication and dissemination (WP11-13)		

Figure 1. An overview of the different Pillars of the FIDELIS project.

The overarching aim of the landscape survey was to gain an understanding of the current capabilities and needs of repositories in the specific areas of focus of each Pillar in the project, and for each Pillar to incorporate these insights in their strategies and planning for the upcoming years. The survey asks questions that identify the areas where support is most needed to tackle the challenges and improve the skills that are currently most pressing for European repositories. The results, therefore, inform

¹ <https://cordis.europa.eu/project/id/101188078>

the efforts and activities of the different Pillars in designing and building the FIDELIS Network in a way that is most attractive to potential members.

As it was important to gain the opinions from as large a group as possible, a survey with widespread dissemination was deemed the most suitable approach for collecting the desired data.

2.2 Creation of the survey

The survey was designed with dedicated subsections for each Pillar and their specific area(s) of focus. Each Pillar worked with their teams to draft the questions relating to their own interests. A dedicated team with delegated members from each Pillar met regularly to work on the overarching questions of interest, streamline the full survey experience, and make methodological decisions. The survey consisted of the following sections:

- Introduction & Consent
- Repository Characteristics
- Activities & Functions
- Training & Support Needs
- Legal Challenges
- Network Interest & Expectations

In the Introduction & Consent section, respondents were asked for consent separately for 1) participating in the survey and consenting to the public sharing of their pseudonymised data, 2) sharing non-anonymised data with EOSC EDEN for potential follow-up activities, and 3) usage of contact information to be invited to the project newsletter once the website had gone live. Consenting to the first was mandatory to participate in the survey, but consenting to the second two was optional. A full copy of the empty survey is added as a separate file to this report to showcase the exact information and questions presented to the respondents.

One of the desired qualities of the survey was to keep it as short as possible. At the same time, there was a desire to obtain a wide scope of information from the different Pillars and their unique perspectives. It was also known and/or expected that other projects would engage similar stakeholders in other outreach activities, including our sister-project EOSC EDEN, to extract their own desired information. To prevent survey fatigue with the same group of respondents, the decision was made to keep everything together in one potentially longer survey, and not split them up into multiple separate shorter surveys on each topic. Because we needed some central insights into the respondents, a small subsection of the questions in the survey were mandatory. The remaining questions were optional: in each section, respondents were explicitly asked whether they wanted to answer the remaining optional questions. This allowed us the opportunity to collect rich additional information from more enthusiastic respondents, albeit from smaller subgroups of respondents.

The questions, sections and the general terminology used throughout the survey followed a core deliverable of the FIDELIS project, the Transparent Trustworthy Repository Attributes Matrix (or

TTRAM)². Some parts of the results presented in chapter 3 are organised with the intention to follow the logic and the general layout of the TTRAM, to present a coherent picture.

The core survey took an estimated 20 minutes to respond to, with the additional optional questions adding up to an estimated 20 more minutes of response time. However, due to the inclusion of a significant number of open questions, the durations were expected to vary quite a lot.

The platform used for the survey creation and dissemination was EU Survey.³ To limit the personal information gathered in the survey process, in line with the general principle of data minimisation in the GDPR,⁴ it was decided to use the EU Survey 'Anonymous mode'. Therefore, it was not possible to analyse the response durations to confirm our estimated duration of the survey or to investigate response validity based on time.

For added information and clarification, the survey was accompanied by a glossary (Appendix A) which explained the definitions used in the survey to the respondents. These definitions were shared in a separate document to be consulted alongside the survey when responding.

2.3 Dissemination of the survey

The survey launched on March 11th 2025. The initial closing date of April 18th was extended to April 25th to allow responses to be finalised around some common European holidays. Dissemination of the survey consisted of several different 'waves' of effort. First, the survey was immediately disseminated in the FIDELIS consortium and the EOSC EDEN consortium for responses and further dissemination. Next, the FIDELIS External Advisory Board was contacted, as well as the FIDELIS 'early adopters' identified during the project proposal phase. All these close contacts were invited to disseminate and promote the survey in their own networks, such as the closely related CESSDA, ELIXIR, and CLARIN consortia.

The joint EOSC EDEN and FIDELIS website was launched during the period of dissemination and could therefore also be used to promote the survey to initial visitors of the website⁵. The last wave of wider dissemination was done through a mass export of over 1500 repository records from re3data⁶ containing repositories across Europe. Members of the different FIDELIS Pillars identified their peers in that list to contact for a personal invite to the survey. The remaining repositories on the list were contacted with a generic invite. This strategy aimed to reach as many repositories as possible, inside and outside of the known networks and connections of the project team.

² L'Hours, H., Kleemola, M., Parkes, O., Recker, J., Duvaud, S., van Horik, R., Alaterä, T. J., Liberante, F., Conzett, P., Kaartinen, H., Bäckman, S., & Esteves, E. (2025). FIDELIS TTRAMatrix v00.02 Guide (v01.00). Zenodo.

<https://doi.org/10.5281/zenodo.15676189>

³ <https://ec.europa.eu/eusurvey/home/welcome>

⁴ https://www.edps.europa.eu/data-protection/data-protection/glossary/d_en

⁵ <https://eden-fidelis.eu/news/fidelis-repository-landscape-survey-available>

⁶ <https://www.re3data.org/>

The main initial focus of the FIDELIS project is the EOSC community, and therefore the main focus of the dissemination strategy, was European repositories or those from Horizon Europe-associated countries. However, the decision was made to allow responses from outside of Europe as well, as topics of harmonisation and support are also universally shared. The FIDELIS External Advisory Board also consists of international members outside of Europe, whose perspectives, insights, and larger networks we greatly value and wished to include in our analyses. Given the survey only reached a part of the community, and that the sample was skewed to more “EOSC-ready” countries, it will be important to continue to consider other perspectives in the community as the FIDELIS Network is developed.

Defining repositories is one of the topics FIDELIS will continue to develop thoughts around as the project advances. To be as inclusive as possible at this early scoping stage, the survey used a working definition of ‘repository’ to encompass all services that manage research objects (data, software, literature or other), regardless of whether they are labelled as a repository, archive, e-depot or anything else. Because repositories function in a wider ecosystem of service providers that can greatly influence the challenges experienced and support needs, other data service providers and registries were also welcome to participate in the survey, though they were informed that some questions could be less relevant to their organisation, operations, and needs.

2.4 Analysis methodology

As each section in the survey was designed for the purpose of a specific Pillar, each Pillar was responsible for the analysis of their own section of the survey results. Each Pillar selected members of their team to be involved in the analysis process, who were given access to the protected dataset cleaned for analysis. The designated members met regularly to discuss the alignment of methodological approaches and decisions, while working collaboratively to create this report.

For the questions related to respondent characteristics (section 3.1) and needs for training and support (section 3.4), R statistical software⁷ was used. No data cleaning was performed. Visualisations were generated with the packages tidyverse⁸, ggplot2⁹, and maps¹⁰. In the case of multi-response questions (in which respondents can select multiple answers), the counts of responses were added up per category. To report the total number of respondents answering the question, the number of respondents is added to the plot.

⁷ R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

⁸ Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686.

⁹ Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.

¹⁰ <https://cran.r-project.org/web/packages/maps/index.html>

For the questions related to repository characteristics (section 3.2) and legal challenges (section 3.3), IBM SPSS Statistics was primarily used for the close-ended questions (where the options were limited), and Microsoft Excel for the open-ended questions. SPSS was used to convert variables into numeric format and to separate textual dichotomies encoded within a single variable into distinct numeric variables. It was also used to process open-ended matrix responses by aggregating all listed items into one variable and their associated descriptions of use into another. Subsequent analysis of these aggregated variables was conducted in Excel. In this section, four kinds of open-ended questions were present.

- For the open-ended questions following controlled ones, a simple, literal analysis was conducted.
- For the open-ended questions following the controlled matrices about encountered challenges, the ideas expressed in the answers were encoded and regrouped whenever possible, taking care not to oversimplify them. These ideas were then aligned to the categories of the matrix to get a focus on each of the topics.
- For the free text matrices, both dimensions (the items, such as metadata schemas, catalogues and registries, etc.; and the purposes) were encoded and regrouped whenever possible. Tables and graphs were then derived from the encoded data.
- For the open-ended questions about licences, the answers were encoded, and tables and graphs were derived from the data.

3 Results

The survey received a total of 159 valid responses, defined as responses to all mandatory questions in the survey. For its optional additional questions, fewer respondents were willing to answer the additional questions as the survey advanced: 92 for digital object management, 78 for organisational infrastructure, 63 for technology and security, and 54 for legal challenges. This suggests some fatigue over the course of the longer survey. As can be expected, open questions generally received a smaller number of responses than multiple choice questions and several open questions in a row showed a decrease in response rates as well.

During the analysis we assumed responses were independent, and there was absence of bias, meaning that we assumed participants understood the questions and they answered them truthfully. To detect possible biases, we investigated the similarity between responses. We observed groups of responses originating from the same country that showed a high similarity amongst the full range of multiple choice questions. Because it is difficult to conclude whether these responses are highly similar due to chance or unknown bias, and because presence of response biases are typically necessary trade-offs in surveys like this, we decided to not directly correct for possible biases in the analysis, while still mentioning and analysing, in the redaction, the occurrences where multiple, similar answer would introduce such biases.

The following sections present the results of the survey per topic. The pseudonymised data and the analysis scripts will become available for reuse as well.

3.1 Respondent characteristics

The 159 responses to the survey came from 27 different countries across Europe (see Figure 2). Seven responses came from countries selected as 'Other', namely Horizon Europe-associated countries in Armenia, Canada, Tunisia, and two other countries such: Uganda, and the United States of America. It can be observed that the eastern part of Europe is less represented in the sample, with the notable exception of Serbia.

Most responses came from institutional, national, or public repositories (see Figure 3), but other repositories and services were also represented, for example those covering multiple countries or institutions. Most repositories in the sample have existed for around 6-10 years ($n = 46$), with a notable skew in the sample to older ages (35 repositories of 11-20 years, and 28 repositories older than 20 years), but also representation from younger repositories (35 repositories 3-5 years, 9 repositories 1-2 years, 4 less than one year, and two still developing).

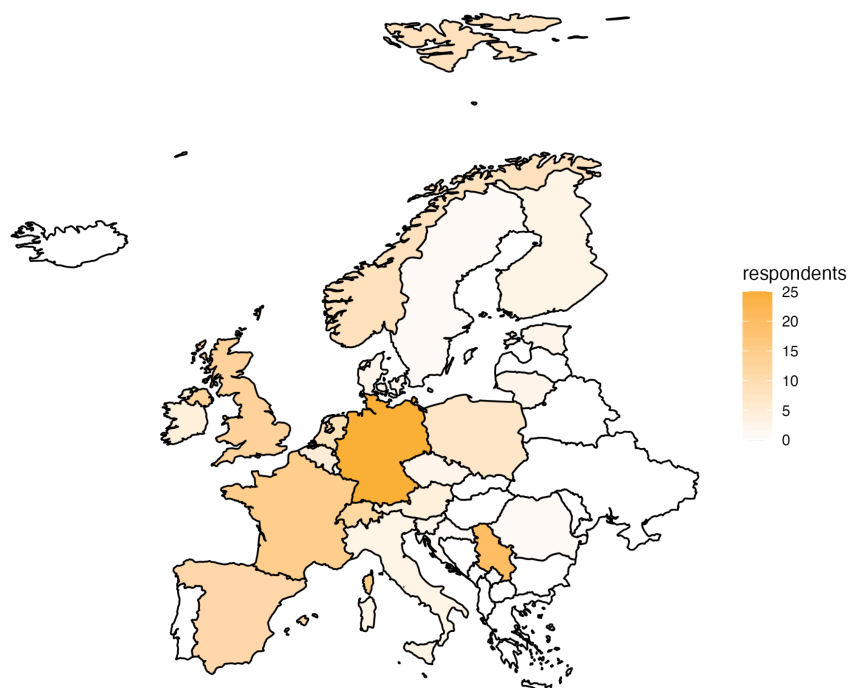


Figure 2. The geographical spread of the survey responses.

Responses from countries outside of Europe: Armenia: 1, Canada: 2, Tunisia: 1, Uganda: 1, United States: 1.

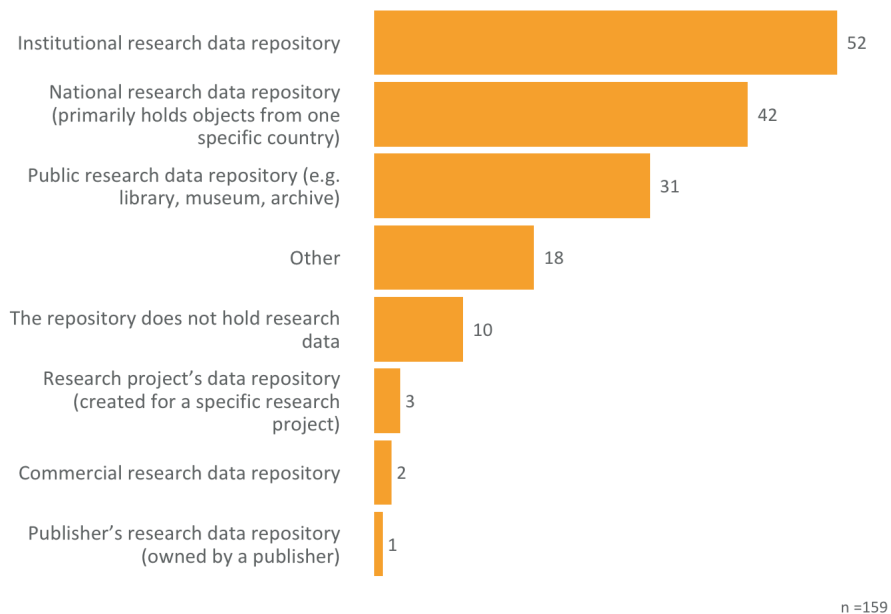


Figure 3. Overview of the different types of repository in the sample.

The largest portion of responding repositories identified as a discipline-specific repository serving one or more specific domains (n = 105, compared to 54 generic repositories), with most of them serving the domain of the Natural Sciences (see Figure 4) and all domains as defined in the survey (according to the Frascati fields of R&D classification¹¹) represented to some extent. Most repositories hold data as their primary content type (n = 139), with literature (n = 42), software/code (n = 30), and other content types (n = 16) much less represented. Respondents most often indicated their role(s) as being repository manager (n = 126) or curator (n = 55). It can be challenging to capture job descriptions with generic titles, and respondents were able to indicate all roles that they felt matched their work. In general, we received some different perspectives in the survey responses, also including those of board members and system engineers.

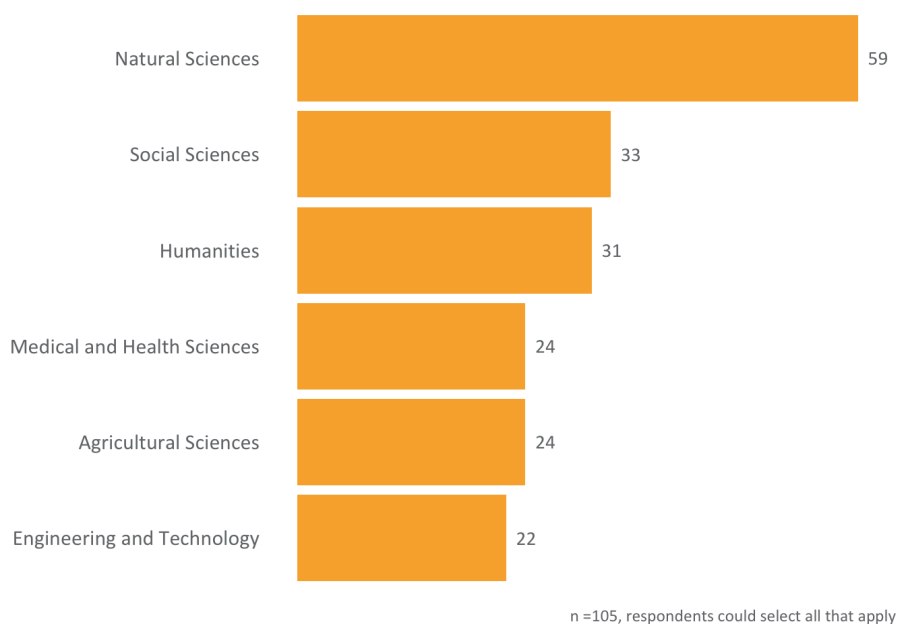


Figure 4. The domains served by the repositories in the sample.

Respondents were asked to include all the persistent identifier(s) their repository can be identified by. As a result, we collected 83 RORs¹², 110 re3data identifiers¹³, 74 FAIRsharing DOIs¹⁴, and a handful of other identifiers, such as OpenDOAR¹⁵ and ISNI¹⁶. This indicates that although the identifiers are in good use, there is still some improvement that can be achieved for repository findability when it

¹¹ OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, <https://doi.org/10.1787/9789264239012-en> (p.59)

¹² <https://ror.org/>

¹³ <https://www.re3data.org/>

¹⁴ <https://fairsharing.org/>

¹⁵ <https://opendoar.ac.uk/>

¹⁶ <https://isni.org/>

comes to using persistent identifiers as well as to synchronise the information presented across the different schemas.

Respondents were asked which communities they consider their repository to be a part of. The communities mentioned most often were RDA, ELIXIR, CESSION, CLARIN, and DARIAH. Other notable communities mentioned were COAR and EOSC. This information is important to take into account when designing a new community for repositories. How can the FIDELIS Network be a unique addition to the kind of membership and support that is already available to the targeted repositories?

Retention, curation, and preservation

Respondents were asked about the levels of retention, curation, and preservation provided by the repository. The levels defined were in line with the ones proposed in the position paper from the CoreTrustSeal board¹⁷:

- **As deposited** - Digital object content and supporting metadata are distributed to users exactly as they are provided by depositors.
- **Deposit compliance** - Digital object content and supporting metadata deposited are checked for compliance with defined criteria, e.g. digital object formats, metadata elements, and compliance with legal and ethical norms.
- **Initial curation** - The digital objects are curated by your repository to meet defined criteria, which may exceed those defined for Deposit Compliance.
- **Active preservation** - In addition to Deposit Compliance and Initial Curation, your repository takes long-term responsibility for ensuring that the digital objects and metadata can be understood and rendered as required by the designated community for reuse.

Repositories were asked to indicate for each level whether it was not provided, planned to be provided in the future, in the implementation phase to provide, implemented to provide, or whether the respondent was unsure about this information. For most levels respondents indicated that it was currently implemented and thus actively provided by their service. The ‘as deposited’ level of retention was most often indicated as not being provided at all (see Figure 5). Active preservation was indicated as planned (n = 20) or in implementation (n = 17) the most of all levels, suggesting this level of care is being seen as desirable by more repositories or their stakeholders.

¹⁷ CoreTrustSeal Standards and Certification Board. (2024). Curation & Preservation Levels: CoreTrustSeal Position Paper. Zenodo. <https://doi.org/10.5281/zenodo.11476980>

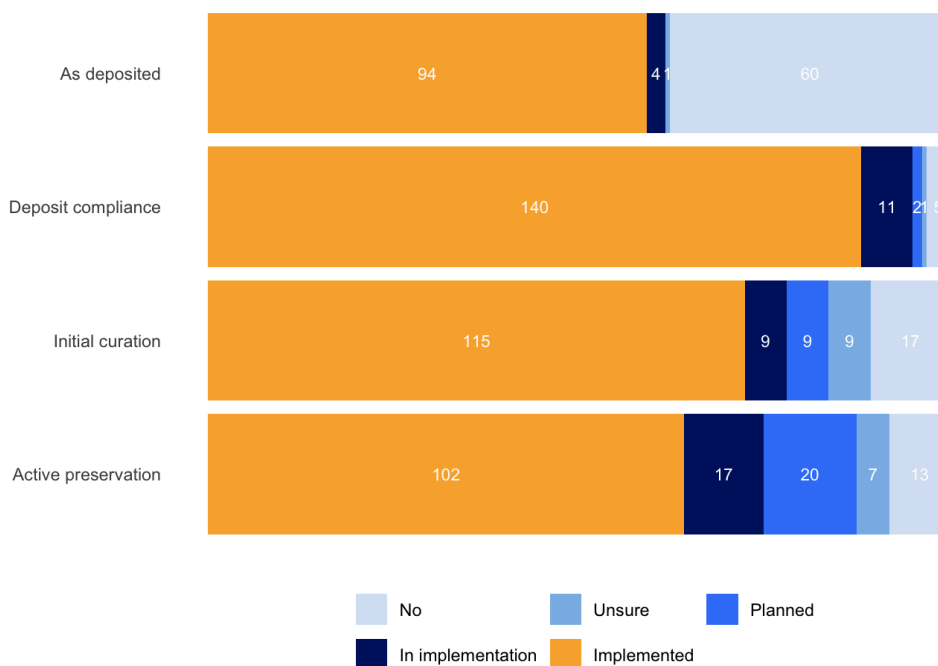


Figure 5. The levels of retention, curation, and preservation provided by the repositories in the sample.

Considering the two main types of repositories (discipline-specific and generalist), it can be seen in Figure 6 that all levels of retention, curation, and preservation are most often already implemented in either type of repository. Two noticeable observations are that discipline-specific repositories seem to not implement ‘as deposited’ level of retention more often, and are more often working on offering active preservation than generalist repositories. This may help identify more specific audiences for certain topics of training or harmonisation.

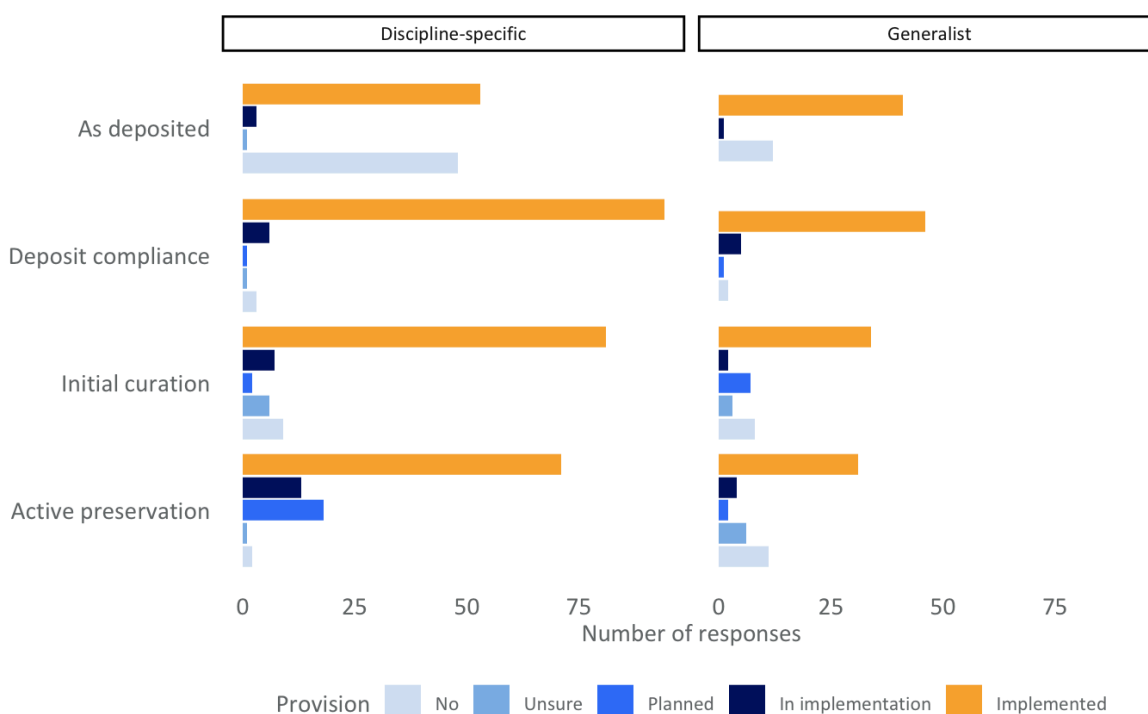


Figure 6. The levels of retention, curation, and preservation provided by discipline-specific versus generalist repositories.

Certification

Respondents were asked about four formal certifications: CoreTrustSeal¹⁸, Nestor Seal (DIN31644)¹⁹, ISO 16363²⁰, and ISO 27001²¹. For each certification, one of the following options could be indicated:

- Do not have, not interested in obtaining
- Do not have, interested in obtaining
- In progress of achieving
- Currently certified
- Expired, in process of renewing
- Expired, not planning to renew
- Do not know /unsure.

The largest part of respondents indicated not being interested in obtaining the Nestor Seal or either of ISO certifications, or being unsure about them. The exception was CoreTrustSeal, with more

¹⁸ <https://www.coretrustseal.org/>

¹⁹ https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html

²⁰ <https://www.iso.org/standard/87472.html>

²¹ <https://www.iso.org/standard/27001>

interest in obtaining the certification and also the most repositories indicating having achieved the certification or being in the process of (re)certifying (see Figure 7).

Besides these four included formal certification schemes, respondents could also indicate other certifications that their repository has or that are of specific relevance or interest to their repository. Some recurring mentions include the community mandates, guidelines, and services, such as the ELIXIR Deposition Database, CESSION Service Provider obligations, and CLARIN B-centre, as well as national initiatives such as the German DINI²², and global resources such as RDA and OpenAIRE guidelines.

