Spring 2020

World Data System
Member Survey 2019

Report by the International Technology Office

Dr. Karen Payne
Associate Director for International Technology
kpayne@oceannetworks.ca

Dr. Alicia Urquidi Díaz
Research Associate for the International Technology Office
aurquidi@oceannetworks.ca

Payne, K. & Urquidi Díaz A. (2020). World Data System Member Survey 2019. Report from the International Technology Office. Victoria, BC, Canada.

DOI: 10.5281/zenodo.3840406

World Data System

International Technology Office

Ocean Networks Canada

University of Victoria

#100-2474 Arbutus Rd Victoria, BC V8N 1V8

P 250 472 4527 | wds-ito.org

World Data System
Member Survey 2019

Report by the International Technology Office

Spring 2020

Contents

Executive Summary

Between 02 April and 31 July 2019, the World Data System (WDS) International Technology Office (ITO) conducted an on-line survey of WDS members to determine the current state of institutions' data management systems and processes, and gauge their interest in technical support for existing repositories. It further served as an opportunity to introduce members of the WDS to the new ITO office. The result of this survey provided a first look at the current state of the WDS members technology, their expertise, interests, needs and expectations.

The survey was sent to 176 contacts, belonging to 139 member organizations[1]. We received valid responses from 77 unique institutions. The questions addressed a range of issues associated with repository technologies and services, along with members' interest in ITO support and ITO initiatives. The survey primarily focused on 3 research data management areas: standards and practices for exposing metadata, persistent identifiers and semantics. All survey questions are included in the Appendix.

Our key finding was an initial set of projects ITO could implement to serve member needs. Ranked by level of interest, members' top choice (1st) was adding semantic markup to metadata, followed (2nd) by harvestable metadata services. A brokering registry seems less interesting to members at this time, but ITO will remain responsive to developments in this area.

Following a brief section on the survey's methodology, responses are presented in thematic clusters and summarized in figures. Based on the responses, a set of developing ITO initiatives to support and serve the WDS community are outlined as well.

The Survey

This survey gives the ITO a first, global picture of the current state of members' research data infrastructure. It was designed to support 3 main goals:

- to introduce WDS members to the new ITO office;
- to learn more about the systems and processes currently employed by WDS members in their data management activities;
- to assess member needs and gauge their interest in potential support projects initiated by ITO.

Table 1: Survey Topics

Repository	Institution
Technology and services used	Contacts at the institution
Standards known and/or used	Relationship to other organizations
Availability of repository	Need for/interest in ITO support
Reporting/clients	Interest in ITO initiatives

Questionnaire

WDS developed a 33-question (20-30 minute) survey with questions covering the topics summarized on table 1. Some questions related to the state of institutional repositories—technologies and services used, general awareness and use of standards, repository availability, and institutions' assessment practices (metrics collection and sharing with WDS-ITO and other partner institutions)—and others to the institution's range of activities regarding their data repositories.

The question formats included closed questions, with binary or scalar options, and open text fields that invited respondents to provide details in more elaborate answers. Members were also asked to identify the technical contact for their organization, and any contacts who may be willing to serve on the ITO Technical Advisory Committee. These contacts have been added to the new ITO contact database for future reference.

All survey questions, tagged according to these categories, can be found in the Appendix.

Sample

The survey form was distributed to 176 WDS contacts using Google Forms.[2] Participants were contacted in April 2019 by the WDS International Programme Office in Tokyo. 3 new organizations joined the WDS during the survey and were not included. The survey was conducted between April and July 2019. It was distributed to 176 WDS contacts at 139 member organizations. It generated 81 responses from 77 institutions, resulting in a 57% response rate (Figure 1).

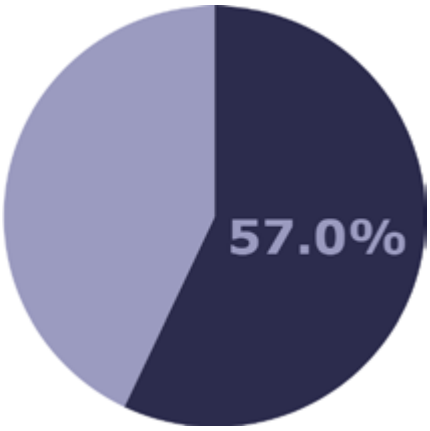

Figure 1. Survey response rate. 57% of contacted institutions participated in the survey.

As can be seen in Table 2 and Figure 2, the response rates varied between different member categories. At the top, Network and Regular members have similarly high response rates (65% combined). The lowest response rates were among Associate and Partner members, which may in part reflect their less "formal" commitment with WDS, but it is worth contemplating more targeted approaches to engaging with these member types and understanding their particular support needs.

TABLE 2: INSTITUTIONS SURVEYED

TYPE	N	n
Regular	*76*	**50**
Candidate	*19*	**10**
Network	*11*	**7**
Associate	*19*	**6**
Partner	*10*	**4**
Total	*139*	**77**

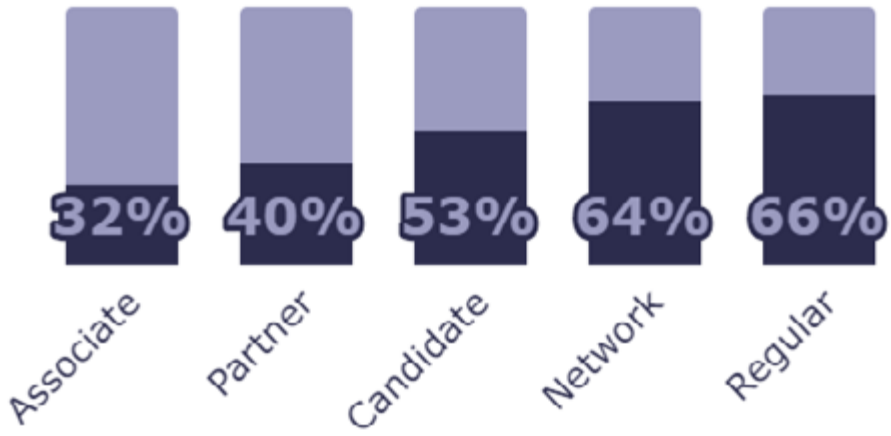

Figure 2. Institution survey response rate by membership type. Columns represent the percentage of responses (n) from each group (N).

Survey responses

Contact and information exchange

Respondents were asked what platforms or channels they would prefer to use to interact with WDS-ITO. By far, email was the preferred communication channel (70 mentions by 86% of respondents). There is also some support to stand up a searchable web forum (12) and a Listserv (11). The respondents are geographically sensitive, with Chinese participants unable to access Google services and preferring Skype for on-line meetings. In addition, one respondent added that they felt a newsletter on key updates from the ITO would be helpful.