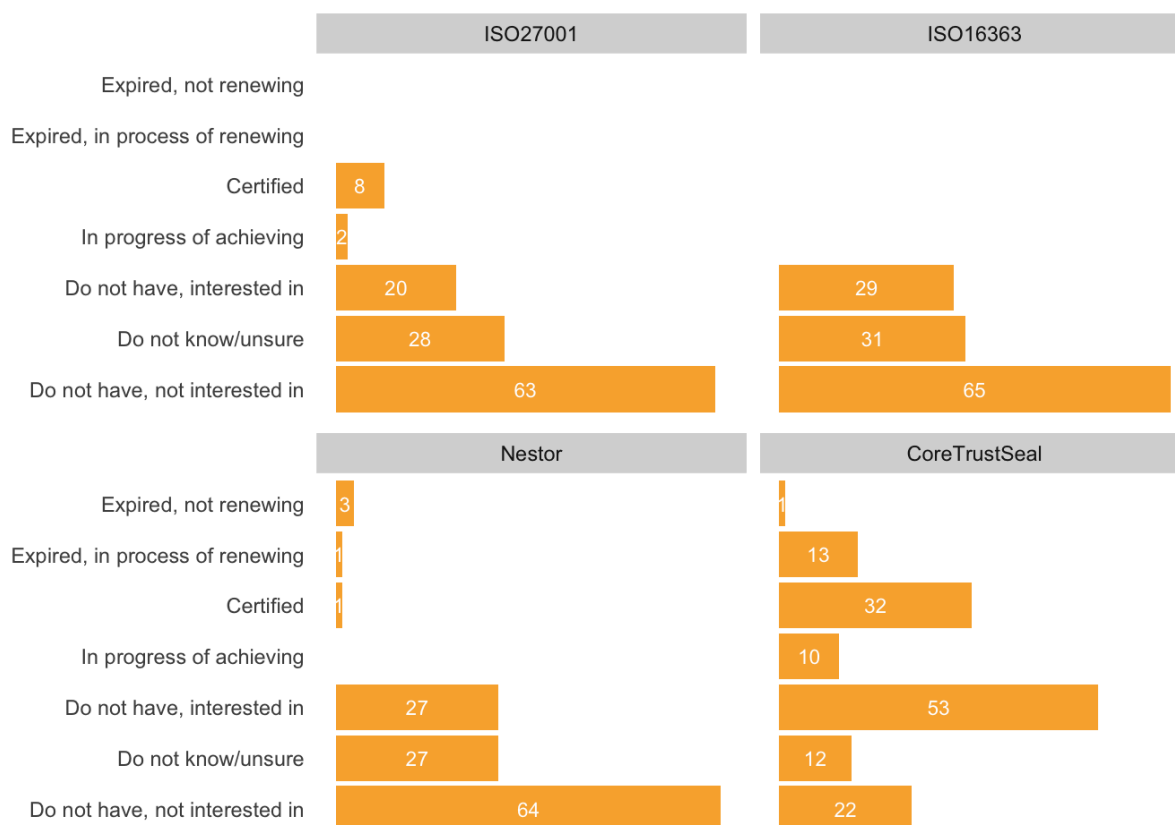


Figure 7. Certification status and opinion on some formal certification schemas.

²² <https://dini.de/dienste-projekte/dini-zertifikat/english/about-the-certificate>

3.2 Repository capabilities and standards

The three subsections of this part of the survey were presented in a similar way: an indication of the challenges and needs, followed by a series of additional questions which were only shown if the respondent agreed to answer them²³.

Digital Object Management

Challenges and needs

Respondents were first asked to indicate the difficulties and needs they encounter in a controlled matrix: on the y-axis, a list of eight categories derived from the TTRAM (Transparent Trustworthy Repository Attributes Matrix)²⁴; on the x-axis, a 5-point scale ranging from “no challenges” to “critical challenges”. Then, an open-ended question (“What are the major challenges and needs your repository faces in terms of digital object management?”) allowed them to identify key challenges and provide more details, if they wished.

For each of the eight TTRAM Activity and Functions (A|F) Figure 8 shows the distribution of responses on a five-point-scale, illustrating the respondents’ assessment of the severity of the challenge that the particular A|F poses.

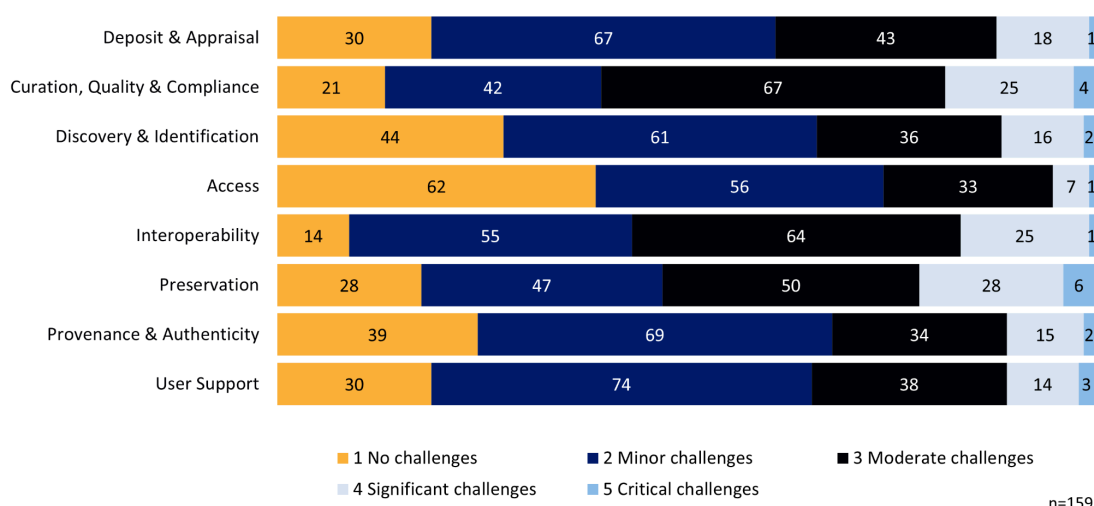


Figure 8. Challenges and needs that repositories face in terms of digital object management activities and functions.

²³ It had been considered to systematically aggregate the multiple entries from one repository. This has however been abandoned because of the mixed nature of the questions: technical, so repository-focused (such as the metadata schemas or PIDs used); and subjective, so respondent-focused, such as the questions about challenges. In the figures, the numbers represent the respondents, except otherwise specified, and for the whole 3.2 section.

²⁴ L'Hours, H., Kleemola, M., Parkes, O., Recker, J., Duvaud, S., van Horik, R., Alaterä, T. J., Liberante, F., Conzett, P., Kaartinen, H., Bäckman, S., & Esteves, E. (2025). FIDELIS TTRAMatrix v00.02 Guide (v01.00). Zenodo. <https://doi.org/10.5281/zenodo.15676189>

Clearly, three challenging activities and functions are emerging from Figure 8: “Curation, Quality & Compliance”, “Interoperability” and “Preservation”. Without overinterpreting these results, it is possible that the least problematic activities and functions (namely, “Discovery & Identification”, “Access”, and “Provenance & Authenticity”), are those for which commonly used repository software tools provide built-in support, such as metadata schemas, DOI minting, or harvesting capabilities. It also could be linked to the lack of expert human resources found across the survey.

A simple cross tabulation was also conducted, offering a focus on institutional (Figure 9) and national (Figure 10) repositories. These focuses do not significantly differ from the global picture, with “Curation, Quality & Compliance”, “Interoperability” and “Preservation” remaining the primary issues. It is however worth noting the even higher degree of difficulty attributed to interoperability issues for national repositories, compared with the general graph.

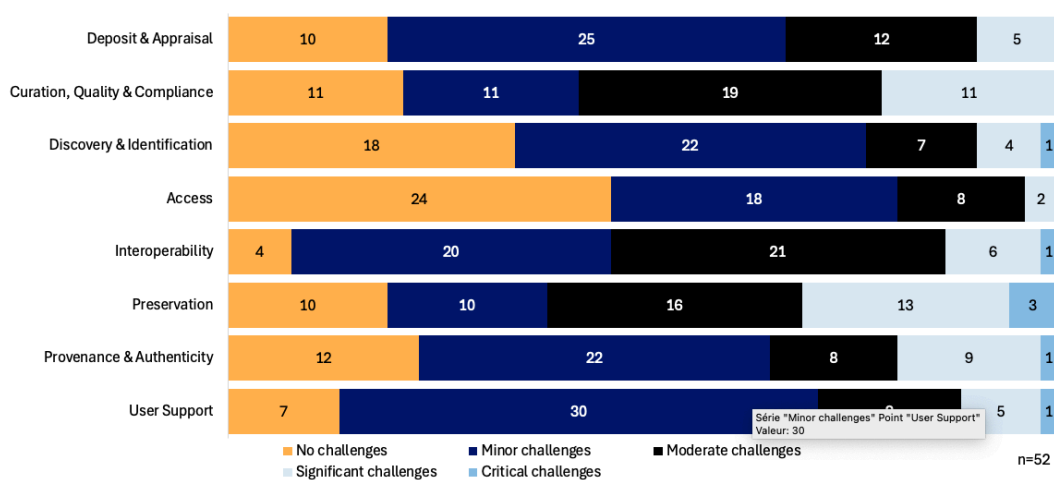


Figure 9. Challenges and needs that institutional repositories face in terms of digital object management activities and functions.

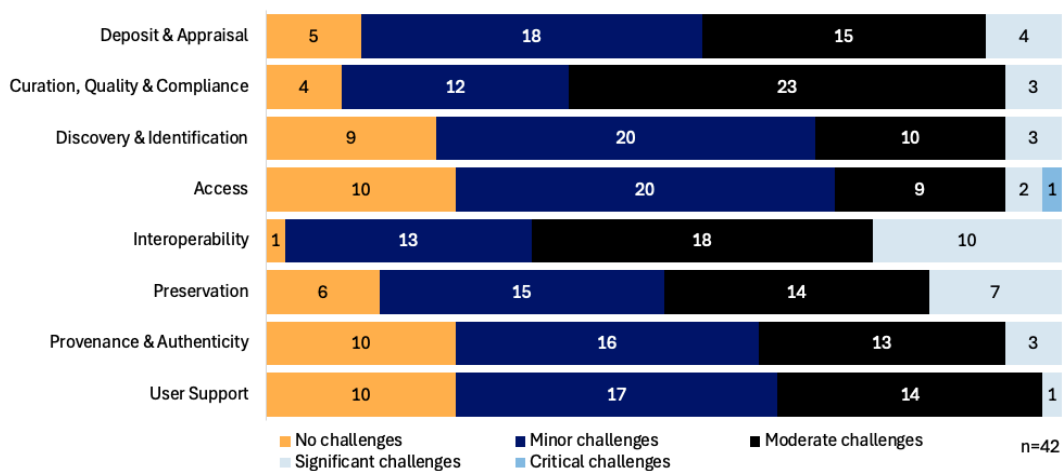


Figure 10. Challenges and needs that national repositories face in terms of digital object management activities and functions.

Figure 11 shows the result of mapping the responses to the open-ended question to the eight TTRAM A|F also applied above. These 78 answers are rich and deserve a deeper dive. Even if the focus of the question is operational or technical, we find already a number of concerns that appear throughout the survey: limited human resources for curation, preservation, and user support; insufficient automation and integration of tools; and the need for recognised guidelines across deposit, appraisal, and preservation workflows. Metadata quality, standardization, and richness are recurring concerns, particularly in enhancing discovery and long-term usability. Legal topics (e.g., GDPR, licensing) and access to high-volume or sensitive data also are present.

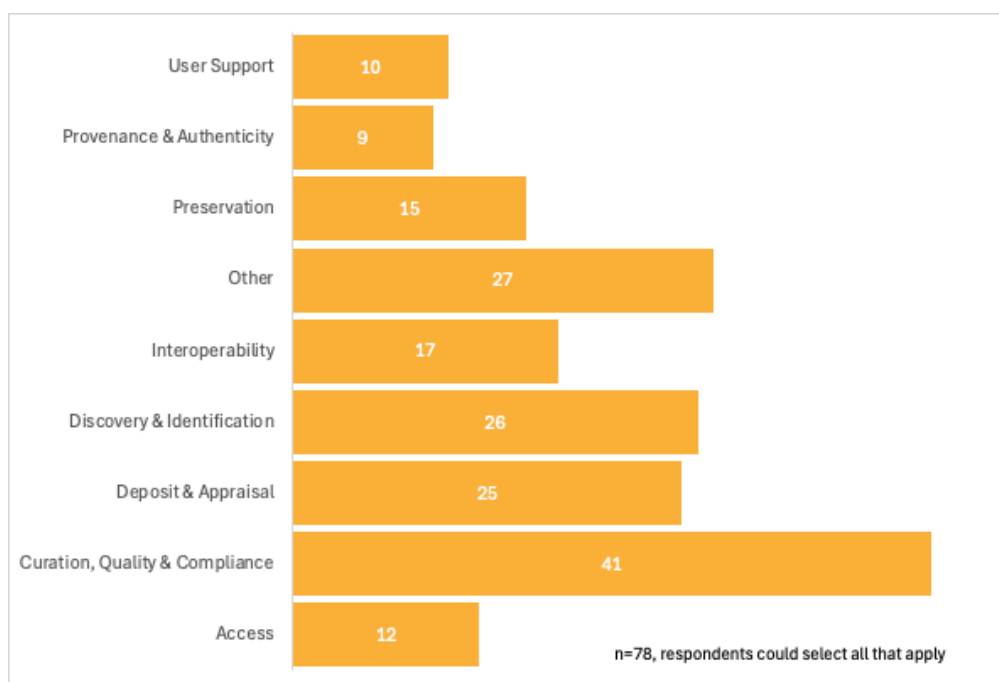


Figure 11. Number of respondents having expressed challenges in each topic.

For each of the TTRAM A|F considered here, Tables 1-9 show the number of mentions of different categories of ideas and challenges induced from the responses. In addition, other TTRAM A|F are cross-referenced if it is relevant to the expressed ideas and challenges. In the subsequent comments and key points, the number of mentions for each idea is indicated in brackets.

Table 1. TTRAM Activities and Functions (A|F): Deposit & Appraisal

Idea expressed	Mentions	Other TTRAM A F
High volume data	4	Storage & Integrity
Automated support during deposit	4	

Idea expressed	Mentions	Other TTRAM A F
Recognised guidelines for appraisal	2	
Legal: GDPR	2	Legal & Ethical
Capacity to expose machine-friendly deposit guidelines	2	
Automated mechanisms for appraisal (during the deposit)	2	
Support during deposit: legal appraisal	1	
Data: focus on raw data while user interest is in derived data	1	
Metadata: lack of user expertise or attention	1	
Data complexity	1	
Lack of documentation for the end user	1	Training
Interoperability with platforms and tools used by researchers	1	Interoperability
Lack of manifest to describe "completeness"	1	
Lack of user expertise	1	Training, External engagement
Learning how to deal with qualitative and relational data	1	
Metadata user guidance	1	
No appraisal because of repository missions to accept all long-tail data from the institution	1	Mission & Scope
Scalability (ever growing datasets)	1	
Support depositors and make them keep the data consumer in mind	1	User training

Key points:

Even if the lack of user expertise is only cited explicitly once, most remarks regarding deposit seem to address this problem indirectly:

- While the most traditional way to address it, user documentation, is also mentioned only once, other ideas emerge. Several respondents suggested integrating **automated solutions during the depositing process**, be it for communicating data appraisal criteria (2), metadata completion (4), or more traditionally by integrated guidance (1).
- The need to **make the depositors aware of the needs of the user** who is going to receive or download the data is interestingly raised (1).
- To ease the deposit, a **better interoperability with platforms and tools used by researchers** (such as automated dataset creation) is also suggested.
- The problems raised by **high volume data** are naturally met at that "early" stage (4) as well as those induced by **complex data** (1) or **legal matters** (1).
- Some challenges can be **repository-specific**: one respondent mentions the scope of their institutional repository (long-tail data from the institution) does not allow them to be too picky regarding the data uploaded.

- An isolated but interesting remark concerns the **gap between the focus of one of the repositories and the expectations of its users**: raw vs derived data, questioning the mandate and acceptance policy.
- Lastly, two respondents outlined the necessity to have **global, recognised guidelines for appraisal**, as well as a collective effort towards the exposition of **aligned, machine-readable guidelines at the repository level (2)**.

Table 2. TTRAM Activities and Functions (A|F) : Curation, Quality & Compliance

Idea expressed	Mentions	Other TTRAM A F
Curation	7	
Human resources: curators	7	People & Expertise
Metadata quality and curation	5	
Metadata: better standardization of metadata schemas	3	
Curation across a federated network	3	Governance, External engagement
Automated mechanisms for curation	2	
Curation: tools integrated in the repository software	2	
Lack of training in partner research organizations	2	Training, External engagement
Recognized guidelines for curation	2	
Human resources: quality control	1	
Lack of community standards	1	
Lack of curator expertise	1	People & Expertise
Legal: GDPR	1	Legal & Ethical
Link to code and software	1	Interoperability
Link to publications	1	Interoperability
Quality certification	1	Analysis & Impact
Standards unclear	1	
Data complexity	1	

Key points:

The number of mentions highlight that curation remains a big challenge, with an explicit emphasis on the **lack of human resources (7)** which reappears often in the survey, and a direct link to metadata curation (5).

- Here also repository software improvements could help: mentions of automated tools (2) and better software integration (2) underline the demand for more efficient systems.

- Also outlined are the lack of recognised guidelines (2), difficulties in federated curation with different stakeholders (3) and/or insufficient training in partner institutions (2), as well as scattered concerns about **quality control** (1), **curator expertise** (1), and **absent community standards** (1).
- Beyond curation, some respondents argue for a **harmonisation of metadata schemas** (3). **Linking data to software** (1) and **publications** (1) could be a quality as well as a discovery topic. Legal compliance under GDPR (1), unclear standards (1), and the challenge of achieving quality certification (1) further complicate repository management.

Table 3. TTRAM Activities and Functions (A|F) : Discovery & Identification

Idea expressed	Mentions	Other TTRAM A F
Metadata: richer metadata	4	
Harvesting: facilitate harvesting by external catalogues	3	
Metadata: richer metadata (for files)	3	
Datacite: more granular resourceTypes in the DataCite metadata schema to distinguish datasets from data files	2	
Separate DOIs for versions (lacking in Dataverse)	2	
Improving search and discovery in the repository	1	
Link to original materials	1	
Metadata	1	
Metadata: domains and data types	1	
Metadata: multilingual metadata	1	
PIDs: improve internal expertise	1	People & Expertise
PIDs: alignment policies	1	
PIDs: support in every concerned field in the repository software	1	
Semantic: vocabularies and thesauri	1	
Version management	1	

Key points:

Challenges are reported at the repository level and at a larger (community, global) scale.

- **At the repository level**, better practices may be impeded by the **repository software limitations**: the wish for **richer descriptors** (multilingual, file metadata, domains and data types); **version management**; **separate DOIs for versions** (a known limitation of the Dataverse software for instance); **improved support for PIDs** (with controlled fields).
- **On a larger scale**, two repositories point to a limitation **in the DataCite schema** (datasets and data files being undistinguished). Concerns are expressed regarding **harvesting facilitation** (3) or **PID alignment** (1). **Human resources** are also mentioned implicitly in relation to PID expertise.

Table 4. TTRAM Activities and Functions (A|F) : Access

Idea expressed	Mentions	Other TTRAM A F
High volume data: access	4	
Sensitive data: access	3	
Storage: interoperability with external (national...) storage	2	
Protect the rights of the owners while opening the data	1	Legal & Ethical

Key points:

Again here are some common and interesting problems.

- High-volume and sensitive data share a common technical challenge in providing access: infrastructure strain for the former, restriction management for the latter.
- Interfacing with an external infrastructure is another one.
- For the data without a clear open license, protecting the rights of data owners while still enabling data access creates tension in a landscape dominated by open data. Repository softwares do not always offer readily available solutions to manage fine-grained access to data.

Table 5. TTRAM Activities and Functions (A|F) : Interoperability

Idea expressed	Mentions	Other TTRAM A F
Interoperability: Integration with DMP tools	4	
Semantic: vocabularies and thesauri	2	
Data interoperability (connect code to data)	1	
Interoperability: connect to other services for automated data transfer	1	
Interoperability: file formats	1	
Interoperability: improve interoperability	1	
Interoperability: too many standards	1	
Data interoperability (at the data level)	1	

Key points:

- Repositories could or should aim to better integrate with Data Management Plan (DMP) tools and external services.
- Semantic interoperability challenges arise due to diverse vocabularies and lack of standard alignment.
- Connecting data to code and managing varied file formats present technical interoperability issues.

- Too many competing standards hinder consistent implementation and cross-platform compatibility.

Table 6. TTRAM Activities and Functions (A|F) : Preservation

Idea expressed	Mentions	Other TTRAM A F
Human resources: preservation	3	People & expertise
“Long term preservation”	2	Preservation
Metadata: PREMIS support	2	
Preservation: file formats	2	
Preservation: funding	1	Resources
Preservation policy: what to keep?	1	Mission & scope
Recognized guidelines and standards for preservation	2	
Recovery of old data	1	

Key points:

- Limited staff and funding affect long-term data preservation capacity and sustainability.
- Support for standards like PREMIS and preservation-friendly formats is not consistently implemented.
- Repositories lack clear policies on what to preserve and struggle with recovering legacy data.

Table 7. TTRAM Activities and Functions (A|F) : Provenance & Authenticity

Idea expressed	Mentions	Other TTRAM A F
Data traceability: at a more granular level (across the repository activities and functions)	2	
Data traceability: lack of provenance citation	2	
Data traceability: lack of provenance in legacy data	1	
Heterogeneity in data coming from other repositories	1	
Lack of clear minimum criteria or ideal practice for provenance and authenticity metadata.	1	
Legal: restrictive conditions when license is not clear (German law)	1	Legal & Ethical

Key points:

- Repositories face challenges in tracing data across workflows and citing provenance correctly.
- Legacy data often lacks provenance metadata; external data varies in format and completeness.

- Unclear standards and legal ambiguity limit consistency in traceability and provenance, crucial in a scientific process.

Table 8. TTRAM A|F: User Support

Idea expressed	Mentions	Other TTRAM A F
Human resources: user support	3	People & expertise
Lack of user expertise	1	Training External engagement
Researcher reticent to open data	1	
User support	1	

Key points:

- Repositories report insufficient staff to provide adequate user support services.
- Users often lack expertise in data management or hesitate to share their data openly.
- Basic support needs remain unmet, creating gaps in data submission and reuse guidance.

Table 9. Other ideas (not explicitly related to digital objects but mentioned)

Idea expressed	Mentions	Other TTRAM A F
Funding	7	Resources
Evolutions of the repository to keep it current	5	Continuity of service
Human resources	3	People & expertise
Sustainability	2	Continuity of service
Community engagement and outreach	1	External engagement
Development: agility	1	People & expertise
Human resources: infrastructure	1	People & expertise
Human resources (development)	1	People & expertise
Legal issues: when a partner can't be contacted	1	Legal and ethical
Licensing for reuse	1	Legal and ethical
Other: "Allowing commercial RDM repositories to be recognized as a repository."	1	Mission and scope
Relations with the community	1	External engagement

These ideas are also developed below. The consistency with which they appear is an indicator of their relevance.

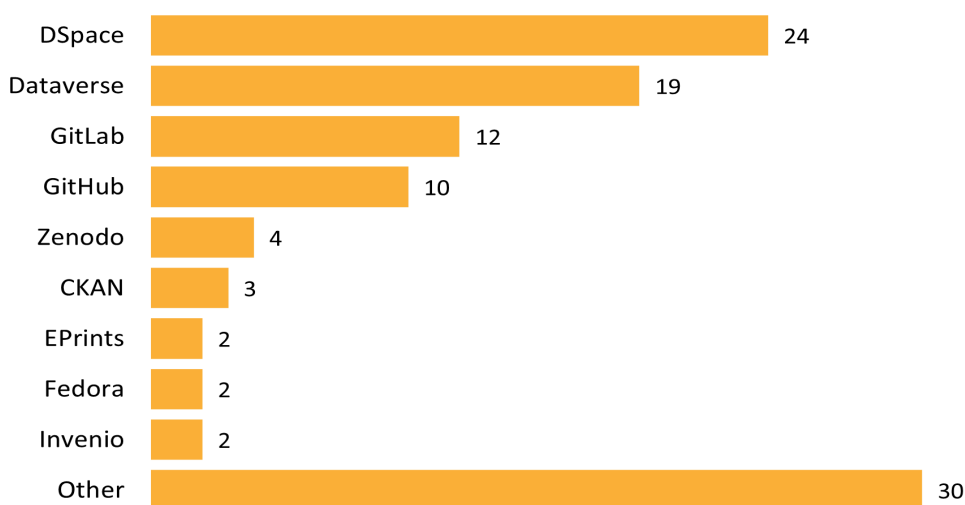
- Funding limitations and staffing shortages affect repository operations and long-term planning.
- Repositories face pressure to evolve and stay current while managing technical and legal complexity.
- Sustainability concerns, community engagement, and licensing issues highlight broader systemic challenges.

Additional questions regarding digital object management

After the two questions regarding challenges and needs, respondents were asked: “*These were the core questions on digital object management. Would you like to answer some additional questions or continue to the next part of the survey?*”. 91 respondents checked “Yes”, triggering the following questions.

Repository platforms and software

Here the question was “*On what platforms/software does your repository operate?*”. Respondents were invited to select all relevant options from a comprehensive list. An “Other” option was also provided, allowing them to specify additional platforms or software in a free-text field.



n=84, respondents could select all that apply

Figure 12. Top 10 platforms or software environments used by repositories.

As Figure 12 shows, DSpace and Dataverse are on the top of the list. DSpace’s position covers actually the answers from Serbian repositories. The Github/Gitlab’s high-ranking may be surprising. However, as multiple answers were allowed, they sometimes came in association with other solutions. Indeed, we notice combinations such as “arXiv; Bitbucket; Figshare; GitHub; Zenodo” or “Dataverse; Fedora;

GitHub; GitLab; Zenodo” in responses. The 30 “Other” mentions come on their own (18) or in conjunction with other solutions (12). For the latter, the entries fall into the shared tools or infrastructure such as Microsoft Azure²⁵, Nextcloud²⁶, or EMBL-EBI²⁷ for example. The former focuses more on the dedicated repository software:

- seven custom (in-house) developments, mostly based on open source software (except one)
- nine specific solutions (OntoPortal²⁸ ; Reposis²⁹ ; Samvera Hyrax or Samvera custom build³⁰ (2); Bespoke; ICAT³¹; a commercial AWS Infrastructure; STAR; GeoNetwork³²).
- a mention of NodeJS which seems out of scope.

Only one respondent mentions Figshare³³ as a repository solution in this particular question. However, Figshare is mentioned by at least two respondents in other questions (on organisational challenges).

The next question asked was “What technologies or services do you use to deploy the platform/software on which your repository operates?” -and was a multiple choice question offering eight options. The responses are summarized in Figure 13.

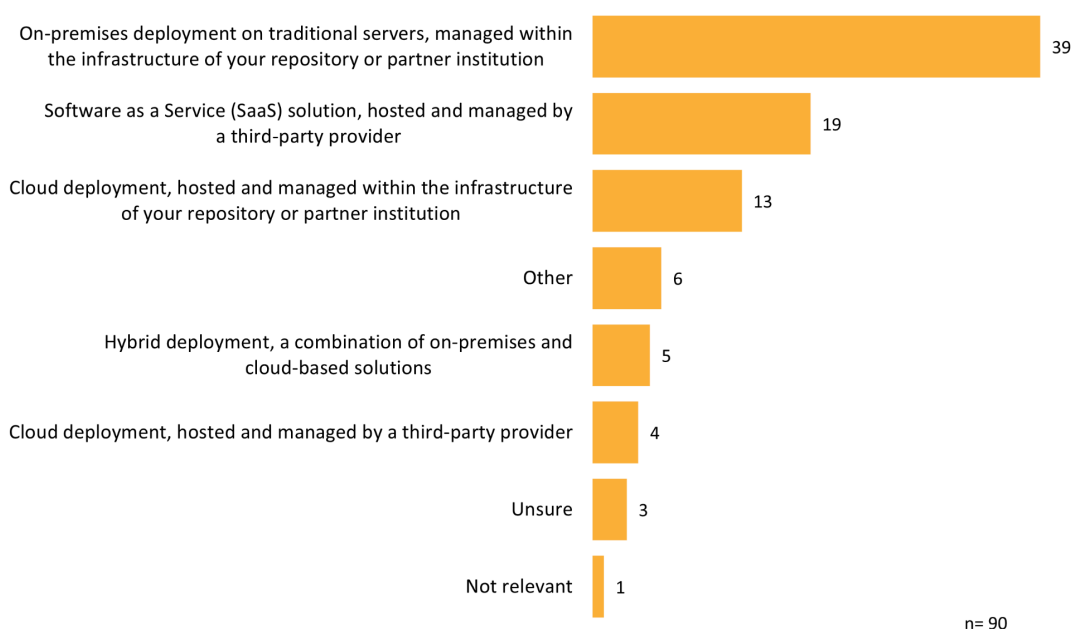


Figure 13. Technologies and services used to deploy repository software or platforms.

²⁵ <https://azure.microsoft.com/>

²⁶ <https://nextcloud.com/>

²⁷ <https://www.ebi.ac.uk/>

²⁸ <https://ontoportal.org/>

²⁹ “Repository Service of the GBV (Gemeinsamer Bibliotheksverbund) based on MyCoRe”

³⁰ <https://samvera.org/>

³¹ <https://icatproject.org/>

³² <https://geonetwork-opensource.org/>

³³ <https://figshare.com/>

Beyond the observation that “traditional” deployment (i.e., as on-premise installation on local servers) is the most favoured option, it is also possible to divide the numbers between the respondents who declare solutions self-hosted or hosted by a partner institution (52), and the ones mentioning solutions managed by a third-party provider, commercial or otherwise (28 if we include the hybrid deployments). When respondents checked “Other”, they reported slight variations on the main choices.

- Cloud deployment, all systems managed within the infrastructure apart from Colectica.
- two wanted to check both “host institution” and “3rd party” for their cloud;
- “Hybrid deployment hosted by a partner institution and managed by the same partner institution and a third-party provider”
- “cloud deployment, all systems managed within the infrastructure apart from us”;
- “on-premises deployment on virtual machines managed within the infrastructure of host institution”
- work in progress: migrating from On-premises deployment to institutional partner’s cloud deployment.

Persistent Identifiers (PIDs)

The respondents were asked whether persistent identifiers (PIDs) were applied to digital objects in their repositories, and if so, to tick in a controlled list³⁴ every type of PIDs which applied. A note was added to avoid confusion with the previous question asking for “persistent identifiers [their] repositories can be identified by”. No distinction was made in the question between digital objects actually held in the repository, and metadata elements pointing to other resources (persons or institutions registries, publication repositories, journals, etc.).

Most repositories (84 out of 91) apply PIDs, 3 are implementing them. Only one respondent checked “No”.

Regarding the persistent identifiers that are used, DOIs (provided through DataCite or other agencies) or, more broadly, Handles, make up the largest part. RORs and ORCIDs are probably used for metadata alignment. The additional identifiers not present in the list but mentioned by respondents include IdHAL; MorpheusModelID, ProteomeXchange, w3id, Wikidata ID, and finally, internal identifiers. Figures 14 and 15 show the implementation status of PIDs stated by the respondents as well as the 15 most commonly mentioned PID systems.

³⁴ The proposed list was rather extensive, with some entries (URLs, URIs, notably) not actually endorsed as PIDs by most recommendations or policies, including EOSC PID policy.

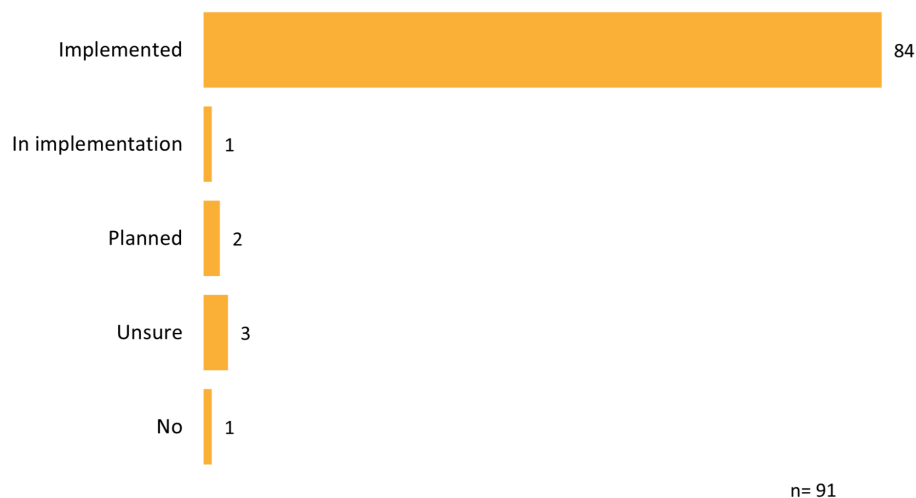


Figure 14. Use of persistent identifiers in repositories for digital objects and metadata elements.

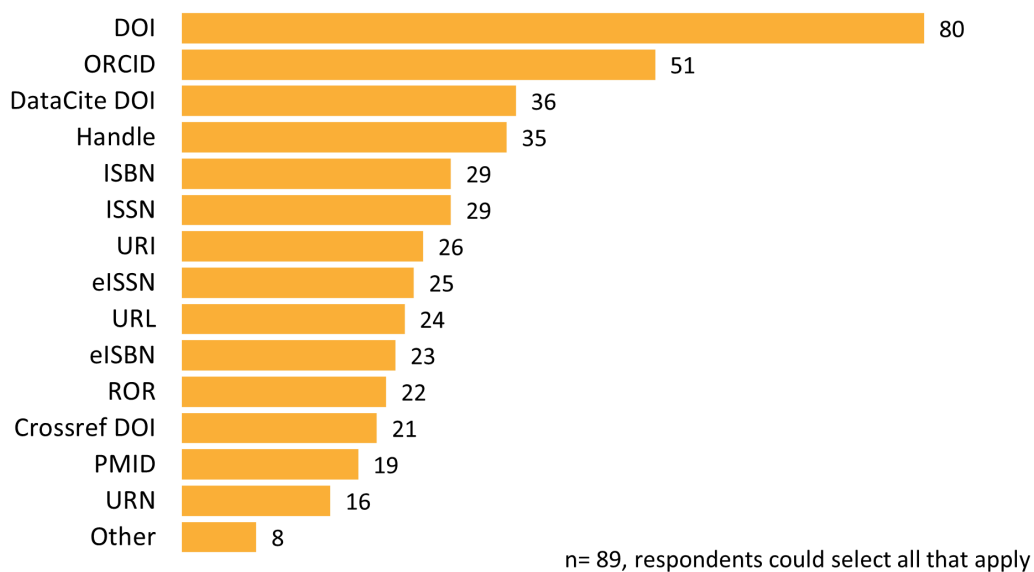


Figure 15. The 15 most commonly used or currently being implemented persistent identifiers.

Metadata schemas

The respondents were asked to cite up to five main metadata schemas used for digital object management, and specify the purpose for each of them in a free text matrix.

Not surprisingly, there is a large variety of schemas mentioned by respondents (56 in total) with a handful of well-known ones that dominate this section.

The most common schemas that are mentioned are ubiquitous. Dublin Core (54) is commonly used as a minimal format for metadata exchange and the DataCite metadata schema (29) is automatically used for datasets registered at that agency. One could infer that these schemas are so commonly used that they are implicitly assumed to be present, and that many respondents did not think of mentioning these, even if they are applied by the repository.

The large number of references to DSpace internal metadata (see Figure 17) comes from a series of DSpace instances used in the same community. As in other questions, it is advised to not overstate these numbers. Putting this aside, next in the ranking come DDI (24) and schema.org (12). In turn, only a small number of responses mentioned DCAT (5).

Some respondents cited tools, protocols (SKOS), formats (such as JSON), thesauri, etc., which are not metadata schemas strictly speaking (19 occurrences). While some of these artefacts may embed metadata, we left them aside³⁵.

Among purposes cited, the identifier registration (DataCite) is also found in the questions about catalogues, registries and services.

As for other free-text matrices, both metadata schemas and purposes have been coded from the answers.

The question received 86 answers, with the distribution as presented in Figure 16:

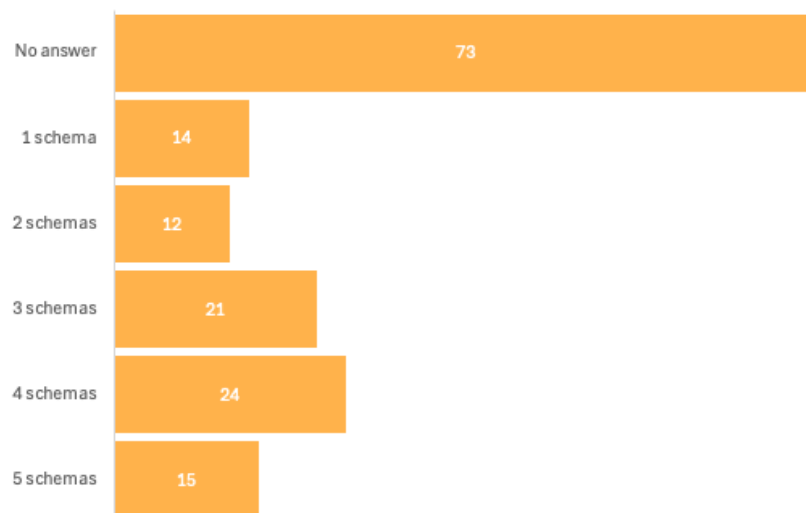


Figure 16. Repartition of respondents by number of metadata schemas cited (n=159)

³⁵ The entries not taken into consideration are as follows: CF conventions, CIF (Crystallographic Information File), Computational Workflow, DIF (Directory Interchange Format), ELSST (European Language Social Science Thesaurus), GDAL/OGR, GeoJSON, JSON, MPEG-21 schema file, OWL, RDF, RDFS, README, RO-Crate, SKOS.

The short list of the most cited schemas shown in Figure 17 does not bear any surprise. The 13 mentions of DSpace internal metadata are due to the presence of the repositories from Serbia.

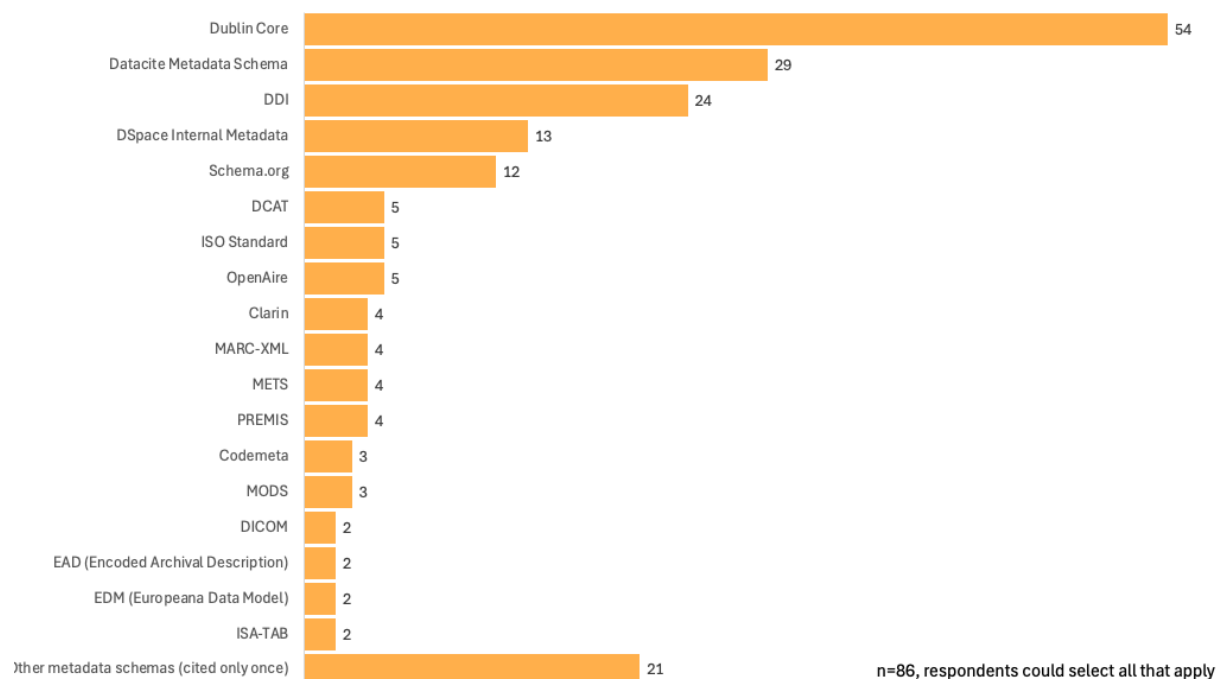


Figure 17. Ranking of metadata schemas cited.

If we break each one of these schemas into the purposes cited, we get a list which can be found in Appendix B.

Figure 18 shows the ranking of most commonly cited purposes. Table 10 links these purposes to the metadata schemas mentioned by the respondents for the given purpose.

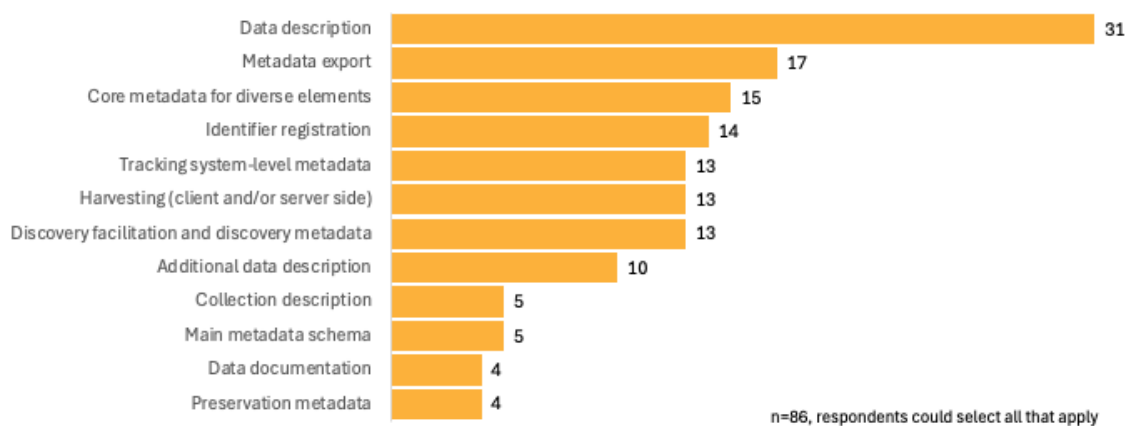


Figure 18. Most commonly cited purposes for metadata schemas.

Table 10 matches the identified purposes with the schemas cited by the respondents. To ease the reading, it can be found in Appendix B.

Respondents were then asked to specify how the metadata schemas were actually implemented. Figure 19 shows the answers.

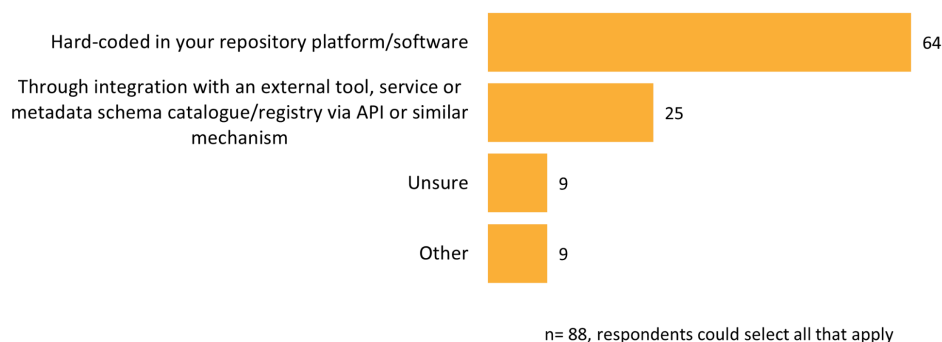


Figure 19. Implementation of the metadata schemas in the repository.

The nine respondents who checked “Other” expressed the following ideas:

- A user interface is provided, in the repository software or elsewhere (e.g., AIMS) (3)
- Configuration at the repository level (Dataverse) (2)
- Depends on the metadata schema (PREMIS/DNX hard-encoded; DC is configurable; mappings on-demand...) (1)
- Interfaced to external thesauri and metadata schema (ELSST) (1)

Semantic artefacts

Respondents were asked to specify if their repository uses semantic artefacts (e.g., controlled vocabularies, ontologies, taxonomies) to annotate, describe or index digital objects, and to indicate the implementation status of the artefacts (see Figure 20).

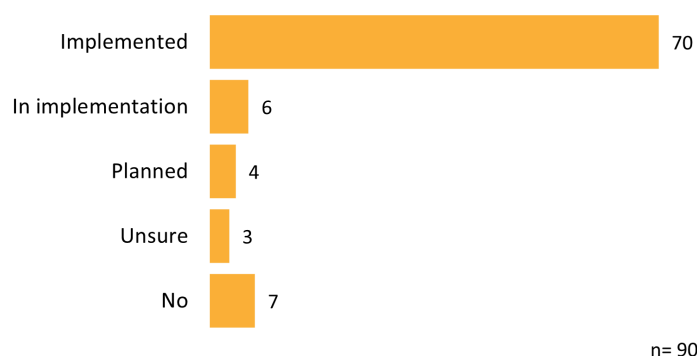


Figure 20. Use of semantic artefacts in repositories.

These responses show the importance of relying on semantic artefacts within data repositories is generally well perceived, 70 respondents over 90 having implemented some form of semantic artefact integration within their data repositories. Seven respondents, however, have not yet taken any action, nor made plans, regarding semantic artefacts, suggesting an area for improvement in the FIDELIS initiative.

The responses regarding implementation status of the semantic artefacts highlight that half of these implementations are done in a way that can possibly be improved to ease their use and maintenance (see Figure 21). FIDELIS in the future could promote the use of external semantic artefact catalogues, e.g. through the use of a standard API such as the MOD-API developed in FAIR-IMPACT³⁶, rather than an embedding or a hard-coding.

Thirty respondents mentioned relying on an external tool, service or semantic artefact catalogue. These responses might need a further inquiry to qualify the different practices (services involved, technical specificities) more precisely.

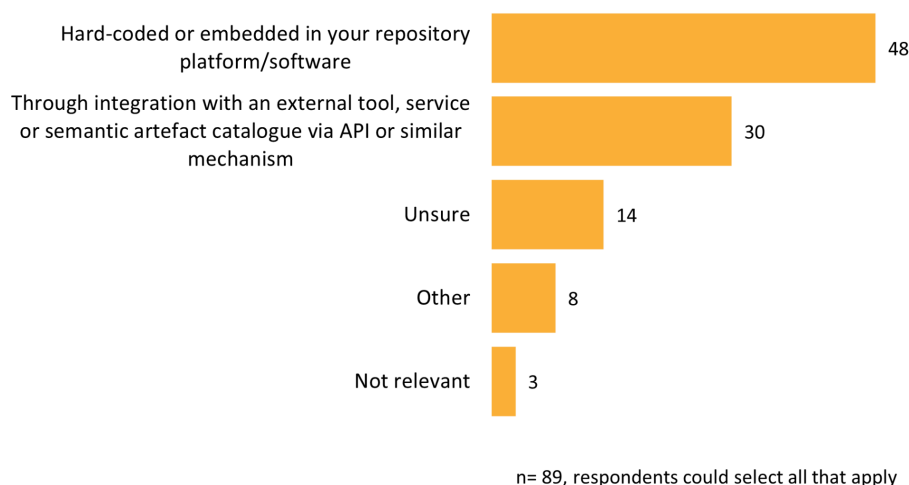


Figure 21. Implementation of the semantic artefacts in the repository.

³⁶ A key takeaway of FAIR-IMPACT was the recommendation to use external vocabulary services or semantic artefact catalogues to manage vocabularies. The reasons are multiple. First, versioning and maintenance: external services specialize in maintaining vocabularies over time, including updates, corrections, and deprecated terms, which ensures long-term semantic consistency and guarantee to have the latest version of the vocabulary. Second, a separation of concerns : decoupling vocabulary management from the repository architecture makes the system more modular, maintainable, and adaptable to change. Third, an access to APIs and tools, since external services often provide API (possibly standard ones), and tooling that enables them to go beyond the use of controlled terms e.g., semantic annotation, vocabulary recommendation, term proposal, etc. (See : Wilson, Antony, et Clement Jonquet. «D4.3 - Specification of Shared Metadata Description of Semantic Artefacts and Their Catalogues Including Common Reference API» 2024. <https://doi.org/10.5281/ZENODO.12579779>)

From the responses of participants who ticked "Other", the following details can be noted:

- Manual entry (2)
- Manual entry in a dedicated Dataverse metadata field which is mapped to DDI (1)
- “Can be used and searched for via the metadata profile generator, more terminologies need to be added (coscine)” (1)
- “Controlled and linked” (1)
- “Harvesting external services into local data storage for optimised search and discovery” (1)
- “Our own software and own vocabulary server that is fed from external authoritative servers when we are not the authoritative source for the vocabulary” (1)
- “We have captured project metadata” (1).

Shared services

Respondents were asked about the shared catalogues, registries, resources, services, and federated systems they used. For ease of reading, the term "service" is used generically to refer to the five object categories addressed in the questions.

The respondents were asked to cite up to five main catalogues or registries used for digital object management, and specify the purpose for each of them in a free text matrix. The same was asked regarding resources, services or federated systems.

The answers to these two questions reveal a certain confusion regarding their respective focus, which some respondents explicitly expressed (“Maybe I don't understand this question, what do you mean by catalogue or registry?”; “I don't quite understand the question above”, etc.). In fact, some catalogues or services are mentioned under both questions, while many entries in response to the second question actually refer to vocabularies or registries (e.g., MeSH, CESSDA vocabulary service, registries of institutions or people, etc.). The mention of DataCite for minting DOIs (coded below as “Identifier registration”) is ubiquitous across the different answers.

This confusion may not be particularly problematic, in the sense that the services listed in both matrices commonly act as third parties supporting various core or additional repository functionalities. In this regard, the top ten list and the indicated purposes are meaningful.

For this reason, we present the data for these two questions together. To do so, we combined the answers to both questions in a single list. The free text answers were coded, both for the items listed and their associated purposes. We checked that there were no identical answers from the same repository. A large number of services were mentioned in the answers; many of which had a limited scope (e.g. discipline-specific or limited to a particular institution or country).

Together, the two questions received responses from 49 repositories. One service (NomadLite, a local service from the University of Belgrade) was mentioned twice in 12 answers to cover two different purposes. To correct this, we reduced these occurrences by half in the following graph.