Overall, the types of projects that generate the most member interest are: adding semantic markup to metadata (1st place), and harvestable metadata services (2nd). There appears to be less interest in a brokering registry at this time, but ITO will remain attentive to WDS members' evolving needs, while keeping abreast of any new developments in this area. Other, less frequently mentioned projects are also included for discussion in their respective sections.

Services

The primary responsibility of WDS members is distributing data. 67 of the 80 responses (84%) to the first question ("Does your Organization have a harvestable public facing metadata catalogue of its data sets?) said they either already have or are currently developing a harvestable metadata catalogue service (Figure 3). 13 respondents (16%) said they did not yet have a metadata service.

Metadata catalogue, services and protocols.

We asked the same question twice in the survey. In addition to question 3, question 10 asks "Do you supply a metadata service based on a standard protocol that can be used to harvest the catalogue content?"

These responses are somewhat inconsistent. 42 members reported offering harvestable metadata services for all of their datasets (some listing support for multiple metadata service endpoints) with an additional 14 supplying a metadata service for some, but not all, of their data holdings. Taken together, the responses to both questions indicate that 56 members have a harvestable catalogue.

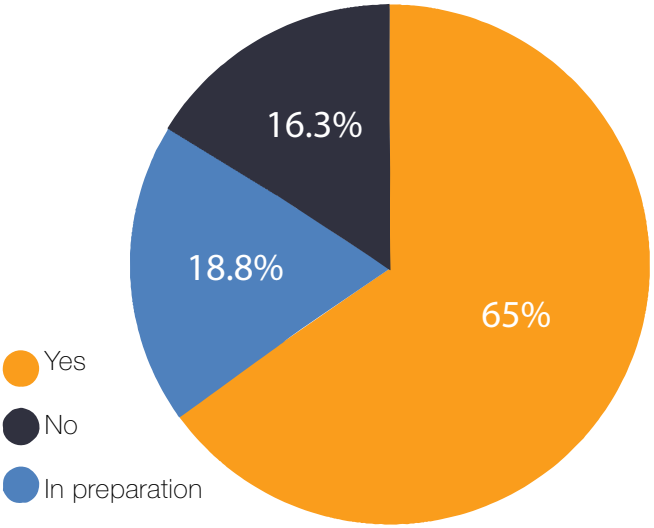

Figure 3. Existence of an institutional harvestable public facing metadata catalogue of datasets.

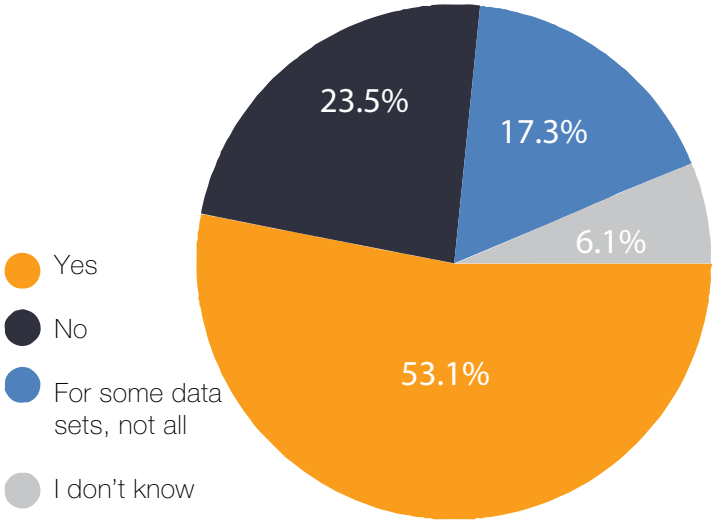

Figure 4. Do you supply a metadata service based on a standard protocol that can be used to harvest the catalogue content?

Question 11 asks which metadata protocols the members support. The most popular harvestable service protocols were OAI-PMH (17) and OGC-CSW (10). Some respondents listed either metadata formats (like ISO19115) or data service endpoints (like WFS), or services that allow users to query but not harvest metadata (like IVOA's TAP - Table Access Protocol) rather than harvestable metadata service protocols. It is possible that either members are not providing a harvestable service and chose to report on how they are exposing their metadata, even if it is not a harvestable service, or are unaware of how the metadata is harvested.