The number of different services mentioned was quite large. In Figure 22 below, we present the distribution of mentions only for the services mentioned more than once³⁷.

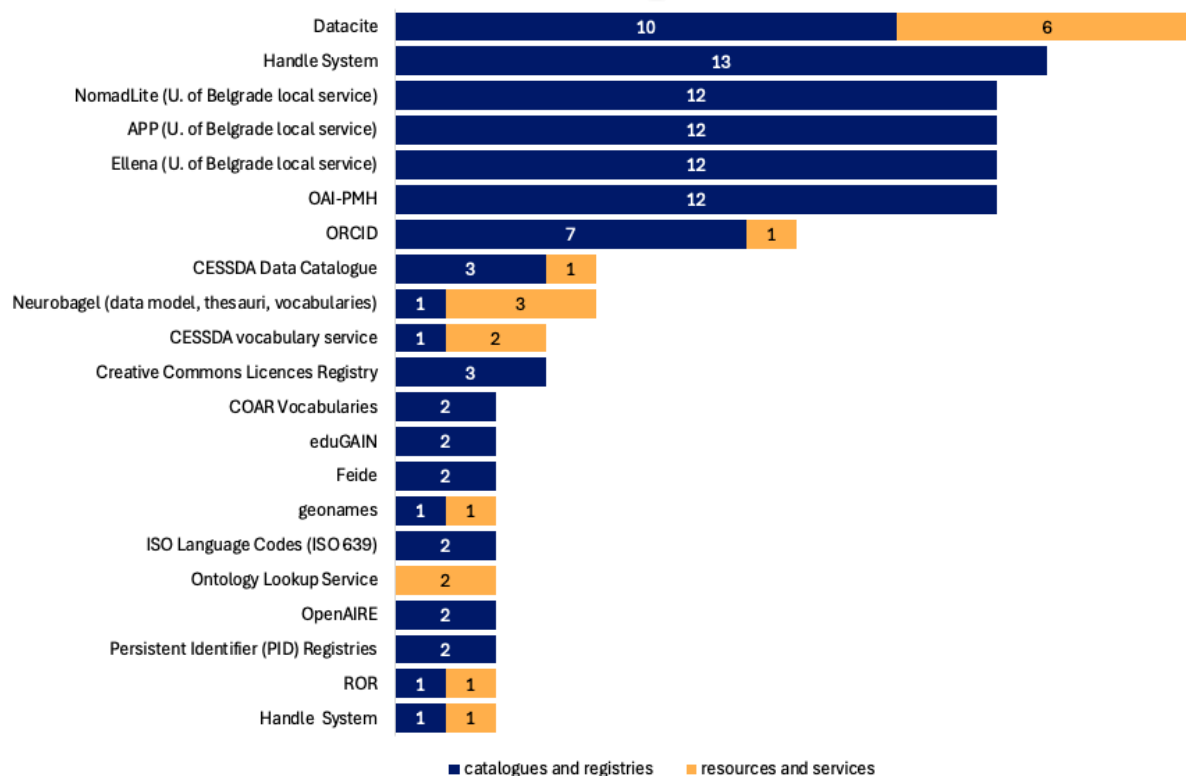


Figure 22. Ranking of catalogues, registries, resources and services most cited (n>1).

A breakdown of these responses by purposes mentioned results in a list which can be found in Appendix B.

Figure 23 shows the ranking of most commonly indicated purposes for the services mentioned.

³⁷ The resources and purpose cited

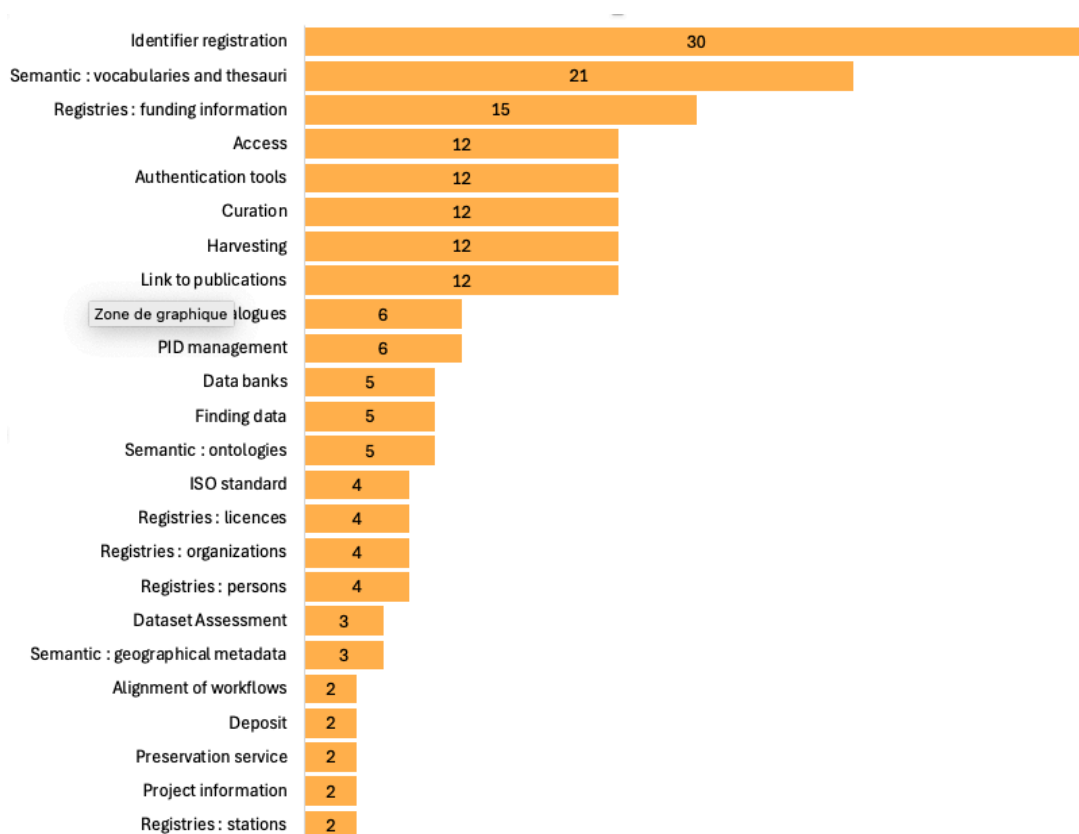


Figure 23. Purposes most cited (n>1) for catalogues, registries, resources, services, or federated systems.

This list of purposes largely summarises some of the functionalities of repositories that are not provided by the repository software alone, but rely on third parties to operate, for example identifier registration or everything that pertains to semantic interoperability (such as ontologies and vocabularies).

Table 11 matches the identified purposes with the services cited by the respondents. To ease the reading, it can be found in Appendix B.

Organisational Infrastructure

Like for Digital Object management, respondents were first asked to indicate the difficulties and needs they encounter regarding organisational infrastructure in a controlled matrix: on the y-axis, a list of sixteen activities and functions derived from the TTRAM; on the x-axis, a 5-point scale ranging from “no challenges” to “critical challenges”.

Then, a free text question asked the respondents to list the major challenges they face and add any clarifications they wished.

For each of the eight TTRAM A|F Figure 24 shows the distribution of responses on a five-point-scale, illustrating the respondents’ assessment of the severity that the particular A|F poses in relation to organisational infrastructure.

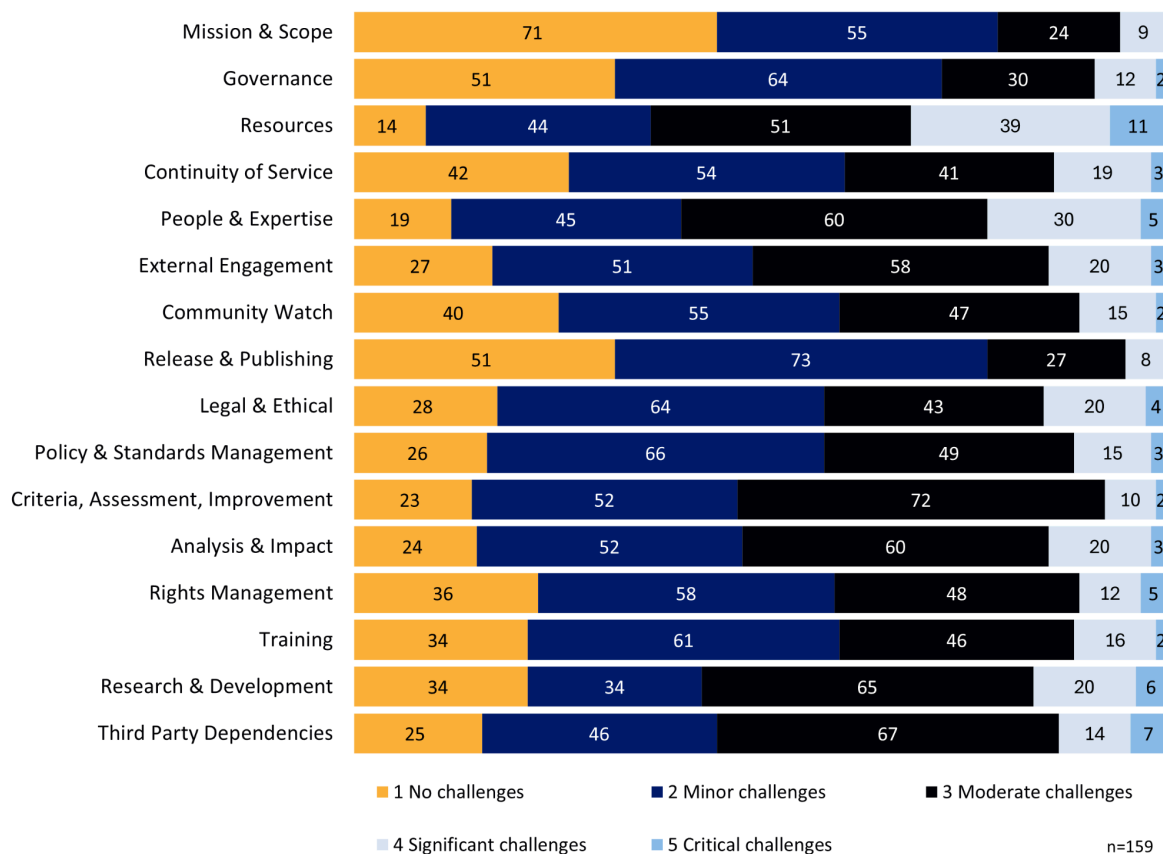


Figure 24. Challenges and needs that repositories face in terms of organisational infrastructure.

Here, we are facing both a clear and contrasting picture. “Mission and scope”, “Governance”, and “Release and publishing” are relatively unproblematic areas. On the other hand, issues related to resources, such as funding, personnel, and expertise, stand out as primary concerns. Finally, less vital activities and functions such as “R&D” or “Third party dependencies” also come clearly to the front.

As for Digital Object Management, a simple cross-tabulation was also conducted, focusing on institutional and national repositories. Figure 25 highlights the focus on institutional repositories, while Figure 26 highlights the focus on national repositories.

For institutional repositories, aspects related to repository administration and political positioning (e.g., “Mission & Scope,” “Governance”) appear to be less problematic compared to the global set of respondents. Similarly, “Resources” and “People & Expertise” seem somewhat less problematic,

though they are far from being non-issues. National repositories, on the other hand, show no such marked differences between repositories functions and activities.

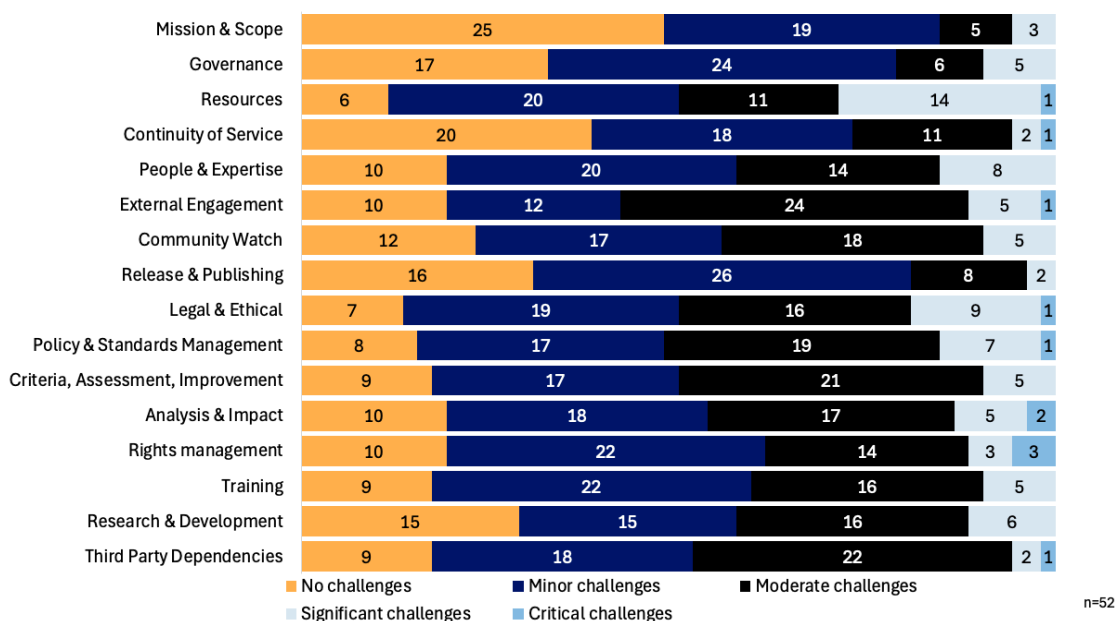


Figure 25. Challenges and needs that institutional repositories face in terms of organisational infrastructure.

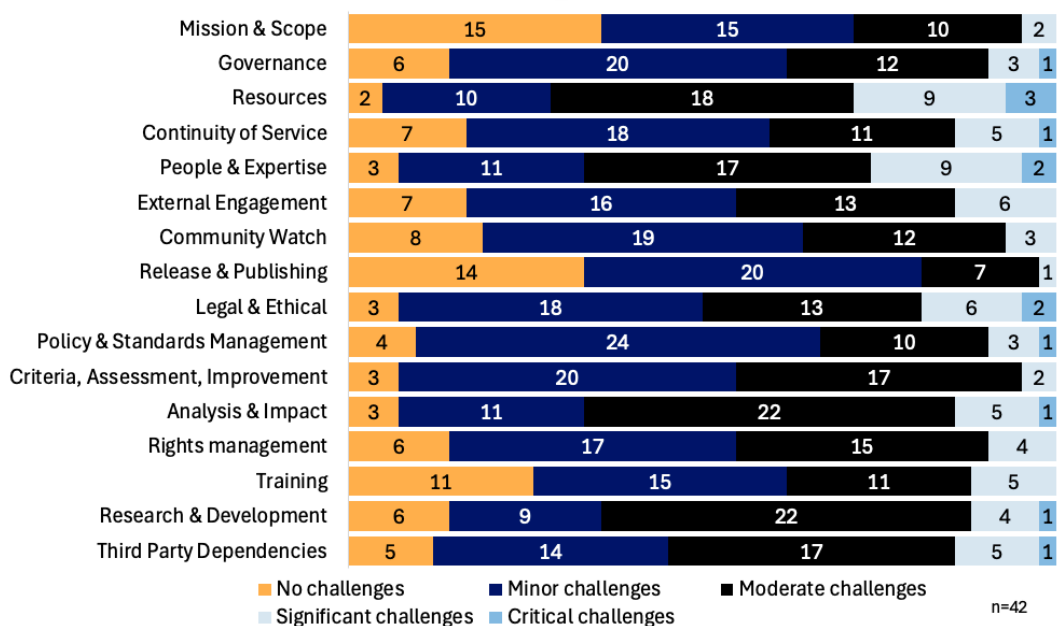


Figure 26. Challenges and needs that national repositories face in terms of organisational infrastructure.

The free text answers confirm this first impression. Concerns about funding and resources are the most pressing issues, with concerns about long-term sustainability, human resources, and infrastructure. Gaps in People and expertise are mentioned, especially in technical and preservation roles, along with a need for training and skill development. Governance is complicated by diverse stakeholders and fragmented ownership, while mission and scope concerns include internal competition and sometimes difficulty for the repository to position itself in a changing environment. Legal and ethical issues—especially around GDPR, licensing, and sensitive data—are frequent. Challenges in external engagement, third-party dependencies, and aligning with user communities also underscore the need for coordination and strategic clarity.

The methodology applied here is the same as for the “Digital object management challenges” question. That is, for each of the TTRAM A|F considered here Tables 12-18 show the number of mentions of different categories of ideas and challenges induced from the responses. In addition, other TTRAM A|F are cross-referenced if it is relevant to the expressed ideas and challenges.

A quick remark about the human resources issues: whenever a speciality (curators, data stewards, etc.) was specified, we considered that it fell under the “People & Expertise” category; otherwise it falls under “Resources”.

Table 12. TTRAM Activities and Functions (A|F) : Mission & Scope

Idea expressed	Mentions	Other TTRAM A F
Assessing the scope / target audience	1	
Increase visibility	1	
Internal competition with another repository, internal dependencies	2	Governance
Multiple other solutions	1	
Potential loss of accuracy inside a larger repository/network	1	Governance

Key points:

In the multi-choice question, it appears that the “mission and scope” activity is not challenging, hence the few occurrences of this activity in this open-ended part of the question. However, for the remaining activities and functions, it is possible to read a struggle for visibility, undesired competition in a field or within the same ecosystem, and perhaps a need to clarify both mission and scope to ensure continuity.

Table 13. TTRAM Activities and Functions (A|F) : Governance

Idea expressed	Mentions	Other TTRAM A F
Diversity of stakeholders, fragmentation of ownership	4	Mission & Scope
Governance	3	
Support from parent / host institution	3	Mission & Scope
Stakeholder management	2	
Change of governance in the near future	1	

Key points:

A clear and unified governance is key to ensure continuity. However, the challenges raised here point in the opposite direction: multiplicity of parent organizations or stakeholders (4), difficulty in managing diverse stakeholders (2); lack of support from parent or host organizations, among others.

Table 14. TTRAM Activities and Functions (A|F) : Resources

Idea expressed	Mentions	Other TTRAM A F
Funding	13	
Human resources	8	
Competition for funding / Project-based funding	3	
Funding: Advocacy support (help to discuss with the funders, recognized arguments)	2	
Funding: ensuring the long-term evolutions	2	
Funding: Recognized models and guidelines for resource management in research repositories	2	
“Increased resource allocations to allow for larger enhancements and long-term planning of repository development, e.g., through increased local, national, European and global investment in and prioritization of open, community-driven research infrastructure, including human capacity”	2	
Resources	1	

Key points:

Not surprisingly, funding is the most commonly reported challenge . While two respondents outline that recognized, established argumentations could help advocate for the continuity of the service, many others report major challenges related to both funding and human resources. Interestingly, several highlight that project-based funding creates instability, not desirable for services supposed to offer stability and longevity.

The need for human resources, when the respondent does not precise the speciality wished, has been listed here as a general “resource”; added to the entries in “Personnel & Expertise” it would attain 27 occurrences.

Table 15. TTRAM Activities and Functions (A|F) : Continuity of service

Idea expressed	Mentions	Other TTRAM A F
----------------	----------	-----------------

Continuity of service: Recognized models and guidelines for succession planning; available frameworks and mechanisms for succession planning	2	People & Expertise
Manage to cover the assigned perimeter (institutional) while keeping the pace technically	1	

Key points:

Regarding “Continuity of service”, both ideas raised are highly valuable. First, the need or at least high value of succession planning. Repository management requires training and skills, which can hardly be replaced without a good environment in place to preserve institutional knowledge and maintain workflows. Second, the tension sometimes emerging between the missions of the repository and the technical evolutions and watch, also mentioned in the “Technology and Security” subsection.

Table 16. TTRAM Activities and Functions (A|F): People & expertise

Idea expressed	Mentions	Other TTRAM A F
Human resources: with expertise (no specification)	7	
Human resources: development	3	
Human resources: data stewards	2	
Human resources: professional and initial training programs	2	
Human resources: skills framework	2	
Human resources skill set (participating institutions)	2	
Improve software	2	
Human resources: preservation	1	

Key points:

Many of the different roles and functions necessary to repositories are represented here, with the exception of curators, the lack of which is mentioned in the Digital object management’s challenge question above. Software improvement is included in this context due to the human resources it entails, although many other activities and functions would have been relevant.

Table 17. TTRAM Activities and Functions (A|F): External engagement

Idea expressed	Mentions	Other TTRAM A F
Closer engagement with the disciplinary community (Clarín)	2	
Closer engagement with the generalist community (RDA, etc.)	2	
Community connection and outreach	3	
Fragmentation of the community	1	
Funding the community outreach	1	

Idea expressed	Mentions	Other TTRAM A F
Keeping up with the user base	1	
Link to the user base	2	
Managing research organizations responsibilities (having them get a model for data governance)	1	Training
Very diverse user base	1	

Key points:

External engagement as defined in the TTRAM is multi faceted, englobing interactions with external partners as well as user base management. The heterogeneity of the user base is a recurring topic. It can be linked here to a need for training or community building and the perceived difficulty to link to the target communities (disciplinary or not). It is possible that limited funding and unclear institutional responsibilities further hinder sustained interaction. The need for clearer models of data governance to support research organizations must be noted.

Table 17. TTRAM Activities and Functions (A|F): Legal & Ethical

Idea expressed	Mentions	Other TTRAM A F
Recognized guidelines, frameworks, templates... for TDR policy & standards management	1	
Implementation of TK (Traditional Knowledge) labels	1	
Support (guidelines...) for GDPR or ethical requirements, including technical implementations, event at the file or element level	1	
Legal or ethical issues about data	1	
Legal: consortium management	1	
Legal: coordinating different rights management / licensing schemes inside the repository	1	
Legal: GDPR / anonymity	1	
Legal: Intellectual property rights	1	
Copyright, Intellectual property, user consent	1	
Regulatory compliance	1	
Support for sensitive data	1	

Key points:

Repositories operate within a complex legal and ethical landscape. They face significant challenges relating to GDPR compliance, intellectual property rights, user consent and regulatory obligations. The responses emphasise the urgent need for clearer, more consistent guidelines, frameworks and technical tools to help repositories navigate these issues.

Critical areas include managing sensitive data, applying appropriate rights and licensing schemes, and ensuring regulatory compliance at organisational and file levels. Repositories also emphasised the importance of bespoke legal and technical support for implementing Traditional Knowledge (TK) labels and managing data privacy. Additional concerns raised included coordinating legal responsibilities within consortia and aligning with national and international standards.

Overall, there is a clear demand for actionable guidance that integrates human legal expertise with technical implementation in order to address the diverse legal and ethical scenarios encountered in digital repository management.

For the TTRAM A|F Analysis & Impact, no table is necessary. Six respondents mentioned the need for impact measurement or meaningful indicators. A respondent wishes there would be a “framework including meaningful indicators of mission achievement for TDRs”, while another one deplors the lack of clear metrics or indicators measuring the reuse of datasets.

Challenges related to training were mentioned twice in this survey section, but training challenges are also apparent in other sections of the survey.

Table 18. TTRAM Activities and Functions (A|F) : Third-party dependencies

Idea expressed	Mentions	Other TTRAM A F
Coordination between a partner institution and a third party provider	1	
Development of and cooperation with the university's CRIS system, which is currently under construction.	1	
Third party dependencies	3	
Third party dependencies (Duraspace)	2	
Third party dependencies (Ex Libris Rosetta)	2	
Third party dependencies [Figshare]	1	

Key points:

Third-party dependencies for a repository service can be a liability, if the agreement with a provider covers core or vital features of the repository service. This advocates in favour of a greater autonomy in managing critical infrastructure components. Some cases outlined here involve vendors of software-as-a-service solutions for repositories by their names. Of the three mentions of third parties dependencies, at least one is also in this kind of situation (“We are relying on 3rd party software, so there would be a challenge if the University stopped funding the subscription, or if the company went out of business”). This reliance may also imply limitations in flexibility,

responsiveness, long-term planning, or access to core or basic services (“we struggle to get any meaningful usage statistics about the data held on our repository”)³⁸.

Some other ideas that do not necessarily pertain to this question are listed in Table 19 below.

Table 19. Additional ideas expressed by respondents linked to TTRAM A|F

Idea expressed	Mentions	Other TTRAM A F
Assessing data quality across new formats and research methods	1	Deposit & Appraisal
Data integrity	1	Storage & Integrity
Framework for depositing data linked to publications	1	Discovery & Identification
Having research data appraisal guidelines for research organisations	1	Deposit & Appraisal
“I cannot legally comment on them.”	1	
No challenges	1	
Preservation	1	Preservation
Semantic interoperability	1	Interoperability
Server maintenance (power outage)	1	Technical infrastructure
Support for reproducibility of data and software, also in relation to Machine Learning (ML) and Artificial Intelligence (AI).	1	Interoperability
Tool integration with the repository for data and software	1	Interoperability

Additional questions regarding organisational infrastructure

After the two questions regarding challenges and needs, respondents were asked: “*These were the core questions on organisational infrastructure. Would you like to answer some additional questions or continue to the next part of the survey?*”. 78 respondents checked “Yes”, triggering the following questions.

Licensing

The respondents were asked to specify which licence(s) they used for four categories of digital objects: metadata, data, code/software, and other digital objects with a request to precise which digital object the licence applies to.