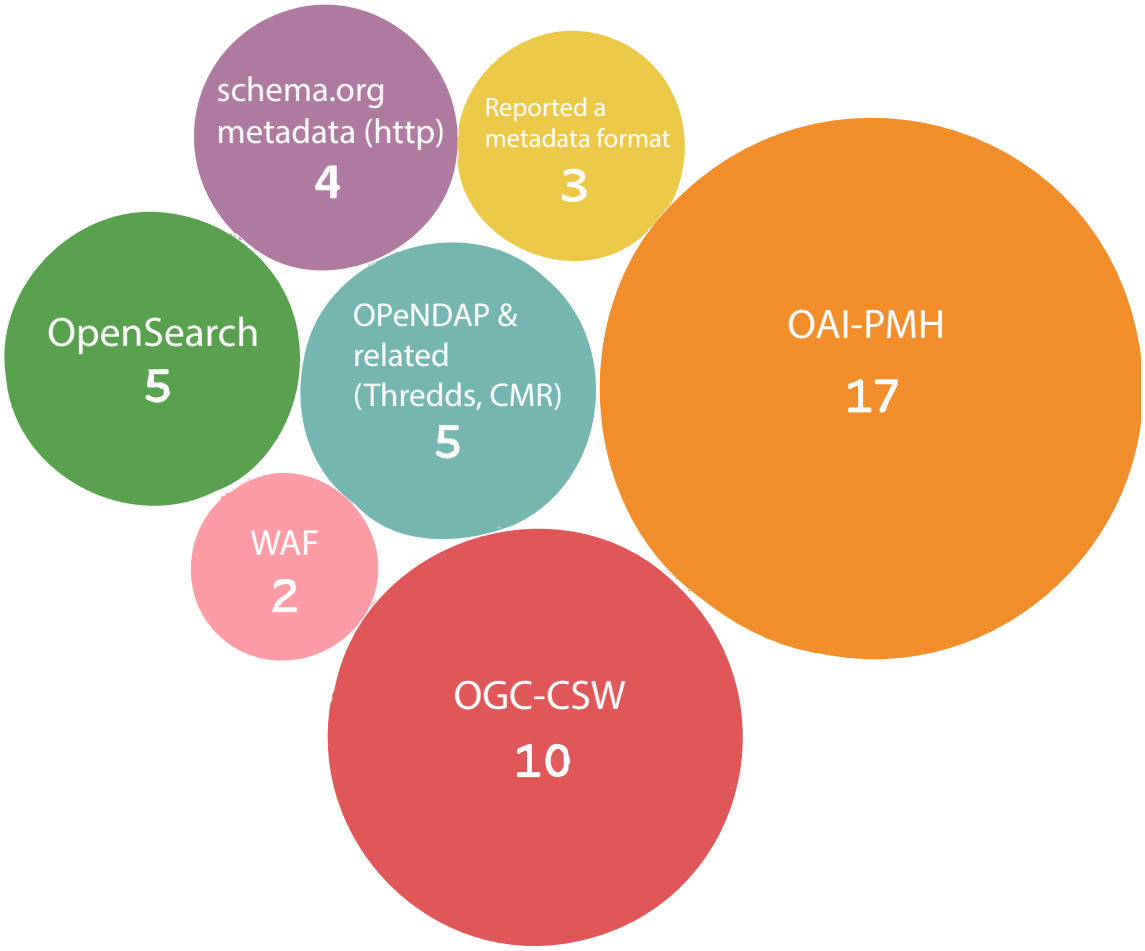

Figure 5. Which specific metadata service protocol or protocols do you provide?

We asked members if they wanted free technical support to create a harvestable metadata service. Our survey shows that overall interest in this topic is high for members across states of catalogue development. This interest is especially high among those who are currently developing or who have no active plans to develop a harvestable metadata catalogue (Figure 6).

Some of the respondents who reported already having a harvestable metadata catalogue at their institutions also expressed some interest in ITO's support (Figure 7, C). Since they already have a service, it is not clear what type of support they may have in mind; perhaps either functional enhancement or potentially migration services. ITO will consider looking further into this question by engaging with respondents directly.

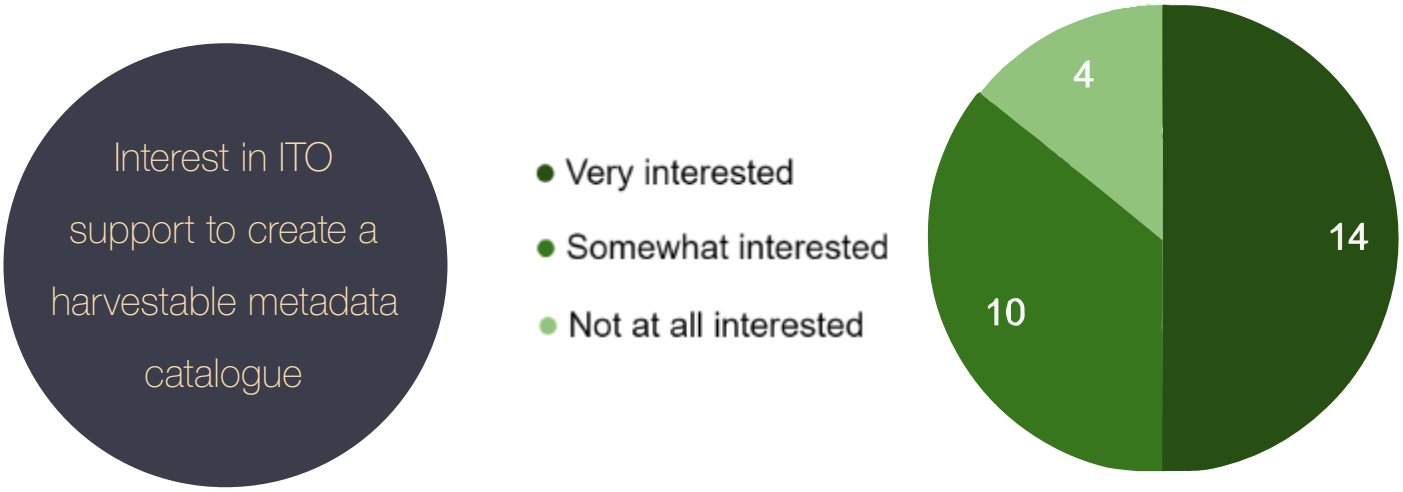

Figure 6. Total interest in ITO support to create a harvestable metadata catalogue.

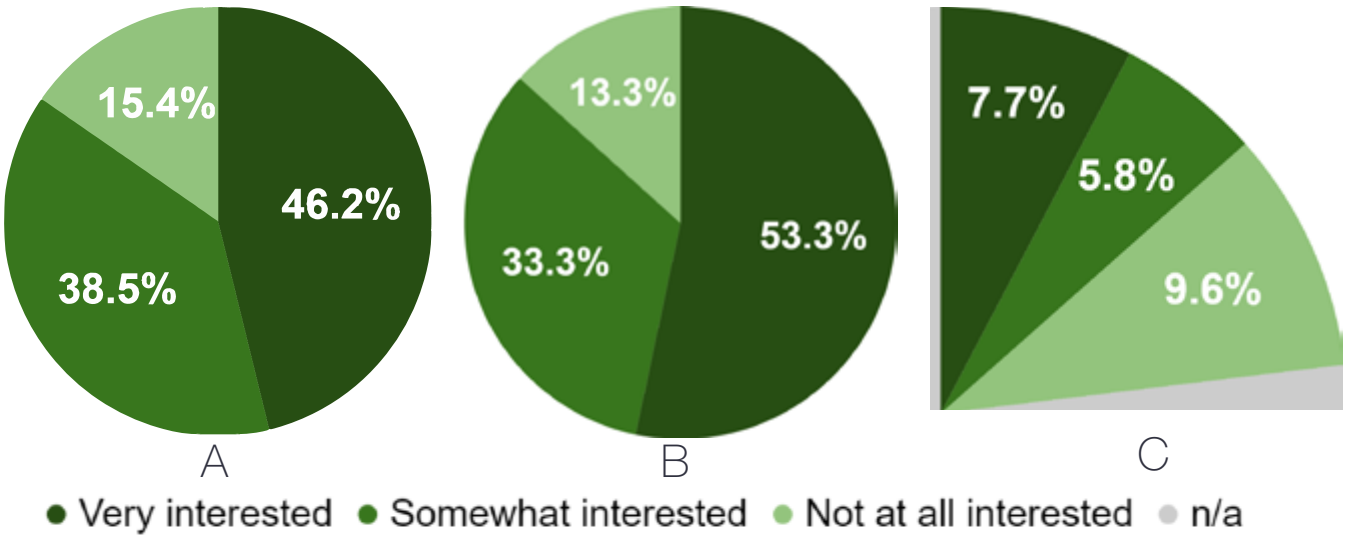

Figure 7. Disaggregated level of interest for respondents in each group. A: No repository at the moment, B: In preparation, C: Existing repository.

These results may be uncovering two minority membership segments:

1. A subset of repositories with advanced RDM capabilities, with experienced practitioners interested in sharing advanced knowledge with the WDS community.
2. A small minority of members without plans for a catalogue or any apparent interest in WDS-ITO support.

It might be worth engaging with the second group of users in the near future, e.g. by reaching out to them for qualitative survey interview.

Only four respondents had no existing or developing harvestable metadata catalogue and expressed no interest in ITO support to develop it. Of the 28 (35%) respondents who did not already have a catalogue (that is, either had none or had one in development), a large majority (24, or 85.7%) reported being either 'very interested' (14, or 50%) or 'somewhat interested' (10, or 35.7%) in ITO's support to develop a harvestable metadata service (Figure 7, A). Consequently, ITO has sought to channel this interest into the creation of a new WDS Working Group on Harvestable Metadata Services.

Unique, persistent identifiers (PID's) for repository datasets

We asked if members generated unique, persistent identifiers (PIDs) for their data holdings. Out of 77 respondents, 42 reported they manage PIDs with their datasets, 19 for some, but not all of their holdings, and 16 said they do not provide PIDs (Figure 8). Of 38 respondents to the follow-up question, asking if they planned to support for PIDs in the future, 18 responded that they were very interested in providing PIDs in the future, 11 were somewhat interested and 9 were not at all interested (Figure 9).

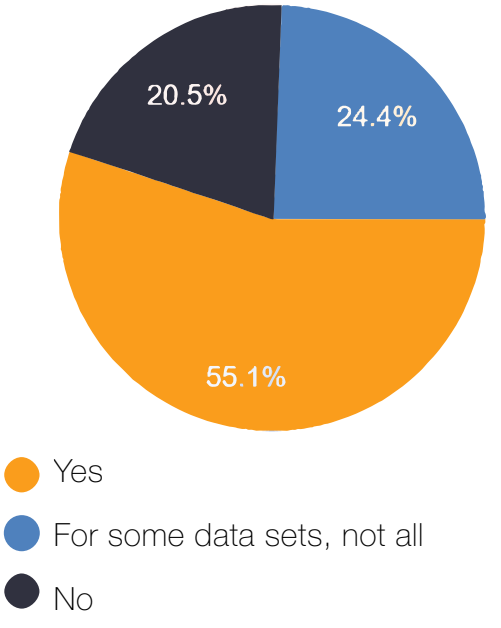

Figure 8. Do you generate unique, persistent identifiers (for example DOIs, UUID, PURL) for your data holdings?

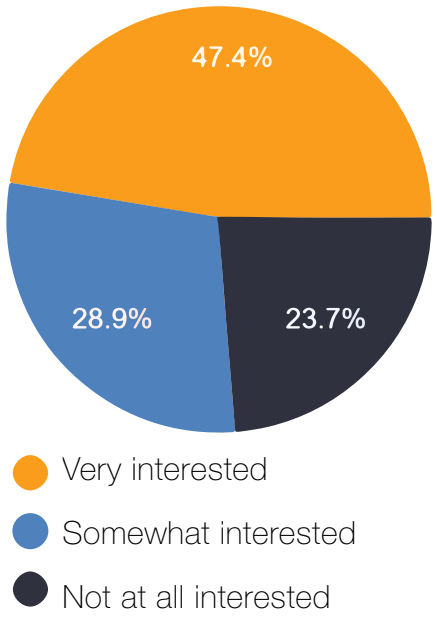

Figure 9. If you do not currently generate unique, persistent identifiers (for example DOIs, UUID, PURL) how interested are you in doing so in the future?

Linked open data/ semantic data

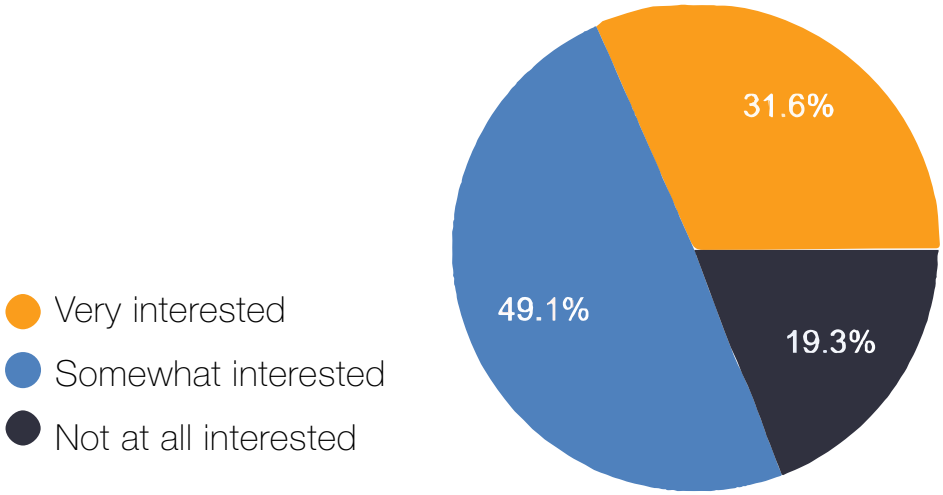

Figure 10. If you do not currently serve data semantically, how interested are you in doing so in the future?

Question 14 asks if the data managers serve Linked Open Data or semantic data. 32 of 76 respondents currently serve linked open data, while 44 do not. We may be responsible for some confusion in the survey in failing to differentiate between activities associated with data development and services, and activities associated with metadata development and services. WDS members are clearly interested in developing semantic markup for metadata (as reported below), but redeploying entire datasets semantically may feel like too large of a project to take on.

Repository Management Software

There does not appear to be any community of practice around specific platforms for serving data and metadata within the WDS. In the 77 responses to a question on this topic (Figure 11: "How does your organization currently serve its metadata catalogue?"), 43 reported using a custom website to serve metadata without giving a specific commercial, off-the-shelf underlying repository management software product (although a few libraries like pycsw and frameworks like Tornado were mentioned by name). 4 respondents had no metadata catalogue.

An additional 12 respondents listed the name of the website where metadata and data are exposed without specifying any underlying technology, implying but not directly stating that these are custom sites.

Most of the respondents that did specify a commercial product also mentioned that they were customized solutions. Other products mentioned were GeoNetwork (6), Drupal (4), CKAN (3, 2 of them custom), Dataverse (2, 1 custom), custom Nesstar (2), Joomla (1) and ESRI's Geoportal (1).