More specifically, the questions were as follows:

³⁸ Another respondent, however, mentions that [they] “do not have any issues with the challenges listed above as we use Figshare’s infrastructure, which is already highly developed and geared towards a generalist repository paradigm.”. As no challenge is listed, it does not figure in the numbers here.

- For the reuse of metadata, which licence(s) does your repository apply?
- For the reuse of data, which licence(s) does your repository apply?
- For the reuse of code/software, which licence(s) does your repository apply?
- For the reuse of digital objects other than metadata, data or code/software, which licence(s) does your repository apply? Please specify the type of digital object and the licence(s).

Respondents were allowed to provide any number of licences for any type of content. Although the answers were in free text, they were relatively easy to aggregate, as most of the mentioned licences are widely recognized and commonly used. Figure 27 shows the distribution of different licences used for different types of digital objects, e.g. metadata, data, code/software and other.

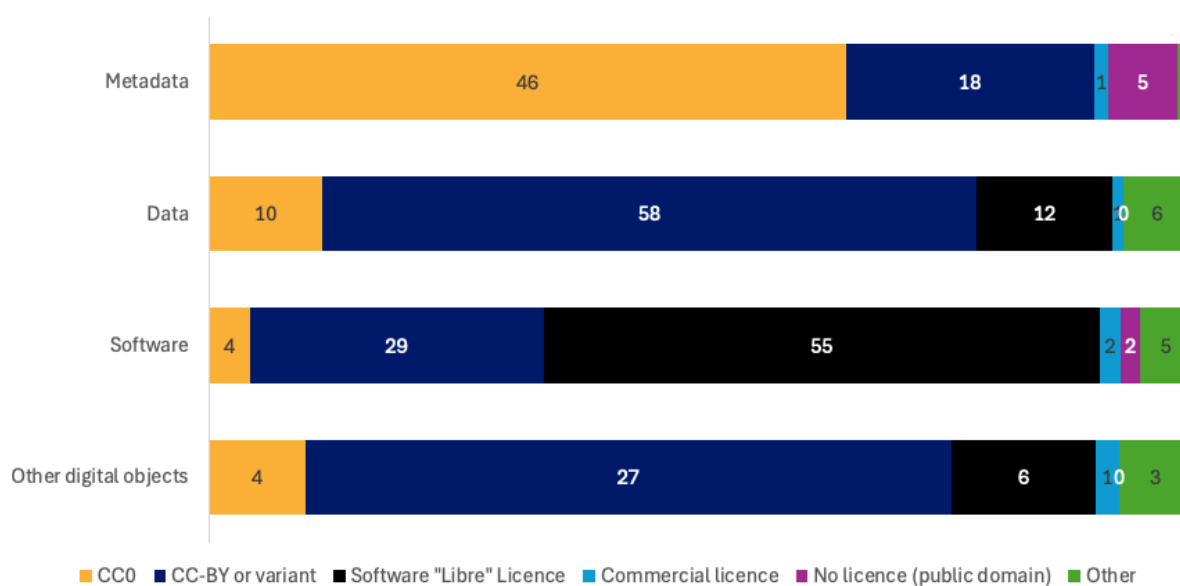


Figure 27. Repartition of types of licences by category of digital object.

These numbers should be interpreted with caution. Some respondents simply copy-pasted a long list of licences in response to each question without refining their own specificity. This may explain the apparent application of CC licences to software, which could be however explained otherwise by the embedding of code in the dataset.

The abbreviations we used for licences are derived from the Linux Foundation Project webpage³⁹. We did not keep the version numbers, as respondents most often did not specify them.

Licences for Metadata

For metadata, i.e. contents usually freely displayed and disseminated, CC0 dominates the answers (see Table 20).

³⁹ <https://spdx.org/licenses/>

Table 20. Licences for metadata

(n=66)

	Licence	Metadata	Total
CC0	CC0	46	46
Creative Commons licence requiring attribution	CC-BY or compliant	13	18
	CC-BY-SA	3	
	CC-BY-NC	2	
	CC-BY-NC-SA	0	

Licences for Data

As shown in Table 21, Creative Commons licences, (licenses demanding attribution), dominate the picture. The two responses containing six software licences each seem to contain copy-pasted identical lists of licenses across all types of items. It is therefore uncertain whether these entries correspond to an actual application.

Table 21. Licences for data

(n=59)

	Licence	Data	Regrouped
CC0	CC0	10	10
Creative Commons licences requiring attribution	CC-BY or compliant	28	58
	CC-BY-SA	14	
	CC-BY-NC	14	
	CC-BY-NC-SA	2	
Software "libre" licence	BSD	2	12
	Apache	2	
	LGPL	2	
	AGPL	2	
	ODC	2	
	MIT	2	
Commercial licence		1	1

Licences for Software

As expected, software-dedicated licences predominate in this section (see Table 22).

Table 22. Licences for software

(n=49)

	Licence	Software	Total
CC0	CC0	4	4
Creative Commons licences requiring attribution	CC-BY or compliant	9	29
	CC-BY-SA	8	

	Licence	Software	Total
	CC-BY-NC	6	
	CC-BY-NC-SA	6	
Software "libre" licence	BSD	8	55
	Apache	9	
	LGPL	12	
	AGPL	4	
	ODC	1	
	MIT	13	
	EUPL	5	
	MPL	2	
	Eclipse Public License	1	
Commercial licence		2	
Other	"Public domain" or "free use" without a (known) licence	2	5
	All rights reserved	0	
	Scientific use file Licence	0	
	Other: ToS, Internal licences...	2	
	Custom license, chosen by user	2	
	ODbL	1	
	CERN Hardware licences	0	
	Open Government License	0	

Licences for Other Digital Objects

There were few respondents for this question, many of whom stated that they do not possess digital objects other than those already discussed, or simply just provided a list of licences, which we summarise below (Table 23).

Table 23. Licences for other digital objects

(n=20)

	Licence	Other digital objects	Total
CC0	CC0	4	

	Licence	Other digital objects	Total
Creative Commons licences requiring attribution	CC-BY or compliant	12	27
	CC-BY-SA	6	
	CC-BY-NC	5	
	CC-BY-NC-SA	4	
Software "libre" licence	BSD	0	6
	Apache	2	
	LGPL	1	
	AGPL	1	
	ODC	1	
	MIT	1	
Commercial license		1	1
Other	All rights reserved	0	3
	Scientific use file Licence	0	
	Other: ToS, Internal licenses...	0	
	Custom license, chosen by user	1	
	ODbL	1	
	CERN Hardware licences	1	
	Open Government License	0	

While the question prompted respondents to specify the type of digital objects to which a given licence is applied, few respondents gave information. Table 24 contains a summary of these responses.

Table 24. Licences for named "other" digital objects.

Object	Licence(s) applied
MorpheusML model descriptions	CC-BY 4.0
Open Hardware	CERN Open Hardware licences: CERN-OHL-P, CERN-OHL-W, CERN-OHL-S
Grey literature	Sometimes no licence, but everything we produce is CC-BY 4.0

Continuity

The respondents were asked whether their repository has a continuity agreement with another organisation to ensure ongoing access to and preservation of their digital objects and metadata, and to indicate the implementation status of the agreement as shown in Figure 27.

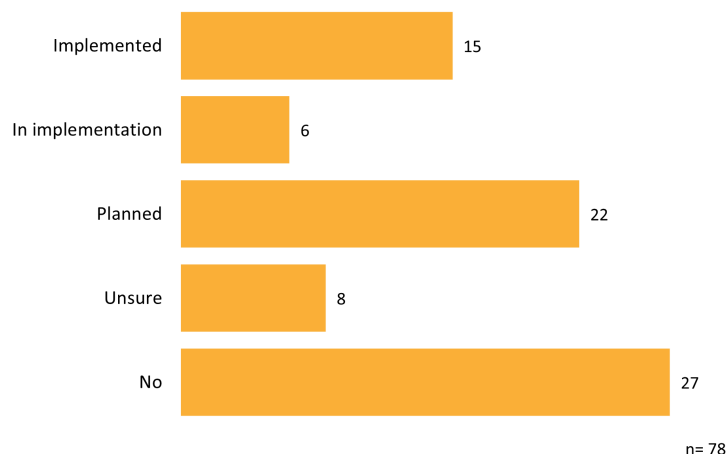


Figure 27. State of continuity agreements.

If the respondents answered that they have implemented or are implementing a continuity agreement, they were further asked to specify in free text what type of organisation the agreement is with (see Figure 28). Most respondents mentioned public entities such as national infrastructures, consortia of research performing organizations (or a combination or variation). Only one respondent mentioned contracts with commercial entities.

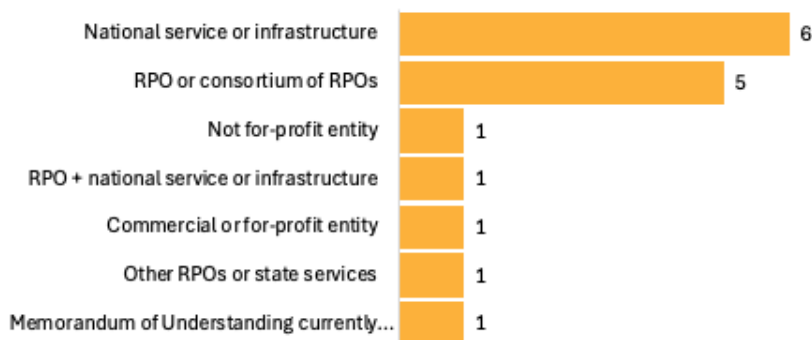


Figure 28. Stakeholders of continuity agreements.

Funding

Regarding funding, the respondents were asked: “To what extent does your repository have sustainable funding to carry out your mission?” As shown in Figure 29, the majority of respondents

(n=29) indicated that they have sustainable funding for 2-5 years. However, 11 respondents answered that they were unsure about their repository’s funding and nine respondents indicated that their repository has funding for one year at a time. So, while many repositories have a moderate level of sustainable funding, there is still uncertainty and short-term funding among a notable portion of respondents. On the other hand, 19 respondents indicated they have sustainable funding extending beyond five years (i.e., nine responses for “6-10 years” and 10 responses for “More than 10 years”).

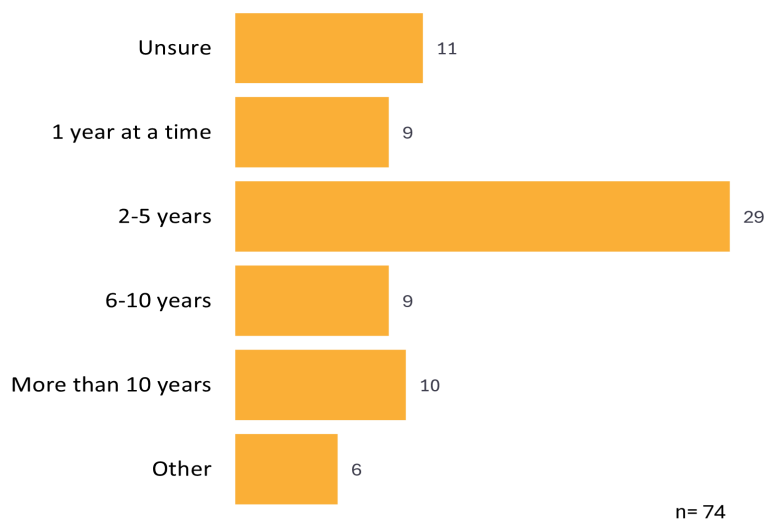


Figure 29. Duration of sustainable funding for repositories.

The following details were identified from the responses of participants who selected “Other” (see Figure 29). One respondent reported receiving joint funding from the government and the university for a period of 2-5 years, with the university committed to continuing support if government funding were reduced or discontinued. Two respondents described funding models based on annual budgeting through their host institutions, without clearly allocated long-term resources. One repository is actively working toward a more sustainable model in collaboration with other national or international partner institutions, while another operates without any funding and is maintained on a pro bono basis. One noted that the repository is currently supported by new, dedicated hardware expected to meet operational needs for the next few years. However, future upgrades and additional staff may be required if data deposits or user support demands increase. One respondent mentioned that the funding is mainly private including sales of resources.

Next, respondents were asked to specify the main funding sources of their repository, with the option to select all that applied (see Figure 30). Based on the responses of 76 participants, the majority (n=53) reported receiving funding from their host or partner institutions. The next most common sources of funding were structural funding (n=33) (i.e. central funding or contract from a research or infrastructure funder that is in the form of a longer-term, multi-year contract) and support through contract services or project funding (n=20) (i.e. charges for contract services to other parties or for research contracts), which indicates a moderate reliance on more temporary

funding. Less commonly mentioned funding sources were data deposit fees (*i.e. in the form of annual contracts with depositing institutions or per-deposit fees*) and funding from access charges (*i.e. charging for access to standard data or to value-added services and facilities*) (n=3 each), which may indicate that user-based funding models are relatively uncommon. Five respondents were unsure of the main funding sources of their repository. Respondents could also indicate other funding sources that their repository has.

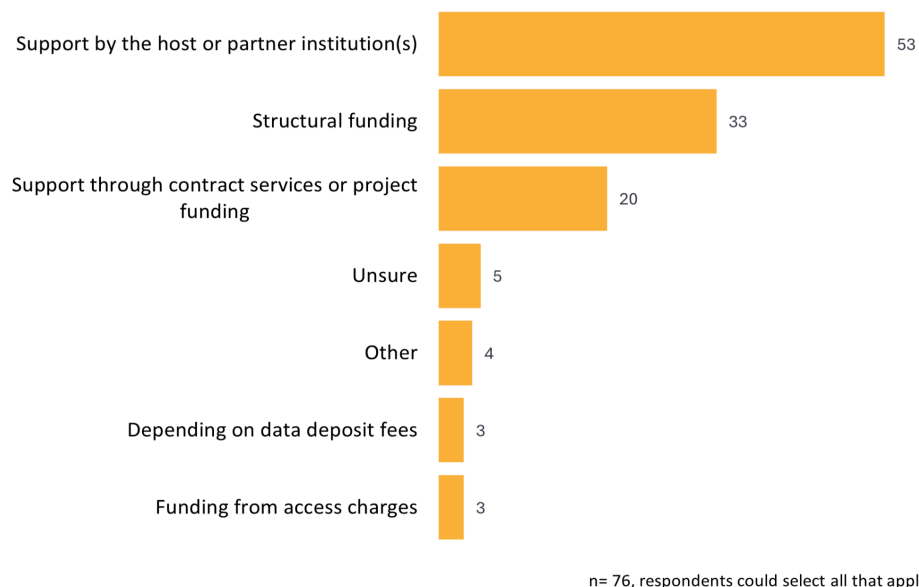


Figure 30. Main funding sources of repositories.

Four respondents indicated their repository’s funding source as “Other” (see Figure 30). Three of these respondents refer to national public funding (national research infrastructure, Ministry of Education and Culture, state budget). One respondent states that the process of receiving payments through annual contracts with the relevant depositing institutions is currently in the planning or preparation phase.

Technology and Security

Like for Digital Object Management and Organizational Infrastructure, respondents were asked to indicate the difficulties and needs they encounter regarding Technology and Security on a controlled matrix : on the y-axis, there was a list of five activities and functions derived from the TTRAM, and on the x-axis, a 5-point scale ranging from “no challenges” to “critical challenges”. Following this, respondents were again presented with an open-ended question, asking them to list the primary challenges they encounter.

Figure 31 shows for each of the five derived activities and functions⁴⁰ how responses were distributed on a five-point-scale, illustrating the respondents’ assessment of the severity that the particular A|F poses in relation to technology and security.

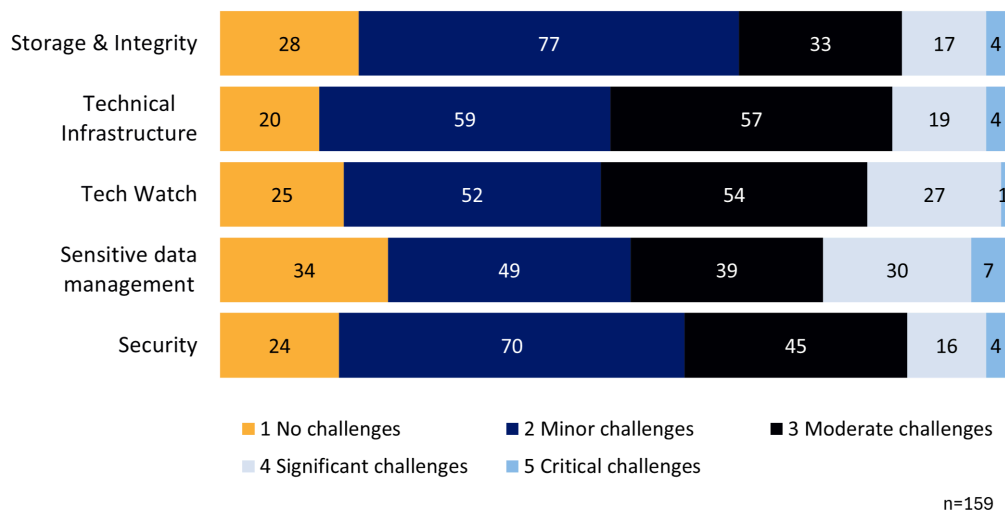


Figure 31. Challenges and needs that repositories face in terms of technology and security.

While no single subject clearly stands out among the entire group of respondents, the overall impression is that of a cautiously concerned community.

The slight dominance of the “Sensitive data management” topic is worth noting, although there is a lack of significant comments from the respondents who selected the “critical” or “significant challenges” options.

“Tech watch” is another topic which seems to raise concerns. These concerns can be associated with issues expressed in the free text comments about keeping the repositories up to date and aligned with changing technical requirements.

The number of mentions for each of the five TTRAM A|F considered here is shown in Figure 32, with “Other” receiving the most mentions.

⁴⁰ Note that “Sensitive data management” and “Tech watch” are not standalone TTRAM A|F

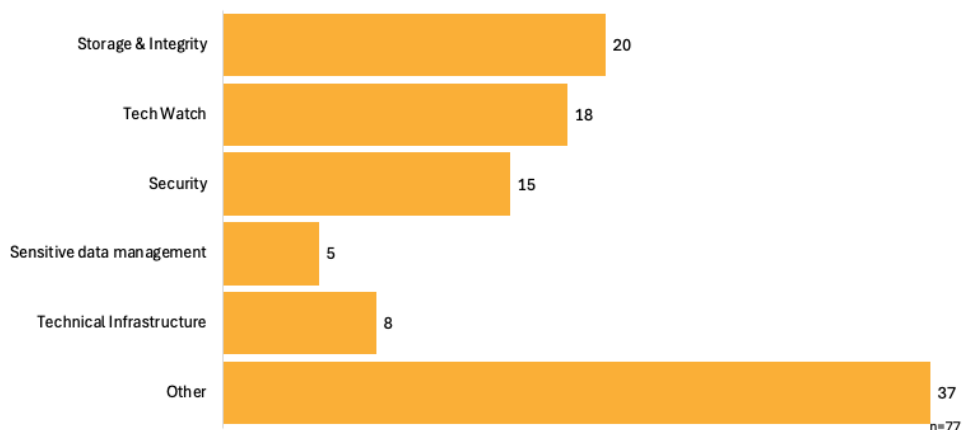


Figure 32. Number of respondents having expressed challenges in each topic.

The methodology applied here is the same as for the “Digital object management challenges” and the “Organizational infrastructure” questions. That is, for each of the TTRAM A|F considered here Tables 25-29 show the number of mentions of different categories of ideas and challenges induced from the responses. In addition, other TTRAM A|F are cross-referenced if it is relevant to the expressed ideas and challenges.

Table 25. TTRAM Activities and Functions (A|F): Security

Idea expressed	Mentions	Other TTRAM A F
Aggressive harvesting by AI companies	3	
Community-driven approaches to security management in TDRs	1	
Metadata	1	
Security	10	

Key points:

Two ideas can be highlighted:

- the concerns about “aggressive” harvesting by commercial AIs (stressing the infrastructure, beyond the challenge posed by the use of the data itself)
- “metadata obsoleted by AI”: a respondent reports the idea that “the evolution of AI means that some taken-for-granted approaches such as the need for metadata or the assignment of subject matter or keywords to existing records in institutional repositories (...) are being called into question and are a technical challenge for them”.

Table 26. Sensitive data management

Idea expressed	Mentions	Other TTRAM A F
Sensitive data: “also starting for proteomics”	1	Legal & Ethical
Sensitive data: making sure coordinates are not published	1	Legal & Ethical
Sensitive data: making sure no sensitive data is uploaded	1	Legal & Ethical
Sensitive data: repository not suitable for some types of sensitive data (personal data for ex.)	1	Legal & Ethical
Sensitive data: risk of overcaution	1	Legal & Ethical
Sensitive data: restrict access	1	Legal & Ethical
Staff training in data privacy	1	People & Expertise, Training

Key points:

As mentioned earlier, this topic is the most problematic challenge according to the results of the controlled matrix. However, the repositories which ticked this choice did not expand their input in the free text field.

Interestingly, the mentions of sensitive data in the free text field are either from repositories not concerned by such data (in which case we did not count their answer in the above table); or, as can be seen above, from repositories which receive sensitive data in submissions without being equipped to receive them (3 respondents). One respondent notes that their discipline is beginning to face the problem, another mentions access management issues. The mention of a risk of overcaution is expressed more broadly as a problem of “coordination between the different departments involved in safeguarding sensitive data [which] leads to overcaution from time to time”⁴¹.

Table 27. Tech watch

Idea expressed	Mentions	Other TTRAM A F
Collaborative tech watch	3	
Evolutions of the repository to keep it current	5	
Keeping up with new standards	7	
Lack of community standards	1	

Key points:

This was the second most cited challenge in the controlled matrix.

Two ideas here are actually quite similar: making the repository evolve to conform to new standards (technical or legal or otherwise); keeping up with these standards. Even if the first is not strictly speaking an idea belonging to the ‘Tech Watch’ theme, we have chosen to list both in the same table.

⁴¹ This mention was also listed in the “Coordination between teams” of the “Other” table, below.

The idea of a collaborative tech watch (3) is of course something that could help alleviate the difficulty, otherwise listed, constituted by the lack of community standards (1). This can be linked with the mentions of staff training, below, also contributing to the building of a community culture.

Table 28. TTRAM Activities and Functions (A|F) : Technical infrastructure

Idea expressed	Mentions	Other TTRAM A F
Infrastructure: high costs	2	Resources
Infrastructure: high-performance servers	1	
Infrastructure: preservation compliance	1	Preservation
Infrastructure: robustness and reliability	1	
Integrating with large data management system(s)	1	Interoperability
More automated integrity check systems	1	
Non-commercial cloud services	1	
Support for containerized deployment	1	
Maintaining compliance certification	1	

Key points:

Technical common requisites (robustness, reliability, high performance) can only be achieved through sufficient funding (high costs) while sticking to non-commercial / open-source solutions as much as possible (to avoid third party dependencies). Some adaptations to existing infrastructure can be required to achieve interoperability, preservation compliance or data integrity (through self-checks systems).

Table 29. Other technology & security challenges

Idea expressed	Mentions	Other TTRAM A F
Human resources: IT	12	People & Expertise
Funding	11	Resources
Human resources: development	3	People & Expertise
Coordination between teams	3	People & Expertise
Funding: large scale storage	1	Resources
Global current situation	1	
Quality certification	1	
Recognized guidelines and standards for preservation	1	Preservation
Sustainability	1	

Key points:

As for the “Other challenges” responses, human resources (15 total) and funding (12) concerns are mentioned most often in that category. It is interesting to note an expression of concern regarding the current global situation.

Additional questions regarding technology and security

After the two questions regarding challenges and needs, respondents were asked: “These were the core questions on technology and security. Would you like to answer some additional questions or continue to the next part of the survey?”. 61 respondents checked “Yes”, triggering the next questions.

Storage and integrity

Respondents were asked to select the processes and documentation their repository uses to ensure appropriate digital object and metadata storage and integrity.

The majority of respondents selected multiple options, indicating that repositories are not relying on a single process but have instead adopted a range of practices to ensure the integrity and proper storage of digital objects and metadata (see Figure 33). Respondents could also indicate other processes and documentation that their repository uses.

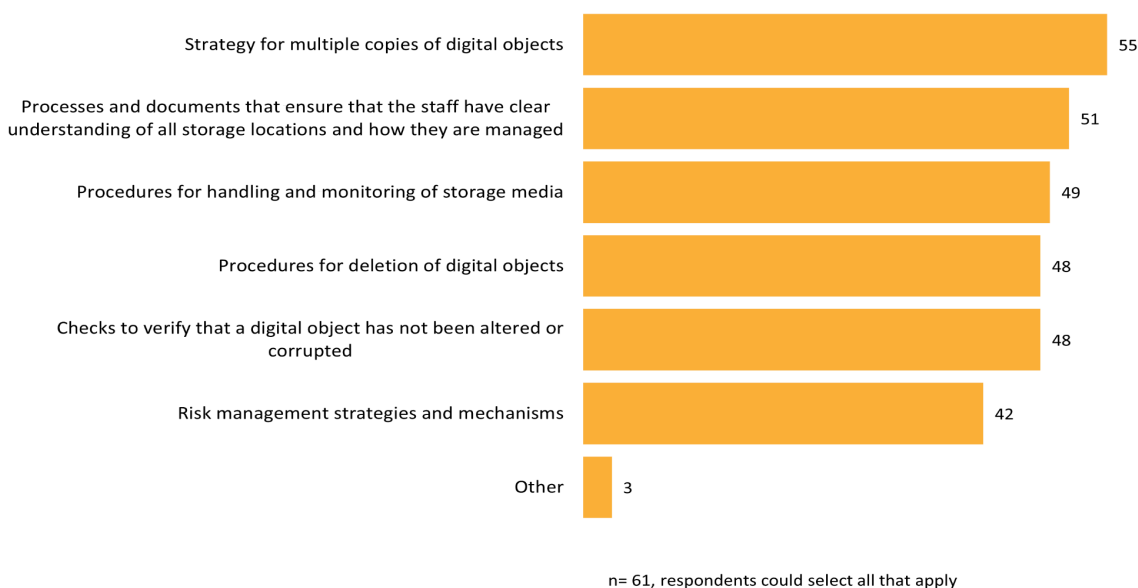


Figure 33. Measures used to ensure the storage and integrity of digital objects and metadata.