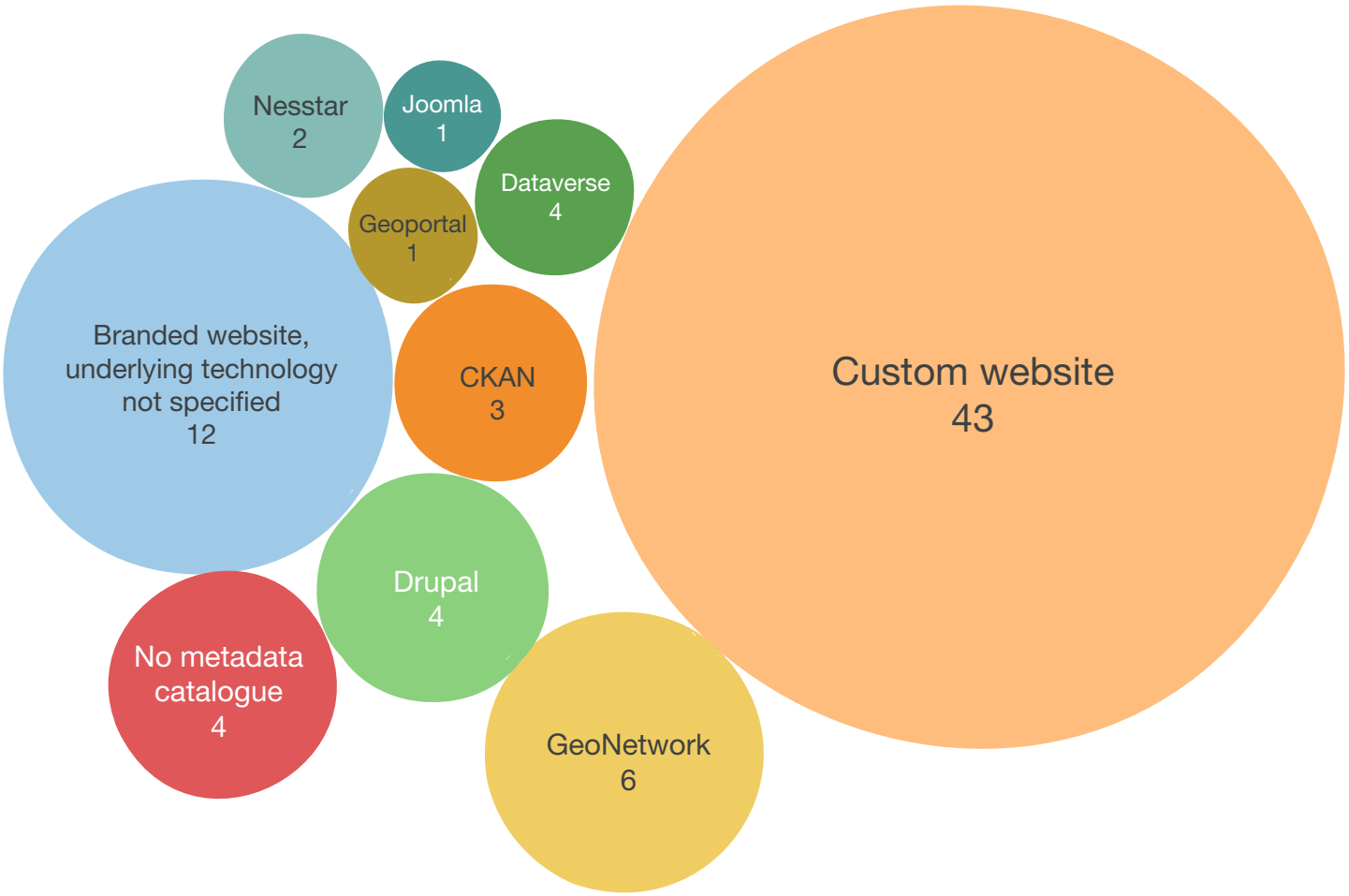

Figure 11. How does your organization currently serve its metadata catalogue?

Similarly, 35 of 68 respondents said they used a custom tool to write metadata, without specifying an underlying product or community (Figure 12). Twelve respondents said they didn't use tools, either because they depended on their data providers to write metadata, they wrote metadata manually with a text editor, or they didn't write metadata at all, and 4 didn't specify any tools, and instead referred to standards or formats like Darwin Core.