Only three respondents indicated the processes and documentation their repository uses to ensure appropriate digital object and metadata storage and integrity as “Other”. Those are audit trails and processes based on services offered by a host platform (Github for the MorpheusML model).

Other standards, best practices, and solutions for technology and security management

The respondents were asked to indicate up to five main standards, best practices and solutions used for technology and security management, and specify the purpose for each of them in a free text matrix. The same model was previously presented for metadata schemas, catalogues and registries and federated services.

There were 34 answers with a total of 112 different mentions.

Unlike we did for these previous free text matrices, we chose to not encode the answers to the “purposes” part of the matrix, as they were too heterogeneous (various specifications, software solutions, etc.). We did this for the standards named, however, and grouped the entries into 9 categories as generic as possible to achieve a general view of the whole set (see Figure 34).

This list is, as one could expect, dominated by the familiar doublet of system administration and repository development. Some other, interesting mentions follow right after.

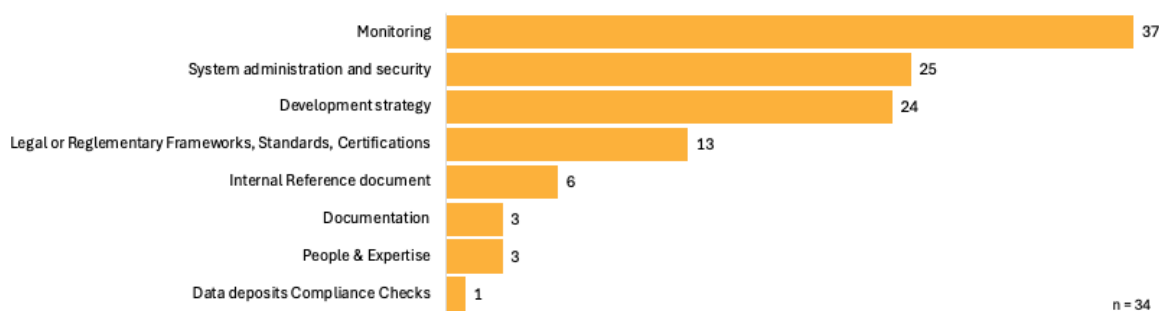


Figure 34. Main other standards, best practices and solutions for technology and security management.

A breakdown of these categories results in the following:

System administration and security (25) and Monitoring (36)

Most of the items mentioned fall into these two categories.

Responses regarding **system administration and security** exclusively mention practices and no products or software solutions, probably due to the fact it is a core activity. The set of services making up the security system of the AMRES (12 mentions) would be the only exception. The number of mentions for each of these practices is probably not significant, as the granularity of the answers is not controlled. For instance, integrity checks are mentioned only once but one could assume that the practice is more widely spread. For consistency, we provide the numbers in parentheses.

- Backup strategy (3 total) with onsite and offsite copies (1)

- DSpace collection policies and group roles, also standardised background scripts and reports for checking space usage, fixity checks, etc. (1)
- Frequent software updates (1)
- Integrity checks (1)
- RBAC (Role-Based-Access-Control) (1)
- Regular security scans (1)
- System virtualization (1), Frequent software updates (1)
- Authentication strategies (4): federated authentication through Shibboleth (1); 2-factor authentication (2); secret management (1).

Regarding the "Monitoring" category, the large number of mentions (36) might be biased due to the high number of solutions provided by the repositories originating from Serbia (28 out of 36 mentions) :

- "In-house monitoring and alerting service MONIT" (11 mentions)
- "ARGO service monitoring": "Uptime and certificate validity monitoring" (8 mentions)
- "In house traffic monitoring" (4 mentions) or "Traffic monitoring": "Monitoring of the quality of service and connections" (9 mentions)

Otherwise, 4 mentions of such services exist, three of which without a name, the last one referencing the Nagios solution.

Development strategy (25)

The sister activity of maintaining and running a repository on a daily basis is its development. Regarding this, we the answers contained the following mentions:

- Agile development or Continuous Integration/Continuous development (4)
- Two respondents used the free text response to mention the use of Composer for dependency management
- DevOps (1) or DevSecOps (1), while they cover more than development, could be mentioned here.
- Git is commonly mentioned, be it on its own (1), or through the mention of an "internal git server" (11), which might commonly be Gitlab (1)
- Version Control (3)
- Software documentation is a practice referred to as "Technical documentation" (1), and two respondents mention they use the markdown or OpenAPI/Swagger formats (2).

Legal or Reglementary Frameworks, Standards, Certifications (13)

We can list here:

- national laws or frameworks, mandatory or not: UK Digital Economy Act and National Digital Preservation Services recommendations (1), or the French ANSSI recommendations (1)
- certifications: CoreTrustSeal (1), Cyber Essentials Plus (1)
- recommendations: Federated IT Service Management (FitSM) (2), ITIL (1)
- standards: ISO 27001 (2) or the Open Archival Information System (OAIS) Reference Model (3)

Reference document (6)

We found several mentions of internal, public documents constituting a summary of the practices of the repository. These documents would typically be publicly accessible and cited in the course of a CoreTrustSeal application:

- Disaster recovery plan (2)
- File naming (1)
- File types (1)
- Long-term data preservation reference document (1)
- Preservation plan (1)

People & Expertise : Staff training (3)

- Staff training: IT security (mandatory) (1)
- Staff training: Records management and archives (1)
- Staff training or certification: FitSM (1)

Other

- Data deposits Compliance Checks (1)

3.3 Legal challenges

Legal challenges pose a significant barrier to effective data sharing between repositories and organizations. To gain deeper insight into these issues, the survey included three key questions: the legal status of the repositories, the specific challenges they encounter, and their willingness to collaborate on addressing these challenges. An additional set of questions further explored the maturity and variation in legal frameworks across repositories.

Of the respondents, 81 indicated that their repository is part of a research performing organization (see Figure 35), such as a university or research institute. This indicates that the majority are tied to established academic or research institutions, meaning that they have access to legal and administrative support, but also that they may be subject to stricter compliance requirements. It will be important that the Network focuses on addressing the needs and legal challenges specific to participants within these institutional frameworks.

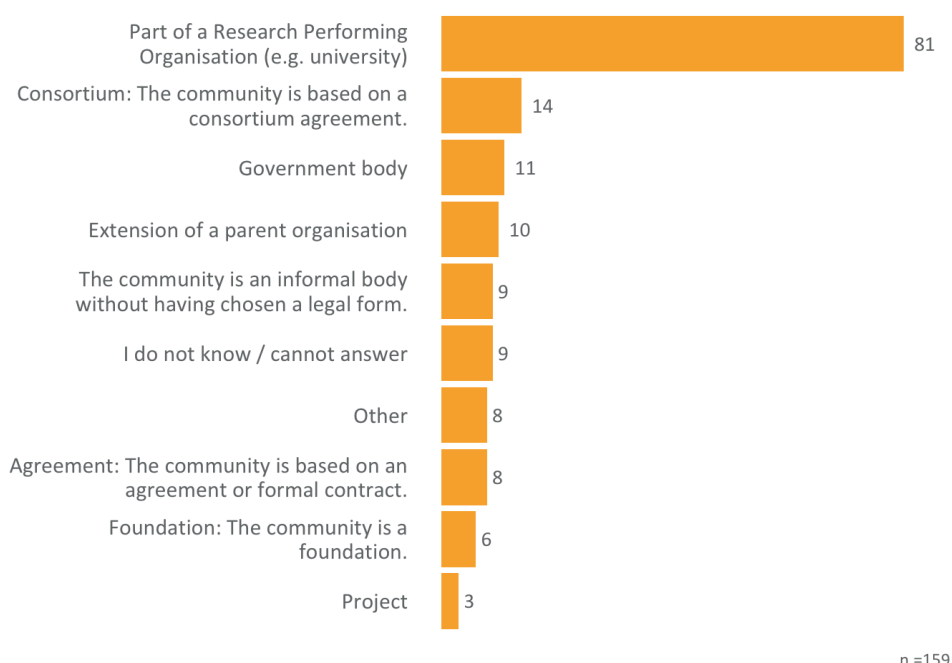


Figure 35. What is the legal status of your repository?

In response to the question about the biggest legal challenges for data sharing, there is a fairly even distribution among the options (see Figure 36). However, the most commonly reported issues were Copyright and Intellectual Property Rights (75) and GDPR (65), followed by data governance and data security (53). These results are aligned with those of another question, where many respondents indicated that their institutions hold expertise in these domains. However, qualitative responses revealed inconsistencies in the implementation and operationalization of this expertise, indicating a need for improved policy alignment and practical guidance.

A recurring comment from respondents highlights a critical aspect of GDPR compliance: the responsibility for ensuring data compliance falls on researchers before the data is submitted to repositories. This means that repositories are not solely accountable for GDPR compliance, as the primary responsibility lies with the researchers who generate and handle the data prior to submission. However, repositories still play an important role in providing clear guidelines, ensuring that their processes align with legal requirements, and supporting researchers in understanding and adhering to GDPR standards.

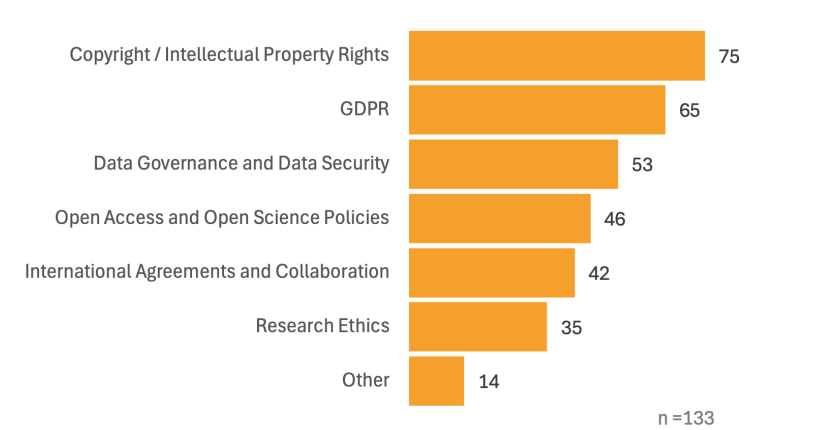


Figure 36. What would you say are the biggest legal challenges your repository faces when sharing data between repositories or with scientists from other research organisations?

When asked about collaborating on frameworks for GDPR and Intellectual Property Rights, 44 respondents expressed interest, 22 declined, and 7 cited lack of time despite their interest (see Figure 37). While there is clear support for collective solutions, practical constraints may limit participation. However, the Network should facilitate the establishment of working groups focusing on these legal issues.

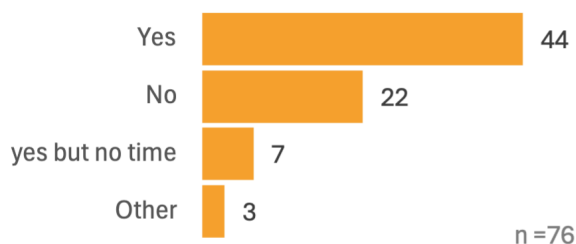


Figure 37. Would your repository be interested in collaborating on frameworks and guidelines regarding GDPR and/or Intellectual Property Rights with others in the Network?

Additional legal questions

Following the three initial questions, respondents were asked if they wished to answer a set of additional questions regarding legal challenges. 24 respondents chose to answer additional questions.

Of the respondents, 38 reported having clear GDPR-compliant frameworks, while 12 did not (Figure 38). This highlights the need for network collaboration to share best practices and ensure all members meet a common minimum standard.

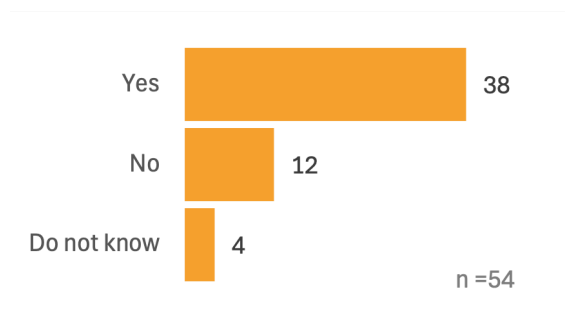


Figure 38. Does your repository have clear legal frameworks, policies and guidelines for handling personal data in accordance with GDPR and other relevant data protection laws?

Of the respondents, 36 indicated that their repository has clear frameworks for sharing data within Europe, while 14 do not (see Figure 39). Similar to the previous question, this highlights an opportunity for the Network to facilitate the reuse of existing frameworks and establish working groups to support members in developing and implementing effective guidelines.

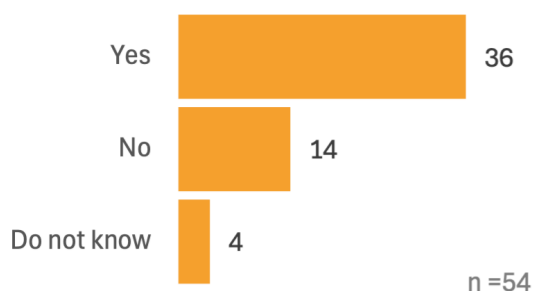


Figure 39. Does your repository have clear frameworks and guidelines for sharing data within Europe?

When asked whether their repository follows specific legal metadata standards for data governance and compliance (Figure 40), respondents most frequently indicated the use of Dublin Core (36),

followed by DDI (14). PREMIS and DCAT were less commonly used indicating that these are used in more specialised contexts. This highlights an opportunity for the Network to explore how metadata standards are applied, and a discussion on how they can promote greater interoperability.

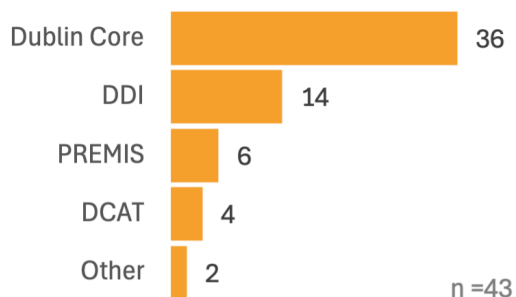


Figure 40. Does your repository follow one or several specific legal metadata standards for data governance and compliance?

Out of 40 respondents, 32 expressed openness to adopting or changing metadata schemas to improve legal interoperability (see Figure 41), highlighting strong community interest in enhancing interoperability and aligning with common standards.

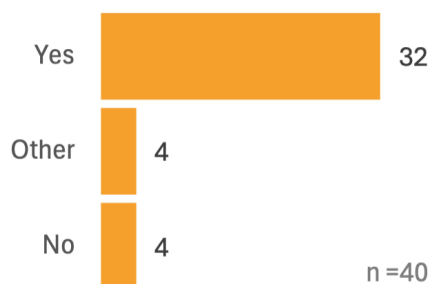


Figure 41. Would your repository be open to change to/adopt/add another core metadata schema if that would improve legal interoperability?

3.4 Training and support needs

FIDELIS will offer different support and training mechanisms throughout the project. Moreover, training and peer-to-peer support is envisioned as an essential part of the long-term future of the FIDELIS Network. To craft a training and support strategy in line with the community needs, the survey included questions regarding topics for training as well as other characteristics such as format and delivery.

Respondents were asked where they find and/or receive training and support on topics related to Trustworthy Digital Repositories (see Figure 42). The most frequently selected options were specific communities (with CESSDA, ELIXIR, RDA all mentioned in a free-text follow-up question), training initiatives from European projects (such as EOSC-Nordic⁴², FAIRsFAIR⁴³, FAIR-IMPACT⁴⁴), and online training offerings. These responses indicate rich sources of existing training initiatives and materials that FIDELIS should not aim to duplicate, but rather to highlight, enrich, or collaborate with. Interestingly, receiving peer support from colleagues at the host institution and from (international) peers in other institutions were both also often selected. This shows an existing habit of peer-to-peer support in the community, which FIDELIS can harness and uplift with our Network initiative.

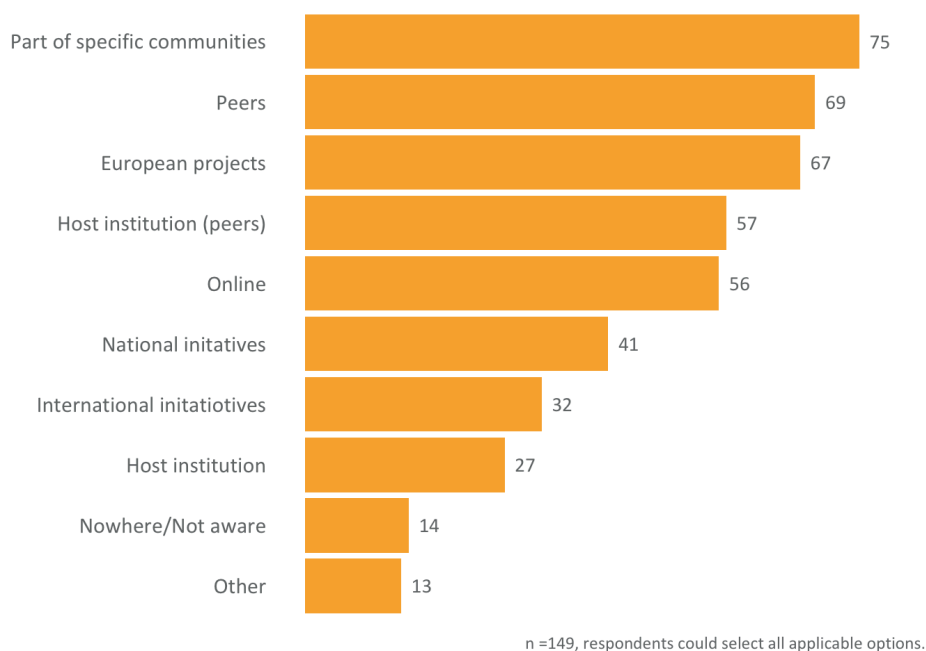


Figure 42. Where do you currently (or have you previously) find/receive training on topics related to Trustworthy Digital Repositories?

⁴² <https://eosc-nordic.eu/>

⁴³ <https://fairsfair.eu/>

⁴⁴ <https://fair-impact.eu/>

Of specific interest are both the topics and the formats in which the community would like to receive training and support. For topics, respondents were asked separate questions about what they currently experience and what their desires are for the future. The data shows that most training and support has already been received on topics related to FAIR-enabling qualities, persistent identifiers, licensing and reuse, and FAIR assessment (see Figure 43, the blue bars). On average, repositories have already received training and support on 4.5 different topics, with answers varying from one to thirteen different topics. Other topics that respondents already had received training and support on ranged from train-the-trainer capability building to different research skills.

Respondents were asked whether the training and support they previously received was sufficient for their needs. Of the 46 responses received, 28 were generally satisfied with the training and support received. However, in many cases these answers also included the additional experience that there are many different topics of relevance for repositories and important tools, standards, and practices develop very quickly, making it difficult for training to remain sufficient for a longer period of time. Repositories with fewer staff and less effort experience challenges in dedicating time to training. When these repositories do dedicate time, it is hard to improve the practical skills of their staff when training is focused on high-level or theoretical topics.

When we examine respondents' interests for receiving (more) training (see the orange bars in Figure 43), development and sustainability of a repository ranks ties with long-term preservation as the leading options on 81 responses each. Interest in long-term preservation training matches the finding presented in section 3.1 that many repositories are trying to implement active preservation in their repository. Other topics that many respondents were interested in include repository certification, the development and sustainability of a repository, and access conditions and sensitive data.

The largest gaps between having received training and holding an interest in training were found with persistent identifiers (87 to 43) and FAIR-enabling (103 to 67). In these cases it seems there has been sufficient training with not much interest for (more) training in that area. Other topics that were suggested by respondents included metadata standards, repository interoperability, and repository governance and management.

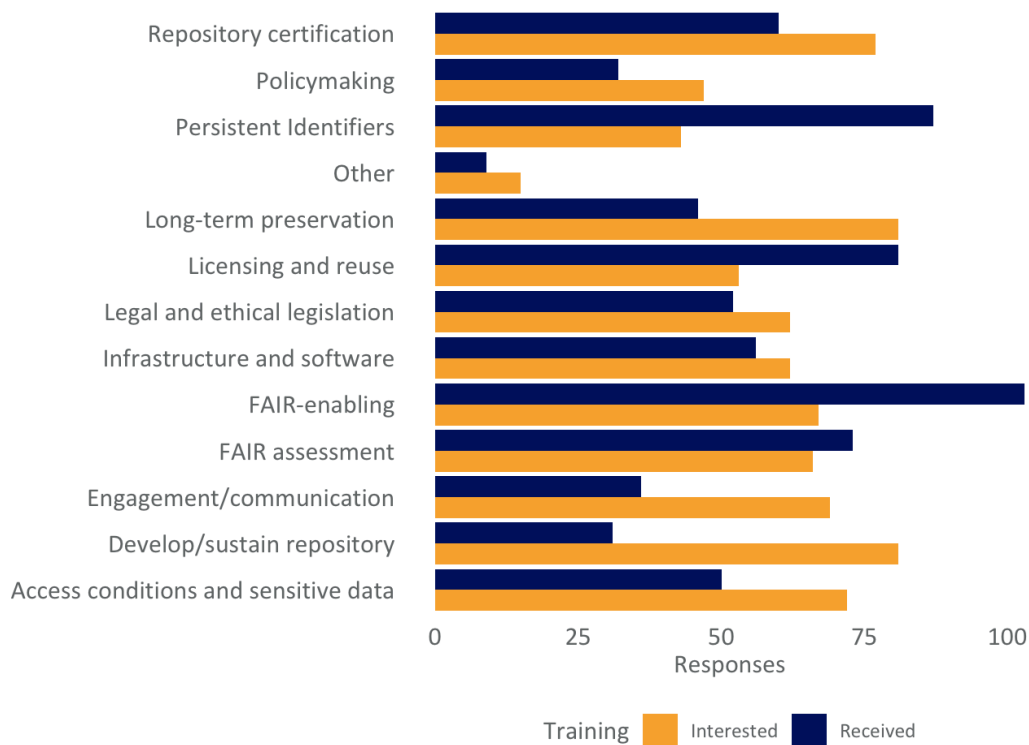


Figure 43. Received (blue) and desired (orange) training and support topics.

Seminars and webinars are the leading format for training and support respondents have experienced. The high prevalence of peer support already noted above can also be observed here, with many respondents indicating that they have received peer support to learn (new) skills. It provided the largest gap between a support format previously experienced and of interest, a difference of 41 respondents. Personal mentoring, open hours, and financial support have been experienced least often.

When asked which formats of training and support respondents would like to receive, seminars, webinars (106), and interactive workshops (97) led the responses (see Figure 44). Expert guidance and peer support were also valued highly, with 81 interested in this. Sixty-three respondents would like to receive financial support, whilst only 19 had received this before. This was the largest gap between interest in a form of support and having received this form of support. Personal mentoring and dedicated time to work independently were indicated to be the support approaches with the least support, though still indicated with interest by around 40 respondents.

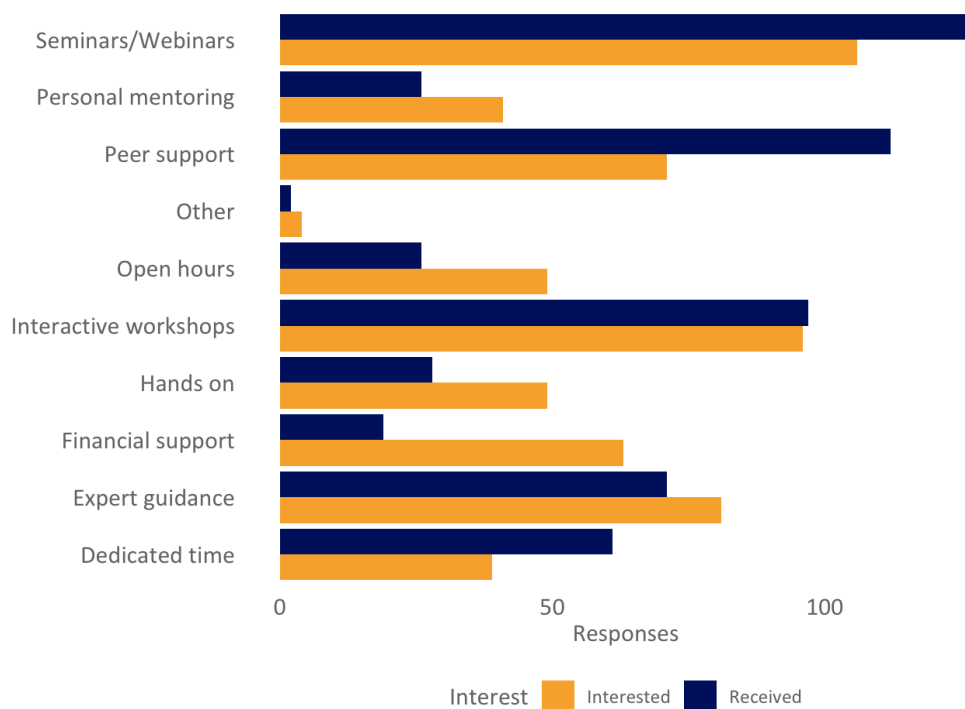


Figure 44. Received (blue) and desired (orange) training and support formats.