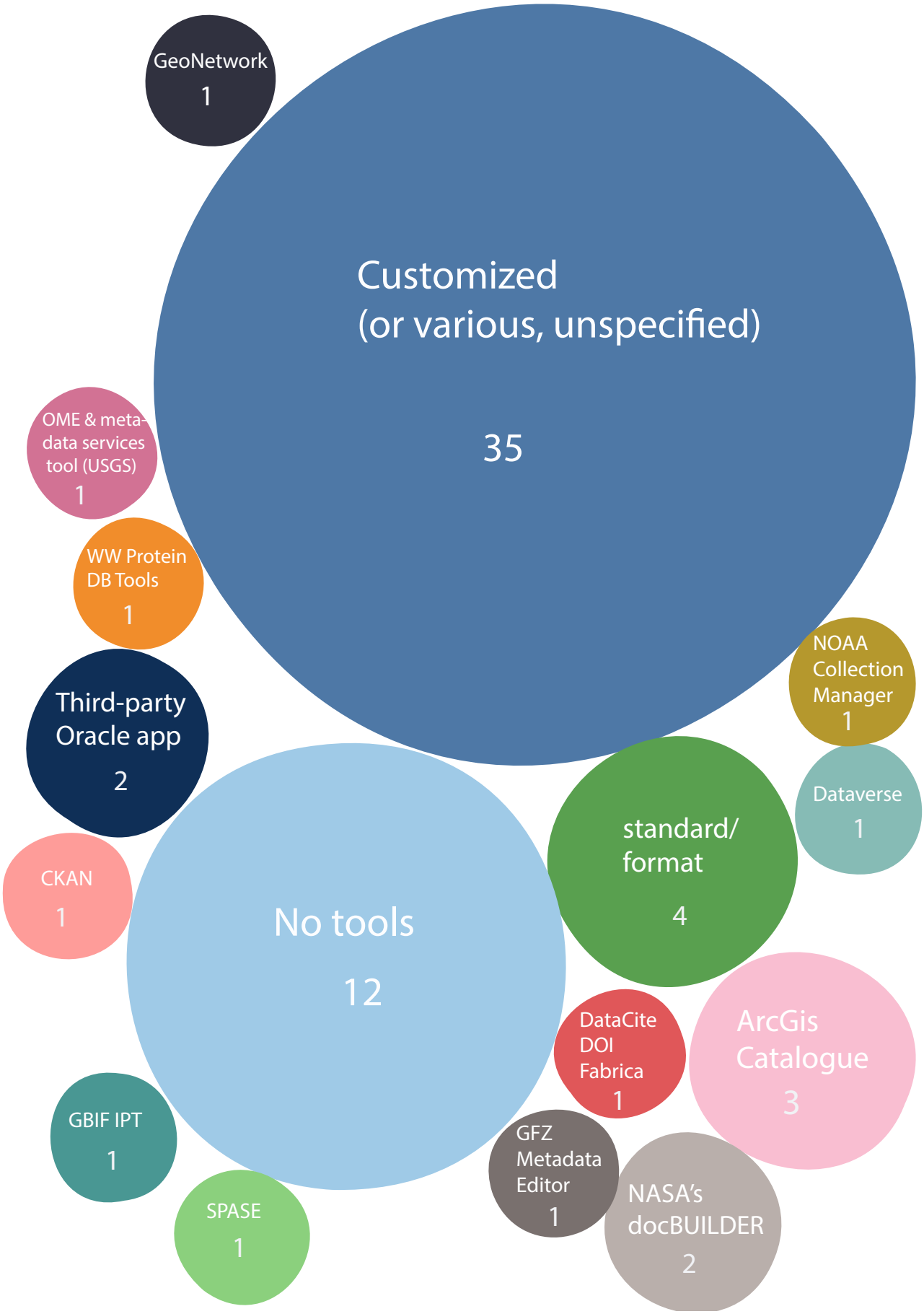

Figure 12. What tools or platforms are you using to write metadata?

The tools mentioned for writing metadata included ESRI's ArcGIS Catalogue (3), DataCite DOI Fabrica built-in tools (1), CKAN (1), Dataverse (1), GeoNetwork (1). 2 reported using an unspecified 3rd party Oracle application. The remainder were largely domain specific tools: NASA's docBUILDER (2), Online Metadata Editor (OME) from the US Geological Survey (1), the Space Physics Archive Search and Extract (SPASE) metadata tools (1), the Global Biodiversity Information Facility Integrated Publishing Toolkit (1), the GFZ Metadata Editor from the German Space Agency (1), NOAA Collection Manager (1), and tools from the World Wide Protein Data Bank (1).

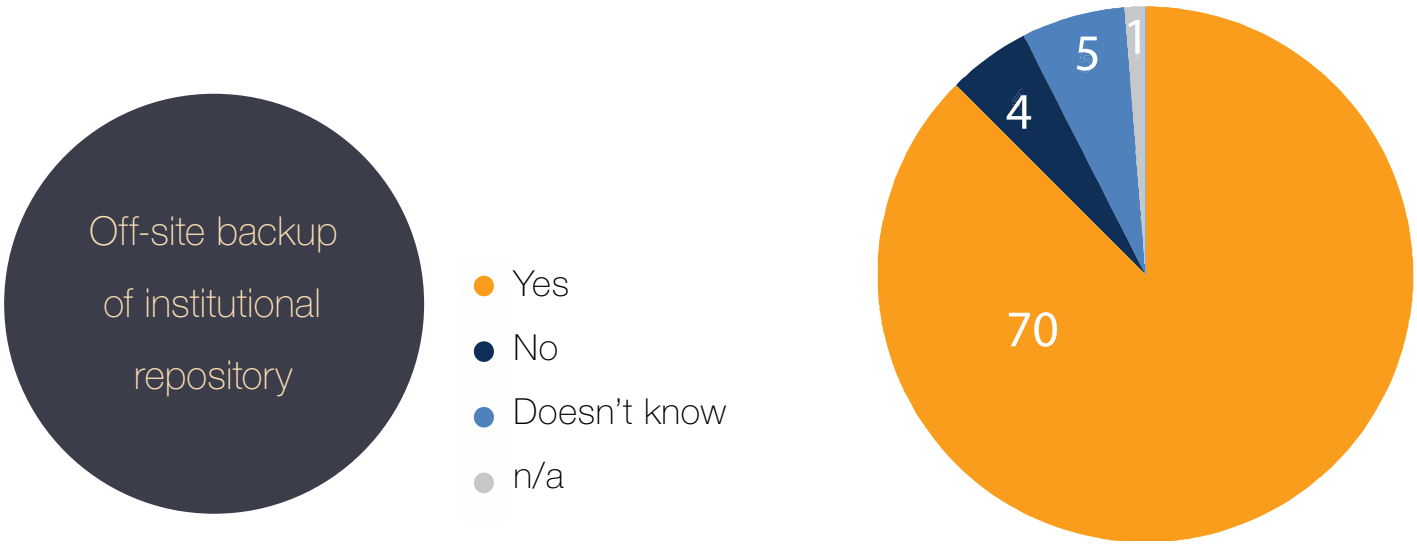

Figure 13. Do you have an offsite backup of your repository?

We asked managers if they had an off-site back up of their repository. Of 79 total respondents, 71 said they had an off-site backup of their repository, while 5 did not know if they did and 4 repositories reported not having an off-site back up at all.

Relationship to other organizations

re3Data, DataCite and Google Dataset Search

Figure 14 shows responses to questions about repositories' presence on re3Data.org, and whether datasets are being harvested by DataCite or Google Dataset Search, respectively. Of 79 respondents, 55 reported having a re3data profile, 37 said at least some of their datasets were harvested by DataCite, and 36 reported the same for Google Dataset.

Slightly more repositories reported being indexed or harvested by DataCite than Google Dataset Search, even though slightly more respondents appeared to be aware of Google Dataset search services. This may be a function of the fact that DataCite works with traditional harvestable metadata services, while getting indexed by Google Dataset Search is a much discussed topic amongst repository managers, it requires repositories to create new publication workflows, adding schema.org to their metadata assets.

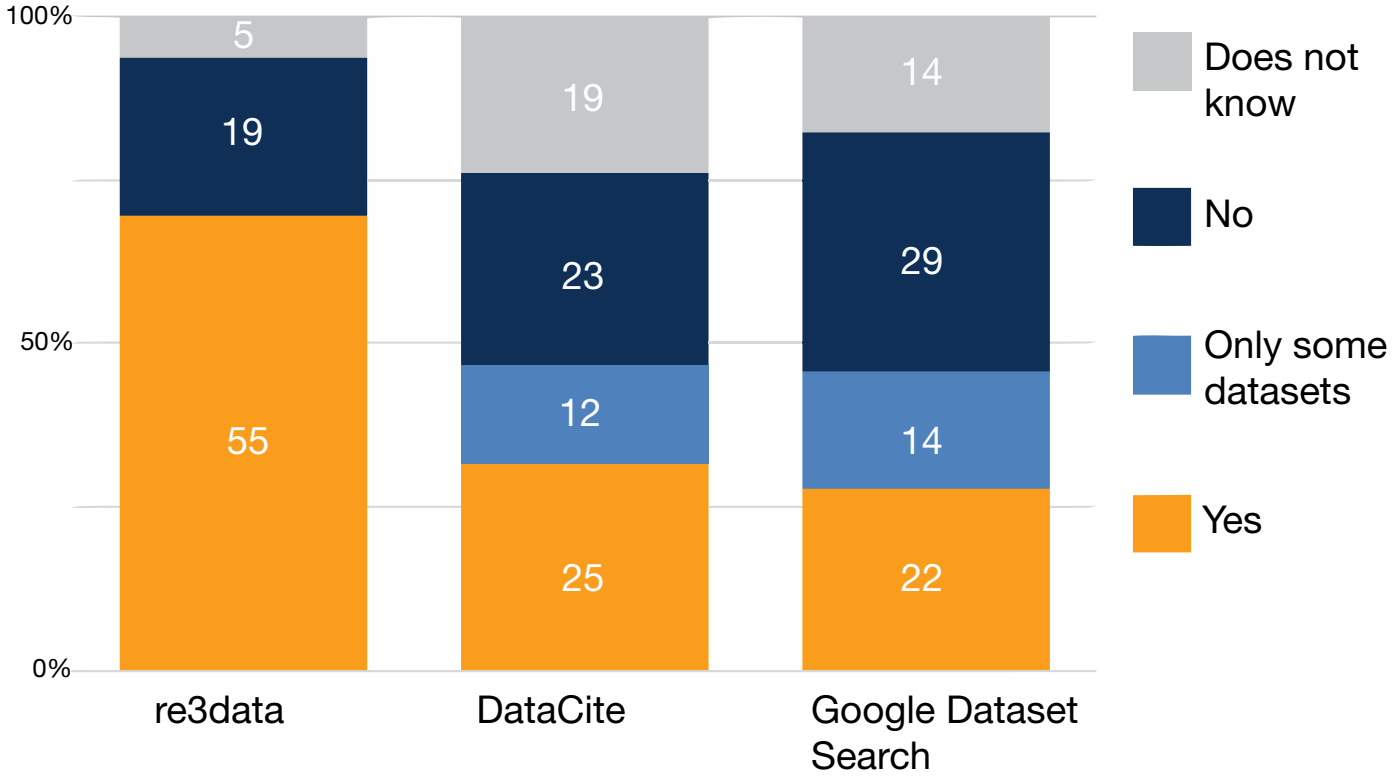

Figure 14. Repositories listed on re3data, harvested by DataCite and indexed by Google Dataset search.

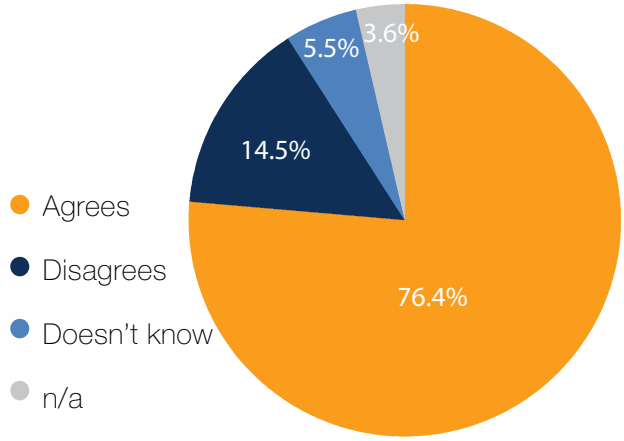

Figure 15. If it is listed, do you agree with the information re3data has presented about your repository?

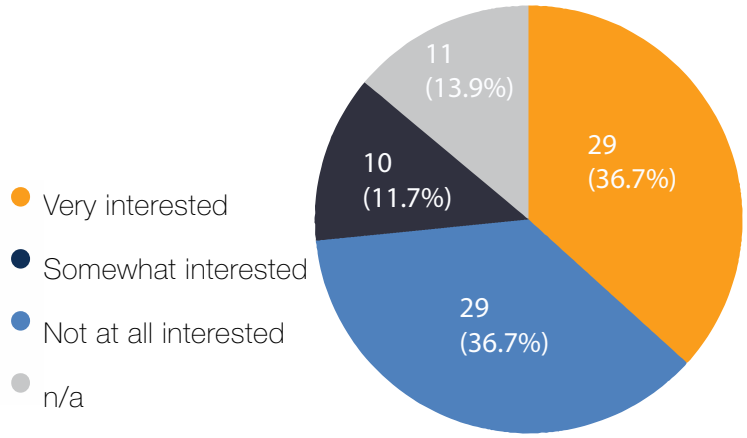

Figure 16. How interested would you be in getting support to help to make sure your data holdings are represented in Google Dataset Search or DataCite?

Most (42, or 76,4%) respondents agreed with how they were represented in their re3data profile (Figure 15), and 8 organizations did not agree with the information presented about their repository. 3 respondents did not know how to answer the question, either because they didn't know how they were represented, or did now know how to judge their re3data profile.

As can be seen in Figure 16, there is a lot of interest in getting WDS repository holdings harvested by Google Dataset Search or DataCite. Almost three quarters of respondents expressed some interest in support to achieve a better representation of their data holdings in these services. As a result we are creating a new work package within the ITO to support data managers interested in having their metadata indexed and searchable in Google Dataset Search.

WDS Members in RDA Groups

46 respondents gave information on their institutions' participation in RDA working or interest groups, including 4 respondents who reported serving or having served as co-chairs on the RDA Advisory committee. Collectively, WDS members have been active in 40 unique interest and working groups (some already completed). These groups are shown in the call out box on page 20, grouped by the number of WDS members in a working or interest group (between 1 and 6 members).