Respondents were asked about the importance of different qualities and characteristics of training and support in two sets of statements. In this first set of answers (see Figure 45), it can be seen that achieving a specific objective or outcome by the end of a training is valued by most respondents (116 agree or strongly agree). This matches the previous finding that higher-level and theoretical training often do not feel sufficient to justify the effort spent on the activity. Other highly valued characteristics, are when training and support include opportunities for peer engagement (97 agree or strongly agree), when training and support is broken up into multiple shorter activities instead of one longer activity (83 agree or strongly agree), and when learners can train at their own pace (76 agree or strongly agree).

Another positive result that can be observed is that many respondents indicate that their organisation encourages them to join training and support activities (98 agree or strongly agree). Responses shift more towards indifference or disagreement when it comes to creating a tangible product in training, receiving digital badges or formally recognised certificates for training, and learning on your own.

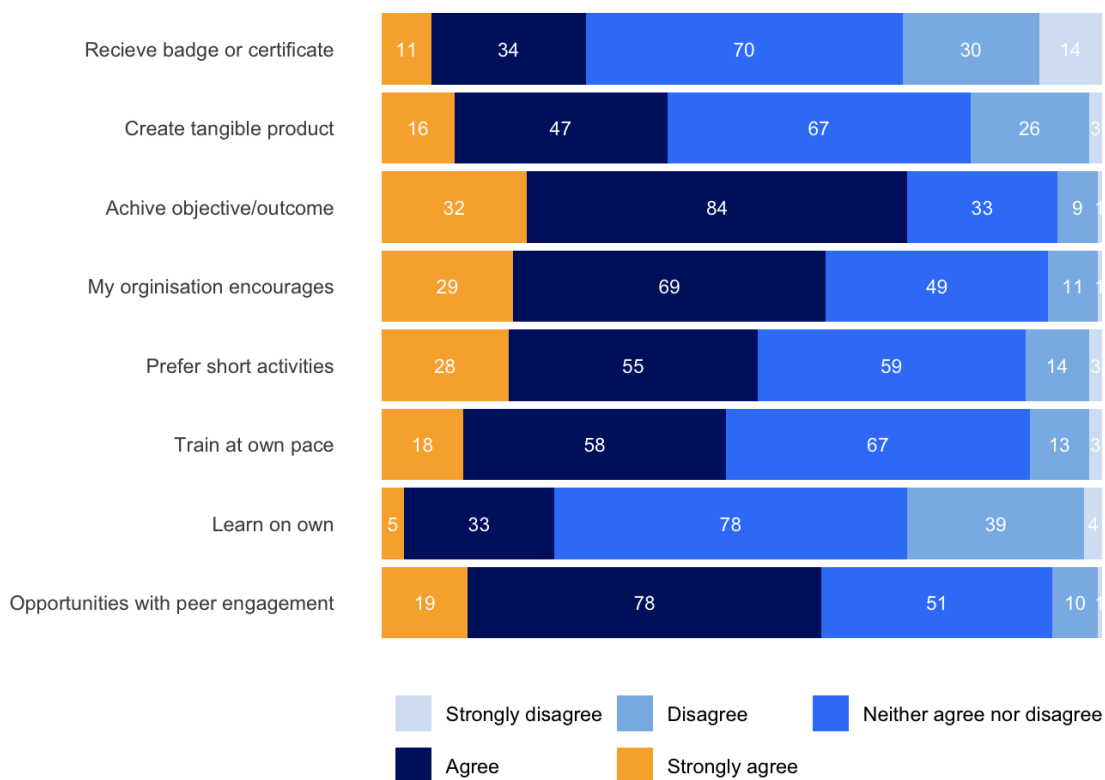


Figure 45. Desirability judgement of some training and support characteristics.

In the next set of questions, respondents were asked to indicate the importance of a set of statements about different characteristics of training and support (see Figure 46). There is a general trend of emphasis on the importance of all elements. Deemed most important are the ability to connect with, learn from, and network with fellow participants, organisers, (guest) speakers, and other experts during the activity, and for the activity to be at a more advanced level. This aligns well with previously observed answers where peer support and granular, specific, and practically-focused training and support is most desired.

Respondents also find it important that the training does not take up too much of their time and that it covers new and cutting-edge topics, following closely the quickly-developing landscape. Slightly less important are the ability to join an activity with a team of colleagues, which could be due to the fact that not all repositories have bigger teams of people on their staff for this to be applicable. A mix of different teaching formats is also deemed somewhat less important, which matches with the previous findings on training formats, where information sharing seminars and webinars were most valued. Beginner level training was still deemed important by 82 of the respondents, with around thirty respondents indicating this is not what they are looking for at all.

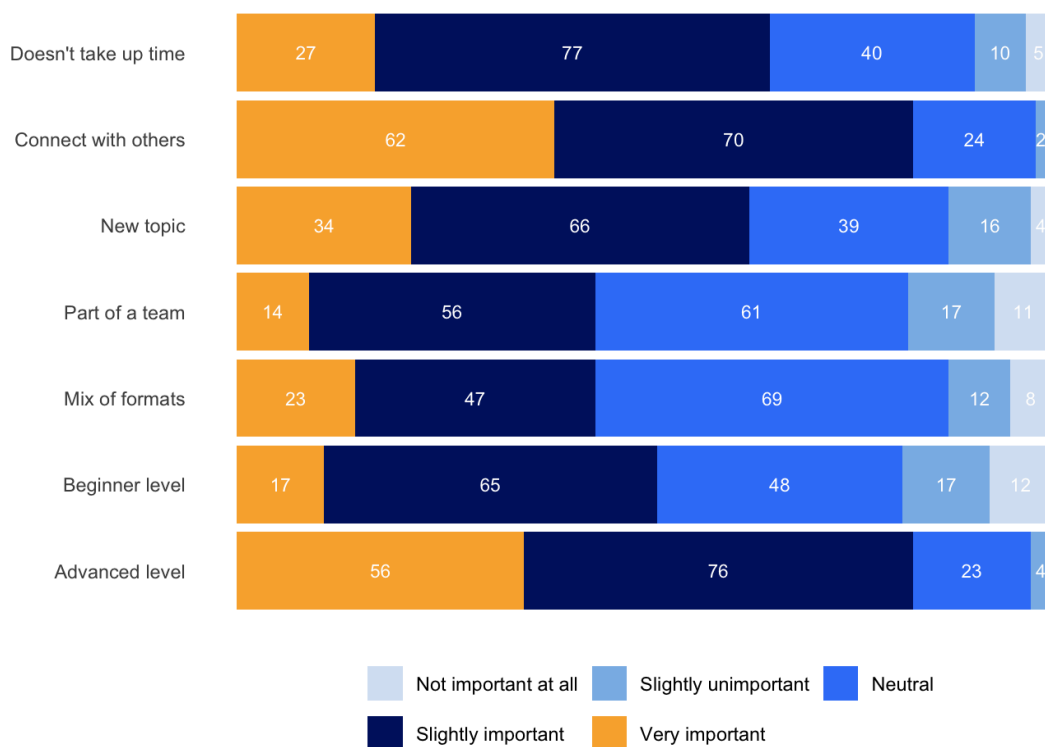


Figure 46. Judgement of importance of some training and support qualities.

Lastly, respondents were informed about the training and support programmes that will be part of the FIDELIS project: open training activities, support for the adoption of solutions, and the mentoring programme⁴⁵. Based on the initial descriptions of these training and support types, respondents were asked about their initial thoughts around potentially joining such activities organised by FIDELIS (see Figure 47). The open training activities were generally valued highly, which matches both the previously observed preference for seminars and webinars, and the indicated desire for support on repository management and governance topics. Also initially well-received is the support for the adoption of solutions. This matches also the indicated importance of interaction with peers and experts, as well as the desired practical focus and specific objectives.

Respondents were less sure about the mentoring programme, with answers more evenly split across the positive, indifferent/unsure, and negative. Looking at the written explanations for respondents' choices, the mentoring programme is the least well known and thus leads to more uncertainty around the content and design, the expected time and effort, and the skills required from a mentor. While some expressed nerves around taking on the role of a mentor, others felt more ready to step into such a role and offer their expertise to the community.

⁴⁵ <https://eden-fidelis.eu/fidelis-training-support-area>

Other often repeating considerations for joining specific FIDELIS training and support activities are that not all topics are of relevance for each person or repository, which understandably leads to some individuals to not participate in training on some of the topics. It was also mentioned often that time and effort are a very limiting factor and it would need to be very clear what value and benefit one would gain from the activity to be able to justify participation. In these answers, the sentiment is again observed that for many high-level topics there are existing resources that should not be duplicated, and that training on more practical and in-depth topics that focuses on the latest developments is most desired.

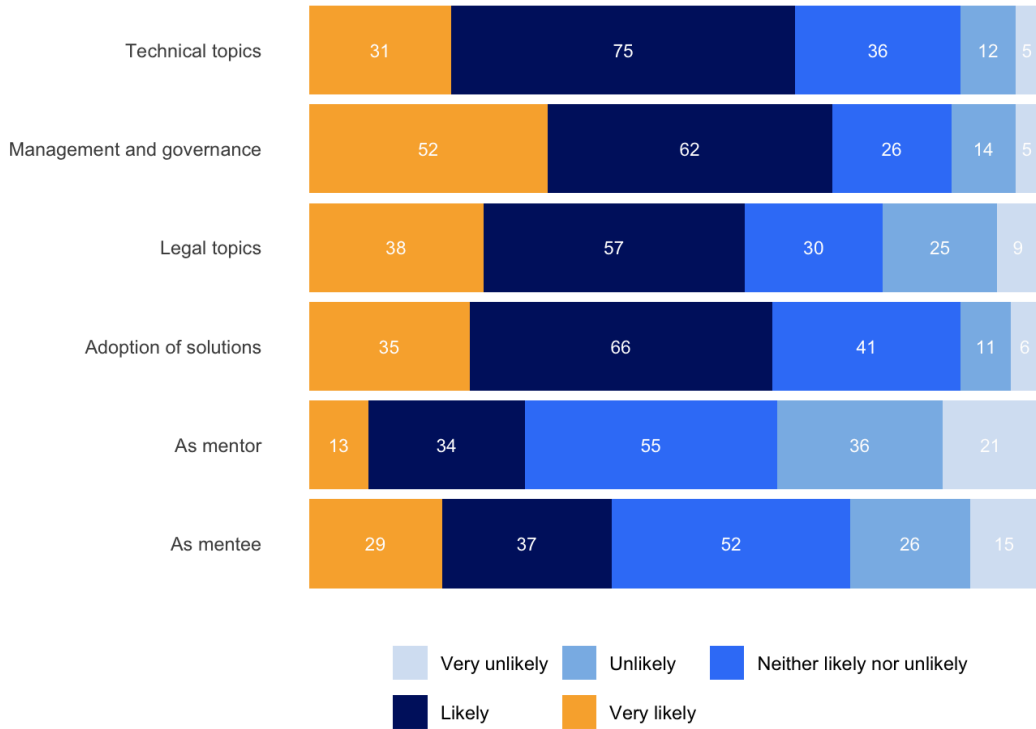


Figure 47. How likely do you currently think it is that you would join or apply for each activity?

3.5 Expectations of the FIDELIS Network

Aside from the valuable information provided for the different Pillars in the FIDELIS project, the survey also asked respondents about their initial thoughts on and expectations of the FIDELIS Network that they were now briefly introduced to. The final questions in the survey captured input of value for the overarching task of building the Network.

The FIDELIS Network was introduced to the respondents as a network of Trustworthy Digital Repositories, which will support the development and growth of TDRs within the EOSC ecosystem; foster harmonisation and interoperability across repositories to enable an EOSC federation of TDRs; and strengthen the upskilling of repositories and expand the Network through an active training and support programme. After this initial introduction, respondents were asked what they felt would be the most important benefits of such a Network for the community. Answers reflected that respondents appreciated the opportunity to connect with other repositories to share expertise, experiences, resources, and work together on addressing common challenges. Some answers reflected on the importance of “finding out where in Europe similar challenges are being grappled with, what solutions are being proposed, and how we can collaborate to create synergy.” Another highlighted the benefit of “bringing people with similar challenges together” to foster collective problem-solving. Interoperability, harmonisation, and the development and implementation of shared best practices, frameworks, and guidelines were also frequently mentioned (see Figure 48).



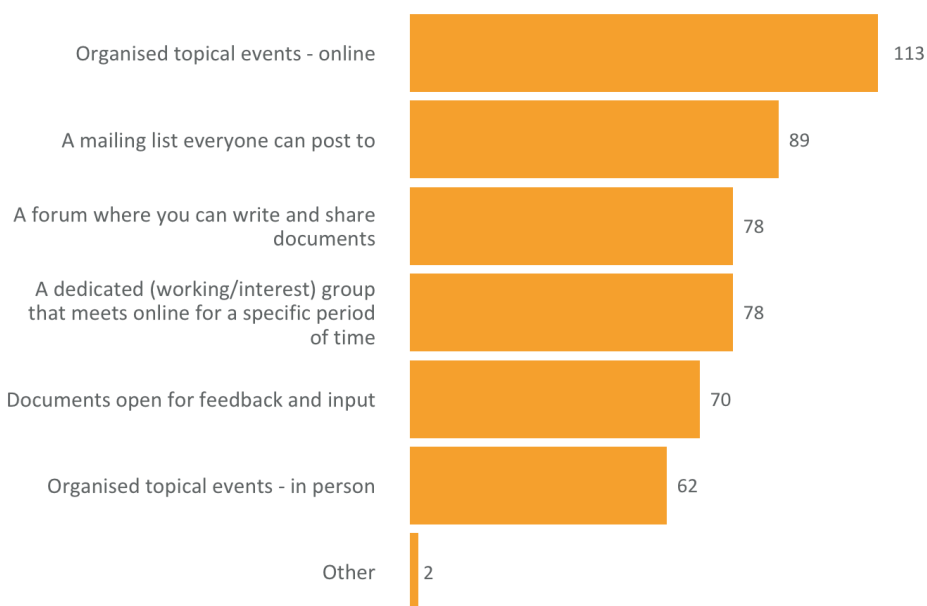
Figure 48. What do you see as the most important benefit(s) this Network should provide to you or your organisation?

Training and support in different formats were appreciated, and respondents could see themselves receiving, but also organising, such efforts. Advocacy was also a recurring theme, where the Network could function as the common voice for repositories to promote and support their needs when it

comes to policy and funding. Being visible as a part of the FIDELIS Network was seen as something that would boost the reputation and visibility of the repository, and would provide opportunities to improve the repository through the different mechanisms. Naturally, some respondents remained somewhat unsure about what any network, or this specific Network, could offer them. However, the answers generally indicated some clear desires and envisioned benefits.

Considering collaborative activities to foster harmonisation, respondents were asked which topics they felt could benefit from such an approach. Some topics suggested included: legal challenges, security, funding, preservation, data discovery, technical development, metadata, curation, repository management, citation, and more. Generally, almost all topics covered throughout the survey were mentioned as potential candidates for collaborative harmonisation efforts. Figure 49 shows the opinions on what such a collaborative mechanism could look like.

Respondents indicated a preference for online events organised around specific topics, followed by a mailing list or forum to share information and documents with each other, or the creation of dedicated working groups focusing on a topic for a specific period of time. Expectations of such mechanisms included many of the aforementioned benefits. Respondents desired the ability to share knowledge and experiences, work together to come to shared best practices, solutions, and guidance, pool together resources, and advance harmonisation and interoperability. The survey asked what respondents felt they would be able to contribute to the Network and such collaboration mechanisms. Responses included: contributing resources, people, expertise, time, support and training, which all aligns well with the type of collaboration that FIDELIS has envisioned.



n =145, respondents could select all applicable options.

Figure 49. What format of collaborative mechanisms would you be most likely to join?

The FIDELIS Network is seen as an enabler for practical collaboration, including the creation of “working groups focused on implementing specific features or certifications”, as well as opportunities for “collaborative applications for funding and projects.” By addressing large-scale issues that are better tackled collectively than individually, aligning resources, and facilitating cooperation, the Network has the potential to become an important platform for building capacity, strengthening connections, and driving innovation within the repository community.

The survey also asked about potential barriers to joining the Network. Figure 50 shows that financial concerns are most often envisioned as a barrier to joining the Network, followed by technical and organisational barriers. Such barriers can be expected as they require some form of investment from the repository (money, time, technical work to implement solutions) which need to be justified in terms of the value they will bring the repository in return. Such justifications will always be challenging to assess. These barriers are important to take into account in the design of the Network.

One existing mitigation to these financial barriers is through the mechanism of the provisional membership that was started in April 2025⁴⁶. This free membership for the duration of the FIDELIS project will allow repositories to get a sense of the value they can get out of the Network without an immediate financial investment. Moreover, provisional members will be able to help build the Network together with the project, helping to ensure that they get the value that they envision.

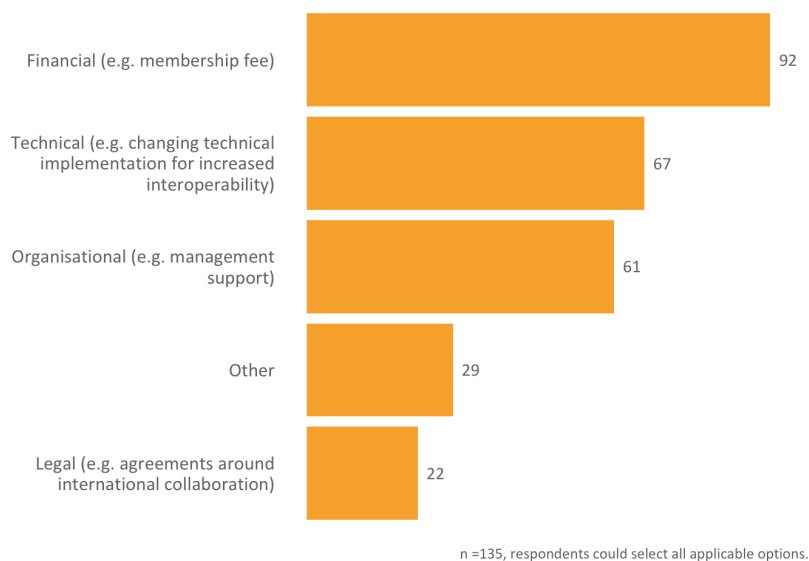


Figure 50. Expected barriers to joining the FIDELIS Network.

Lastly, respondents were asked what other ideas they had about how the FIDELIS Network could help them. Answers mostly emphasised that a practical and focused approach would likely reap the most benefits, working on one challenge at a time with very practical and tangible outcomes before moving on to the next. Offering clear guidelines, frameworks, standards, and tools that can then be

⁴⁶ <https://eden-fidelis.eu/news/fidelis-network-trustworthy-digital-repositories-has-officially-launched>

endorsed, adopted, and adapted by the Network members was also envisioned. Other answers demonstrated a desire for clear definitions of important terms, such as 'Trustworthy Digital Repository' and 'preservation'. Connecting and onboarding to, and informing about, EOSC was also mentioned as an area that the FIDELIS Network could support.

4 Conclusions and next steps

The aim of the FIDELIS landscape survey was to collect community input that would inform the different parts of the project and the building of the FIDELIS Network for Trustworthy Digital Repositories. Considering the results from the previous sections, this last chapter describes how FIDELIS will take the findings on board in its developments.

Repository capabilities and standards

The survey results confirm that the repository landscape is very diverse. Sometimes this diversity can be explained by specific disciplinary, governance or other factors; sometimes there are possibilities to converge practices, streamline processes, leverage shared resources and even achieve cost efficiencies.

General information on this variety combined with the detailed knowledge about the actual repository practices, as well as the challenges that repositories are facing, will be used by FIDELIS Pillar 3 Trustworthy Digital Repository Harmonisation and Interoperability to plan for the next two years.

The survey results provide information about the gaps where FIDELIS can provide recommendations or solutions, and enable Pillar 3 to identify and select standards, frameworks, policies and practices that can be used, established or developed further to enhance interoperability and federation in the context of EOSC. Common areas for cooperation and alignment between the repositories in the Network can range from common policy templates to reducing burden, challenges, and risks through shared services.

The survey results will also be used to improve the TTRAM model and its activities and functions, and to inform our work on a tiered approach to capability and maturity of repositories.

In terms of repository capabilities and standards, the survey analysis revealed the following key conclusions:

1. **Curation, Preservation, and Metadata Quality:** These areas emerged as critical challenges, with repositories facing resource constraints (particularly in terms of qualified personnel), desirable but scarcely present automation, and a lack of standardized practices. Metadata quality and standardization are particularly pressing concerns for enhancing discovery and long-term usability.
2. **Interoperability and Semantic Alignment:** Repositories face challenges in integrating external platforms, aligning metadata standards, and synchronizing metadata records. While common metadata schemas and persistent identifiers (PIDs) are widely adopted, there is a clear need for richer, multilingual metadata to enhance discoverability and usability. Additionally, repositories often embed or hardcode semantic artefacts within their systems rather than integrating with centralized semantic artefact catalogues. This limits

standardization and complicates the management of artefact complexity, such as versioning and heterogeneity. Addressing these issues will be critical for fostering interoperability and enabling seamless networking across repositories.

3. **Governance and Sustainability:** Governance fragmentation, funding instability, and legal compliance (e.g., GDPR) are significant difficulties. Repositories require clear governance models, sustainable funding mechanisms, and shared policies to ensure long-term success.
4. **Technical Infrastructure:** Robust, scalable infrastructure is essential for managing sensitive data, ensuring preservation compliance, and maintaining repository reliability. Non-commercial solutions and automated integrity checks are particularly needed.

Legal Challenges

A significant outcome of the survey is the fragmented legal landscape within which repositories operate. Many repositories lack formal frameworks for GDPR compliance or cross-border data sharing, and responsibility for legal compliance is often delegated to researchers without sufficient institutional support. Intellectual property rights (IPR) and inconsistent licensing practices also emerged as recurring challenges, limiting interoperability and data reuse.

While a few respondents described promising workflows involving intake appraisal, consent templates, and use of open licences, the majority indicated a lack of formalised processes for legal interoperability. The FIDELIS Network can provide valuable support by promoting shared legal templates, facilitating peer-to-peer learning, and coordinating working groups on GDPR, licensing, and metadata standards. However, the success of these initiatives will depend on the willingness of members to actively contribute to shared solutions and engage in collaborative development. Creating a sustainable, legally interoperable data-sharing ecosystem will require collective investment from across the community.

Repository Training & Support

The Repository Training & Support pillar will use the findings from the survey to help create our training strategy. This informs choices around the design and formats of the training offered, the topics covered, and how to communicate effectively with the community about the activities on offer. Aside from the topics indicated in the questions in the dedicated section, we also take away interesting findings from the other sections to potentially craft training and support around important standards and challenges experienced in the different topics. Such training and support efforts will be a collaborative effort with the other parts of the project and potentially external projects and organisations. A clear desire for more advanced, practical, and granular activities was expressed across different parts of the survey, which we will take on board in designing our training offerings. Seminars, webinars, and workshops were most appreciated and respondents' desires for technical and repository management topics map well to the areas of focus that we already envisaged for our training and support activities.

Peer support was highly valued and desired, while there was also some uncertainty around the mentoring programme where this type of support will be showcased most. It is possible that these uncertainties originate from a lack of understanding about what mentoring entails. This could be alleviated by explicit information and instruction around the type of support, which we will be able to provide on our website during the launch of the call. The survey has been useful in building our awareness of some of the worries and assumptions that we should address in advertising the call. In general, it is important that we provide clear, extensive, and complete information around the support offered to allow repositories to accurately judge the value they will get out of each activity, compared to the expected time and effort to spend.

The survey results also highlighted again the existence within the repository landscape of a wealth of existing training resources as well as communities which already offer training and support to repositories. Harnessing what already exists instead of duplicating effort is essential to meet the current needs of our potential members. Contributing to the rich landscape through creation and collaboration was the aim of FIDELIS from the outset, but the survey answers have pointed to some specific communities and resources to investigate.

When using the results of the survey to inform the drafting of the FIDELIS training strategy, we should remember that the responses were skewed towards more “EOSC-ready”⁴⁷ countries. It will be important to continue to consider the other perspectives in the community in the development of the training strategy, for example, repositories that have a need for beginner-level training, and the opportunities that are created by financial support for repositories with limited resources. FIDELIS is committed to targeting less developed repositories and countries where lower rates of certification are attained to enable them to become active members of our Network.

Expectations of the FIDELIS Network

The final part of the survey returned results that inform us about community expectations regarding the FIDELIS Network. Respondents’ answers showed a general enthusiasm for the potential benefits of the Network, and contributed a high number of ideas for such benefits, collaboration mechanisms, and topics of relevance. The already-stated aims of the FIDELIS Network align well with the survey respondents’ desire for connection, advocacy, and knowledge exchange.

Respondents stated that they believed that the most effective way to establish the Network would be through granular and practical work on specific topics with tangible results, which can then be endorsed or adopted throughout the Network. Communicating through, for example, a forum, alongside opportunities to attend (online) events and meetings is most highly appreciated. Repositories deal with barriers to participation in a membership-based network, most notably if there is a membership fee. FIDELIS’ proposed free provisional membership provides a useful intermediate solution for members to participate in the creation of the Network.

⁴⁷ <https://www.eoscobservatory.eu/eoscreadiness/2023/practices/nationalMonitoring/all>

This report, the data, and the relevant analysis scripts will become available for public reuse. The outcomes will also be shared with the EOSC EDEN project⁴⁸, relevant EOSC Association Task Forces⁴⁹, existing and forthcoming EOSC Nodes⁵⁰, and other relevant partners in the EOSC ecosystem to share this extensive snapshot of information on the current repository landscape. These different initiatives may use these insights to inform their own activities and areas of focus for the future.

⁴⁸ <https://cordis.europa.eu/project/id/101188015>

⁴⁹ <https://eosoc.eu/eosoc-association/eosoc-task-forces/>

⁵⁰ <https://eosoc.eu/eosoc-about/building-the-eosoc-federation/>

5 Appendices

Appendix A - Survey glossary

This glossary accompanied the survey to support respondents in understanding the terms used.

Access: Activities/Functions that include determining the appropriate access routes (e.g. download, secure remote environment, etc.) and mediating the process of providing access to users. Access activities/functions are defined by digital object characteristics and user characteristics that comply with Rights Management (see below). (Adapted from L'Hours & Bell 2023)

Activity/Function: An activity or function provided by a (research) repository. In this survey, Activities and Functions are grouped into three thematic areas: Digital Object Management, Organisational Infrastructure, and Technology and Security.

Analysis & Impact: Activities/Functions that include the analysis of internal business information and externally available information to ensure and demonstrate that the organisational entity is fulfilling its mission (see Mission & Scope). (Adapted from L'Hours & Bell 2023)

Archive: Curated collection or repository containing physical or digital static records, objects, metadata and data deemed suitable for permanent retention, set up and managed to established standards and models, such as ISAD(G), CoreTrustSeal, and the OAIS reference model, that ensure long term integrity, security, authenticity and accessibility of the records, objects, metadata and data. (Source: CODATA Research Data Management Terminology. Archive. <https://terms.codata.org/rdmt/archive-noun>)

Authenticity: The degree to which a digital object is what it appears to be, with evidence to prove it. Also see Provenance.

Catalogue: Index describing, indicating the location of, and recording other details of resources, materials works, etc. Curated and organised using a formal metadata schema such as MARC, ISAD(G), Dublin Core, DataCite etc. (Source: CODATA Research Data Management Terminology. Catalogue. <https://terms.codata.org/rdmt/catalogue>)

Community Watch: Activities/Functions that cover engagement with (meta)data reusers to ensure their current needs (see ReUse above) and future needs (see Preservation) are met. (Adapted from L'Hours & Bell 2023)

Conceive, Create, Collect: Activities/Functions that cover varied pre-repository phases of the research project lifecycle, e.g., engaging with (meta)data producers and (meta)data owners during

the research project planning phase - whether or not they become (meta)data depositors - to help them improve the quality of deposited digital objects and communicate the value of services, including preservation and reuse. (Adapted from L'Hours & Bell 2023)

Continuity of Service: Activities/Functions that ensure defined business continuity in line with the expected 'business as usual'. This includes all non-technical organisational, digital object management (see 'workflows') and security processes. It also encompasses provisions to recover from interruptions to the business, including disasters, as well as plans for succession when a service is discontinued. (Adapted from L'Hours & Bell 2023)

Controlled Vocabulary: List of standardised terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain. (Source: CODATA Research Data Management Terminology. Controlled vocabulary. <https://terms.codata.org/rdmt/controlled-vocabulary>)

Criteria, Assessment, Improvement: Activities/Functions that manage the processes by which internal compliance with criteria (including standards, policies and standard operating procedures) is assessed, and the measures undertaken to ensure continuous, targeted improvement. (Adapted from L'Hours & Bell 2023)

Curation, Quality & Compliance: Activities/Functions that cover steps taken by your repository to ensure deposited digital objects reach a defined level of quality and standards compliance before they are made available for reuse. (Adapted from L'Hours & Bell 2023) **Curation:** Managing and promoting the use of assets from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials. (Source: CODATA Research Data Management Terminology. Curation. <https://terms.codata.org/rdmt/curation>)

Data Compliance: Ongoing processes to ensure adherence of data to both enterprise business rules (government department, university, industry, or agency), and to legal, regulatory and accreditation requirements. Includes five areas: controls, audit, legal compliance, regulatory compliance, and accreditation conformance. (Source: CODATA Research Data Management Terminology. Data compliance. <https://terms.codata.org/rdmt/data-compliance>)

Data Quality: Reliability and application efficiency of data. Perception or assessment of a dataset's fitness to serve its purpose in a given context. Aspects of data quality include: Accuracy, Completeness, Update status, Relevance, Consistency across data sources, Reliability, Appropriate presentation, Accessibility. Data quality is affected by the way data are entered, stored and managed. Maintaining data quality requires going through the data periodically and scrubbing it. Typically this involves updating, standardising, and de-duplicating records to create a single view of the data, even

if it is stored in multiple disparate systems. (Source: CODATA Research Data Management Terminology. Data quality. <https://terms.codata.org/rdmt/data-quality>)

Deposit & Appraisal: Activities/Functions that cover the phase where the custody of digital objects is transferred from depositor to repository. Digital objects that are offered or requested for deposit may be appraised to ensure they meet defined criteria. (Adapted from L'Hours & Bell 2023)

Discovery & Identification: Activities/Functions that cover the application of persistent identifiers and associated resource discovery metadata to digital objects, the provision of resource discovery systems, and providing harvestable metadata to other resource discovery systems. (Adapted from L'Hours & Bell 2023)

External Engagement: Activities/Functions that include the means by which the organisational entity engages with external actors, including individuals, organisations, funders, partners and other bodies. Specialist engagement is required for Conceive, Create Collect, Deposit & Appraisal, Access, Support, and Training activities. (Adapted from L'Hours & Bell 2023)

Governance: Activities/Functions that cover the organisational hierarchies and processes by which the entity's mission is managed and executed. (Adapted from L'Hours & Bell 2023)

Interoperability: Activities/Functions that manage the protocols and processes by which people, processes, technologies and digital objects effectively interact with others, both within and across organisational entities. (Adapted from L'Hours & Bell 2023)

Legal & Ethical: Activities/Functions that cover the framework of legal and ethical requirements that the organisational entity must be aware of and comply with. (Adapted from L'Hours & Bell 2023)

Mission & Scope: Activities/Functions that are about defining and communicating the purpose (mission) of the organisational entity and the boundary (scope) of the activities it is responsible for. These activities may involve partners or outsourcing but your repository or data service is the entity that takes ultimate responsibility for the actions and outcomes. (Adapted from L'Hours & Bell 2023)

Network: The planned FIDELIS Network of trustworthy digital repositories (TDRs).

Ontology: Shared and standardised list of words, terms and phrases to describe components of a particular discipline or domain, along with a taxonomy of their relations. Compare this to a controlled vocabularies, which tend not to include a structure of relations between their terms. Ontologies are typically developed by domain-specific institutions or communities to aid in the precise referencing of elements. (Source: CODATA Research Data Management Terminology. Ontology. <https://terms.codata.org/rdmt/ontology>)

People & Expertise: Activities/Functions that cover the management of human resources to fulfil the mission. Includes ensuring that sufficient skills are available, internally or externally (see External Engagement). (Adapted from L'Hours & Bell 2023)

Policy & Standards Management: Activities/Functions that cover the adoption and creation of policies and standards to guide practice (see Governance and Workflows). Complies with the Legal and Ethical framework in place. (Adapted from L'Hours & Bell 2023)

Preservation: Activities/Functions that include monitoring the technology landscape (see Technical Infrastructure below) and the user community (see Reuse above and External Engagement below) for changes that affect the use or understanding of digital objects. If necessary, preservation actions are taken on the data (e.g., format migration), or on the metadata (e.g., updated ontologies), or on the whole object (e.g., emulation of the environment in which the digital object is rendered and used). These preservation actions then ensure preservation outcomes i.e. the continued viability of the digital object. (Adapted from L'Hours & Bell 2023) **Preservation:** An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology. (Source: CODATA Research Data Management Terminology. Preservation. <https://terms.codata.org/rdmt/preservation>)

Provenance & Authenticity: Activities/Functions that ensure that information is provided about the digital objects that describes their history (provenance) and provides evidence that the digital objects are what they claim to be (authenticity). Provenance and authenticity information can be sourced at the point of deposit, generated during curation and preservation actions, and shared with (meta)data users. (Adapted from L'Hours & Bell 2023) **Provenance:** A type of historical information or metadata about the origin, location or the source of something, or the history of the ownership or location of an object or resource including digital objects. For example, information about the Principal Investigator who recorded the data, and the information concerning its storage, handling, and migration. (Source: CODATA Research Data Management Terminology. Provenance. <https://terms.codata.org/rdmt/provenance>)

Registry: Database containing information about trusted repositories that are provided by repository managers and are useful for human and machine users. These registries do not contain information about all metadata descriptions of digital objects, nor do they offer a list of PIDs of all stored digital objects. They do offer information based on standardised types on how to retrieve such information (e.g., the port under which OAI-PMH can be accessed to offer metadata). A registry requires the assignment of a permanent, unique and unambiguous identifier to each item. (Source: CODATA Research Data Management Terminology. <https://terms.codata.org/rdmt/registry>)

Release & Publishing: Activities/Functions that manage the processes by which information - including papers, policies and other content - is released from your repository or data service entity. This is a 'push' activity by the entity, in contrast to the 'pull' process by which users request Access to

digital objects. It includes release and publishing of the information necessary for External Engagement, and to provide evidence during external assessment (see Criteria, Assessment, Improvement). (Adapted from L'Hours & Bell 2023)

Repository: In this survey used to include research repositories and archives as well as repository and archive services. **Repository:** Physical or digital storage location that can house, preserve, manage, and provide access to many types of digital and physical materials in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the physical and digital material to be authentic, reliable, accessible and usable on a continuing basis. (Source: CODATA Research Data Management Terminology. <https://terms.codata.org/rdmt/repository>)

Research & Development: Activities/Functions that include any work, internal or external, often delivered through projects, that falls outside 'business as usual'. Includes - but is not limited to - work that identifies and develops new and novel approaches to data, metadata, business processes, technology, security and research infrastructure. Where assessment indicates a need for improvement that falls outside of normal maintenance upgrades, this may be delivered through Research and Development (R&D). (Adapted from L'Hours & Bell 2023)

Resources: Activities/Functions that cover the organisational hierarchies and processes by which the organisational entity's human and financial resources are managed. This includes a knowledge of and an effective deployment of human, financial and energy assets. (Adapted from L'Hours & Bell 2023)

Reuse: Activities/Functions that ensure that the outcome of Deposit & Appraisal, Curation, Quality & Compliance, Access, and Preservation activities/functions result in digital objects being able to be used and understood by the (meta)data users for as long as your repository or service has promised. This includes ensuring that digital objects are furnished with sufficient information to support understanding and use over time. In cases where (a copy of) the digital object is not handed entirely over to the user (e.g., direct download to a researcher's computer), your repository or data service mediates reuse. This mediation includes the provision of remote secure access systems, safe rooms or other tools where the (meta)data remains fully or partially under the control of the service provider. Supporting reuse depends on an understanding of the community of users including, but not limited to, targeted 'designated communities'. This understanding is gained through External Engagement (see below). (Adapted from L'Hours & Bell 2023)

Rights Management: Activities/Functions that cover the management of the permissions, prohibitions and obligations related to the interactions between digital objects and actors (such as individuals and organisations), both inside and outside your repository/data service. (Adapted from L'Hours & Bell 2023)

Security: Activities/Functions that include measures that ensure the appropriate physical and virtual protection of the digital objects held by the entity and of other related information including staff and user information. (Adapted from L'Hours & Bell 2023)

Semantic Artefact: A machine-actionable and -readable formalisation of a conceptualisation, enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from loose sets of terms, taxonomies, thesauri to higher-order logics. Moreover, semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, and OWL, using XML (RDF) and JSON-LD. (Source: Franc et al. 2022) Here also used as a broader term to include controlled vocabularies, ontologies, terminologies, taxonomies, thesauri, metadata schemas and standards.

Semantic Artefact Catalogue: A broader term to include libraries, registries, listing or repositories of semantic artefacts and also platforms often named terminology/vocabulary service/server.

Storage & Integrity: Activities/Functions that cover the means by which digital objects data and metadata are stored, persisted and replicated, and by which replicated copies are validated as identical, or restored from a backup in the event of errors. (Adapted from L'Hours & Bell 2023)

Taxonomy: A controlled vocabulary with a hierarchical structure used to classify things or concepts. Terms within a taxonomy have relations to other terms (parent/broader term, child/narrower term). (Source: Franc et al. 2022)

Tech Watch: Activities/Functions that cover the process by which the wider landscape of hardware, software and information technology services are monitored. Drives changes to the provision of technology for delivering services internally and externally. Aligns with Community Watch and informs Curation and Preservation. (Adapted from L'Hours & Bell 2023)

Technical Infrastructure: Activities/Functions that include the overall provision of hardware, software and information technology service management to support the entity. (Adapted from L'Hours & Bell 2023)

Thesaurus: A controlled vocabulary following a standard structure, where all terms have relationships of three kinds to each other: hierarchical (broader term/narrower term), associative (related term), and equivalent (use/used for or see/ seen from). Some terms in thesauri might have additional explanatory notes, such as scope notes (brief explanations about the coverage of the term or of how it should be used in indexing) or history notes. Thesauri are defined in the ISO 25964. (Source: Franc et al. 2022)

Third Party Dependencies: Activities/Functions that cover the identification of and management of any third party organisations, hosts, partners and other actors that are dependencies for the entity, including for delivery of (meta)data services. (Adapted from L'Hours & Bell 2023)

Training: Activities/Functions that leverage internal expertise to train (meta)data producers, owners, depositors and users on a range of research and digital object management issues. These issues span the initial point that a research is conceived up to the point of (meta)data reuse. Also includes internal training and training delivered to peer organisations and other partners and third parties. (Adapted from L'Hours & Bell 2023)

User Support: Activities/Functions that include providing expertise and guidance and responding to requests from depositors and users around deposit and appraisal, discovery, access and re-use of (meta)data. (Adapted from L'Hours & Bell 2023)

Workflows: Activities/Functions that cover all of the processes undertaken to manage digital objects and their associated data and metadata, guided by legal, ethical, policy, and other standards and operating procedures. (Adapted from L'Hours & Bell 2023)

References of the Glossary

CODATA Research Data Management Terminology, available at <https://vocabs.ardc.edu.au/viewById/685>.

Franc, Y. L., Bonino, L., Koivula, H., Essen, J. P., & Pergl, R. (2022). D2.8 FAIR Semantics Recommendations Third Iteration. <https://zenodo.org/records/6675295>

L'Hours, H., & Bell, D. (2023). Repository & (Meta)Data Services Activities & Functions Overview. Zenodo. <https://doi.org/10.5281/zenodo.7689090>.

Appendix B - Additional tables and lists relevant to repository characteristics

To ease the reading, some elements belonging to section 3.2 (presenting the results about repository characteristics) are reported here.

The comments are in the relevant subsection.

Metadata schemas

Metadata schemas mentioned and purposes cited for them (page 37)

- **Dublin Core (54):** Core metadata for diverse elements (14); data description (11); collection description (3); discovery facilitation and discovery metadata (3); harvesting (3); metadata export (3); harvesting (client and/or server side) (2); metadata export, citational metadata (2).
- **Datacite Metadata Schema (29):** Identifier registration (14); data description (4); discovery facilitation and discovery metadata (3); harvesting (2).
- **DDI (24):** Data description (7); Data documentation (4); Data discovery (3); Discovery facilitation and discovery metadata (2); Variable description (2).
- **DSpace Internal Metadata (13):** Tracking system-level metadata (Serbian repositories)
- **Schema.org (12):** Data description (3); Citational metadata; Data dissemination; Discovery facilitation and discovery metadata; Exposing metadata on the landing page; Harvesting.
- **DCAT (5):** Data description (1); Harvesting (client and/or server side) (1); Metadata export (1); Metadata schema used for specific elements (1); Ontology/Semantic artefact description (1).
- **ISO Standards (5):** Additional data description (1), Feature description (1), Harvesting (client and/or server side) (1), Metadata description (1), Quality and Imagery Description (1).
- **OpenAire (5):** Data description
- **Clarín (4):** Additional data description
- **EAD (4):** Collection description, data categorization and description
- **METS (4):** Container for other metadata schemas, including PREMIS.
- **PREMIS (4):** Preservation metadata
- **Codemeta (3):** Metadata export, Software metadata
- **MODS (3):** (Meta)data and collection description
- **DICOM (2):** Additional data description (Medical data)
- **Europeana Data Model (EDM) (2):** Interoperability with Europeana
- **ISA-TAB (2):** Experimental and sample details; additional (Life sciences) metadata

Other schemas, with only one mention, were :

- **ABR (Archeologisch Basis Register):** Semantic: vocabularies and thesauri
- **ACDD (Attribute Conventions for Dataset Discovery):** Discovery facilitation and discovery metadata
- **BibTeX:** Metadata export
- **BIDS (Brain Imaging Data Structure):** Dataset organization

- **C4D (CERIF for Datasets):** Main metadata schema
- **CESSDA: Semantic:** vocabularies and thesauri
- **Citation File Format:** Metadata export
- **Darwin Core:** Additional data description
- **Dataverse Metadata Schema:** Main metadata schema
- **EngMeta/metadata4ing:** Engineering metadata, process metadata
- **JATS (Journal Article Tag Suite):** Publications and other digital objects description
- **MOD (Metadata for Ontology Description):** Ontology/Semantic artefact description
- **NeXus:** Data description (proton and neutron)
- **OAI-ORE:** Metadata export
- **ProteomeXchange XML:** Data description (dataset level)

Purposes mentioned and metadata schemas employed for them (Table 10) (Page 38)

Table 10. Purposes mentioned and metadata schemas employed for them

Some commonly cited purposes	Metadata schemas used
Data description	Datacite Metadata Schema, DCAT, DDI, Dublin Core, EAD, MARC-XML, METS, Other: CF conventions, Schema.org
Metadata export	BibTeX, Citation File Format, Codemeta, Datacite Metadata Schema, DCAT, DDI, Dublin Core, MARC-XML, OAI-ORE, Other: GeoJSON, Other: JSON, Other: RO-Crate
Core metadata for diverse elements	Datacite metadata schema, Dublin Core
Identifier registration	Datacite Metadata Schema
Tracking system-level metadata	DSpace Internal Metadata
Harvesting (client or server side)	Datacite Metadata Schema, DDI, Dublin Core, MARC-XML, Schema.org, DCAT, MARC-XML, Other: ISO standard.
Discovery facilitation and discovery metadata	ACDD (Attribute Conventions for Dataset Discovery), Datacite Metadata Schema, DDI, Dublin Core, Schema.org
Additional data description	Clarín, Darwin Core, DICOM, Dublin Core, ISO standard, README....
Collection description	Dublin Core, EAD, MODS
Main metadata schema	C4D (CERIF for Datasets), Dataverse Metadata Schema, DDI
Data documentation	DDI
Preservation activities and metadata	PREMIS

Shared services

Shared services mentioned and purposes cited for them (Page 41)

- Datacite (16): Finding data (1); Identifier registration (15)
- Handle System (14): Identifier registration (13)
- NomadLite (U. of Belgrade local service) (12): Link to publications; Registries: funding information
- APP (U. of Belgrade local service) (12): Access (12)
- Ellena (U. of Belgrade local service) (12): Curation (12)
- OAI-PMH (12): Harvesting (12)
- ORCID (8): Authentication tools (2); Deposit (1); PID management (1); Registries: persons (4) (2)
- CESSDA Data Catalogue (4): Data catalogues (3); Finding data (1)
- CESSDA vocabulary service (3): Semantic: vocabularies and thesauri (1); Standardize descriptions in metadata (1)
- Creative Commons Licences Registry (3): Registries: licences (3)
- OpenAIRE (3): Dataset identification (1); Finding data (1); Project information (1) (1)
- COAR Vocabularies (2): Semantic: vocabularies and thesauri (1)
- eduGAIN (2): Authentication tools (2)
- Feide (2): Authentication tools (2)
- geonames (2): Semantic: geographical metadata (2)
- ISO Language Codes (ISO 639) (2): ISO standard (2)
- Ontology Lookup Service (2): Semantic: ontologies (2)
- Persistent Identifier (PID) Registries (2): Registries: persistent identifiers (1)
- ROR (2): PID management (1); Registries: organizations (1)

Purposes mentioned and shared services employed for them (Table 11) (Page 42)

Table 11. Purposes of specific catalogues, registries, resources and services mentioned by respondents.

Purpose mentioned	Mentions	Services cited (number of mentions)
Identifier registration	30	da ra (1), Datacite (15), Handle system (14)
Semantic: vocabularies and thesauri	21	aat getty (1), CESSDA vocabulary service (2), COAR Vocabularies (2), DDI CVs (1), ELSST (European Language Social Science Thesaurus) (1), EuroSciVoc (1), GEMET (1), GND (Gemeinsame Normdatei [Integrated Authority File by the German National Library]) (1), HESA vocabulary (1), Hierarchical Event Descriptors (HED) (1), IFREMER vocabulary (1), MeSH (1), NERC vocabulary server (1), Neurobagel data model terminology (1), Neurobagel data model vocabulary (1), Pactols (1), perio.do (1), The European Language Social Science Thesaurus (ELSST) (1), TIB Terminology Service (1), Wikidata (1)
Registries: funding information	15	Agence nationale pour la Recherche registry of funded projects (1), CORDIS (2), NomadLite (U. of Belgrade local service) (12) OpenAIRE Graph, creative commons api (1)
Access	12	APP (U. of Belgrade local service) (12)

Purpose mentioned	Mentions	Services cited (number of mentions)
Authentication tools	12	CrossAsia user administration (Shibboleth authentication) (1), eduGAIN (2), Feide (2), Lithuanian E-Government Gateway (1), LITNET FEDI (federated identity service for research and education institutions in Lithuania, https://fedi.litnet.lt/en/) (1), National HAKA and VIRTU user authentication methods (1), ORCID (2), RENATER register (1), SURFconext (1)
Curation	12	Ellena (U. of Belgrade local service) (12)
Harvesting	12	OAI-PMH (12)
Link to publications	12	NomadLite (U. of Belgrade local service) (12)
Data catalogues	6	B2Find (1), Belgian GNSS Data Repository (1), CESSDA Data Catalogue (3), ESRF data portal (1)
PID management	6	DOI registry (1), HandleSystem (1), ORCID (1), PID services (1), ROR (1), URN registry (1)
Data banks	5	Human Organ Atlas (1), IANA media types, PRONOM registry (1), Koodistot (1), Language Bank Rights service (1), Paleontology data portal (1)
Finding data	5	BASE (1), CESSDA Data Catalogue (1), Crossref (1), Datacite (1), OpenAIRE (1)
Semantic: ontologies	5	AgroPortal / OntoPortal instances (1), EarthPortal (1), Finto-ontology service (1), Ontology Lookup Service (2)
ISO standard	4	ISO 3166 - Country name codes (1), ISO 639 - Language codes (1), ISO Language Codes (ISO 639) (2)

Purpose mentioned	Mentions	Services cited (number of mentions)
Registries: licences	4	Creative Commons Licences Registry (3)
Registries: organizations	4	EDMO (European Directory of Marine Organizations)(1), Research Information System of the institution (1), ROR (1), ROR, ORCID, European Directory of Marine Organisations (EDMO) (1)
Registries: persons	4	ORCID (4), SURFconext (1)
Dataset Assessment	3	BIDS Validator (1), F-UJI (1), Neurobagel (1)
Semantic: geographical metadata	3	geonames (2) Open Street Map (1)
Alignment of workflows	2	Common policies and guidelines for repository consortium (1), Common policies and guidelines with DataverseNO (1)
Deposit	2	ORCID (1)
Preservation service	2	APTrust (1), National Digital Preservation Service (PAS) (1)
Project information	2	https://gepris.dfg.de/gepris/OCTOPUS (1), OpenAIRE (1)
Registries: stations	2	https://gnss-metadata.eu/ (1), "IGS site log (standard to store GNSS station information - text based)" (1)
Access content through a federation	1	AgroPortal/OntoPortal federation (1)
Data archive	1	Aila Data Service (1)
Data models	1	Neurobagel data model (1)
Data repository	1	CrossAsia Search (based on PICA format, part of the GBV framework database) (1)
Dataset identification	1	OpenAIRE (1)

Purpose mentioned	Mentions	Services cited (number of mentions)
Discovery	1	VerbundFDB (German Network of Educational Research Data) Search (1)
DOI, diffusion and preservation	1	Nakala (1)
Enable submission to users	1	ProteomeXchange/PRIDE standalone submission tool (1)
Federated search for data across multiple data repositories of photon data	1	PaNOSC federated search portal (1)
Metadata editor	1	COMEDI (1)
Multiple purpose	1	Reposis Service of the GBV (1)
None	1	None (every service is embedded) (1)
Publication catalogues	1	EuropePubMed Central (1)
Registries: persistent identifiers	1	Persistent Identifier (PID) Registries (1)
REST API standardization	1	Optimade (1)
Standardize descriptions in metadata	1	CESSDA vocabulary service (1)
Storage	1	Fedora software (1)