WDS members have been active in 40 unique RDA interest and working groups

(shown grouped by number of WDS members in each group)

6 

- PID IG
- Publishing

4 

- Certification of Digital Repositories IG
- RDA Data at Risk IG/Data Rescue/Long Tail
- RDA Technical Advisory Board or chair

2 

- Marine Harmonization
- RDA national groups
- RDA/CODATA Legal Interoperability IG/Legal issues
- Data Versioning WG
- Data Discovery Paradigms IG
- Social Science Interest Group
- Data Usage Metrics WG
- Domain Repositories IG
- WG Data Fitness for Use
- Disciplinary Collaboration Frameworks

1 

- Software Source Code IG
- Data Granularity
- Harmonizing FAIR descriptions of observational data
- Research Data Repository Interoperability WG
- Active DMPs
- DMP Common Standards WG
- Exposing DMP WG
- Metadata IG
- Brokering Framework
- Federated Identity Management
- Sharing Rewards and Credit (SHARC) IG
- Provenance Patterns
- InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT WG)
- Repository Platforms for Research Data I
- Early Career and Engagement
- IG Physical Samples
- Chemistry
- Vocabulary Services IG
- Biodiversity
- Data Fabric
- Data Security and Trust WG
- IG Earth and Environmental Sciences
- Maturity Models
- Structural Biology IG

Although an exhaustive categorization in mutually exclusive, non-overlapping[3] groups is likely not possible for this set of responses, Figure 17 below shows a rough attempt to group some of these activities into common themes. This rough view shows two main foci for RDA activity, namely a focus on platform development and interoperable services, and another on work within domain groups, where researchers from associated fields collaborate, acquainted with each other's data management needs, to develop activities, practices and standards to meet those needs.

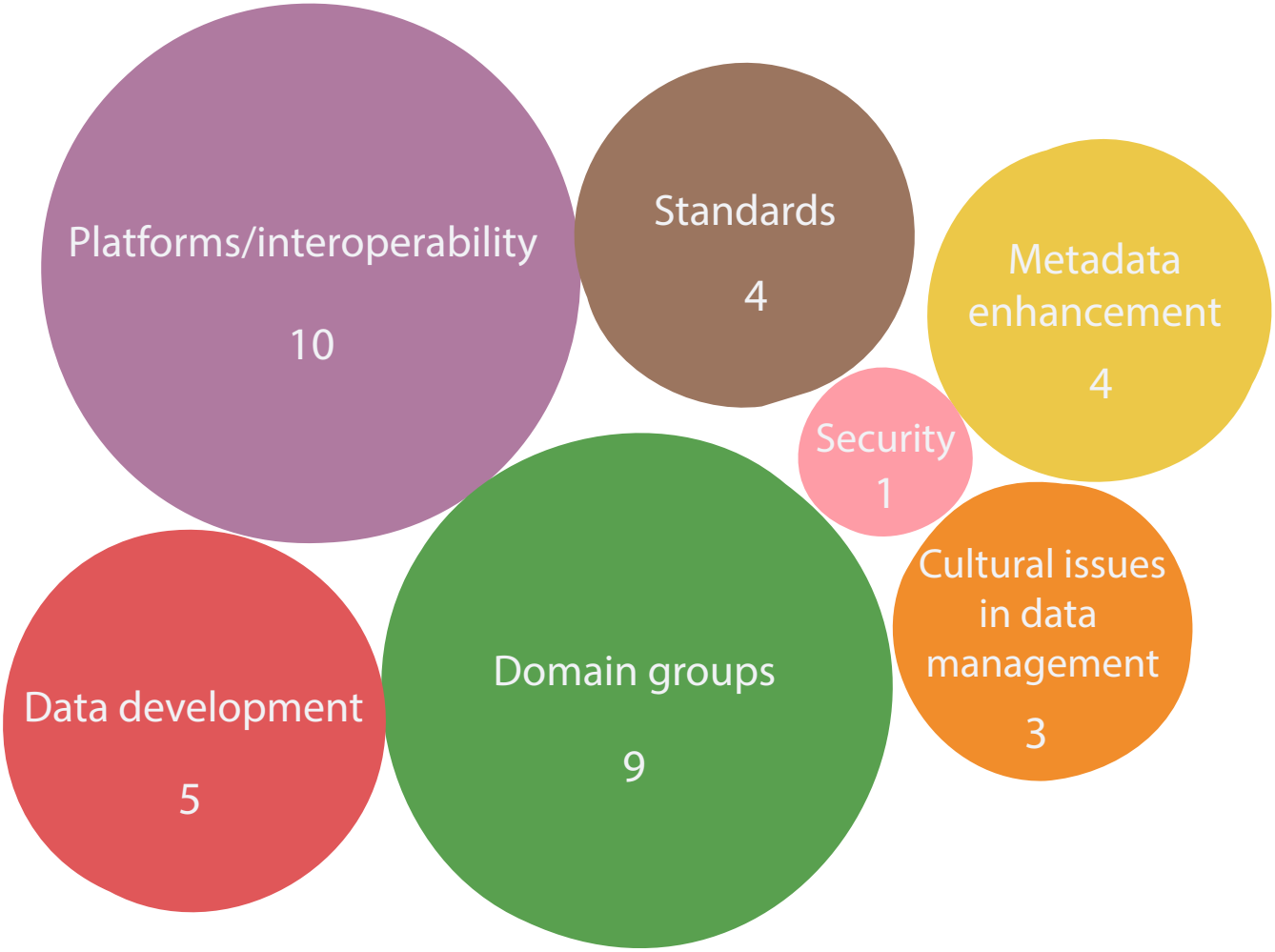

Figure 17. Topics of Working Group/Interest Group focus for WDS members.

Interest in WDS initiatives

We asked members several questions to gauge their level of interest in various WDS initiatives, specifically a brokering registry, a WDS catalogue, WDS collective data statistics and a knowledge network.

'GEOSS Data-CORE' tags in WDS metadata

We asked members about their interest in using a brokering registry, designed to help users find and re-use data brokering tools. Members were also asked about their interest in having all WDS member repositories tag metadata as 'GEOSS Data-CORE', in order for WDS datasets to be grouped and shared within the GEOSS Data Collection of Open Resources for Everyone, "a distributed pool of documented datasets with full and open unrestricted access at no more than the cost of reproduction and distribution".[4]

As shown on Figure 19 (opposite page), 19 of 78 respondents said they were 'very interested' in utilizing a brokering registry, 45 said they were 'somewhat interested' and 14 said they were 'not at all interested'. 18.4% of respondents expressed no interest in 'GEOSS Data-CORE' tagging (Figure 18), while 14.5% said they were very interested. Over half of respondents stating they were 'maybe' or 'somewhat interested' in both initiatives (57.7% and 53.9%, respectively), the implication may be that WDS members aren't yet sufficiently familiar with these developments. In fact, one respondent used the open text field to disclose that they did not understand what exactly was involved in tagging metadata as 'GEOSS Data-Core'.

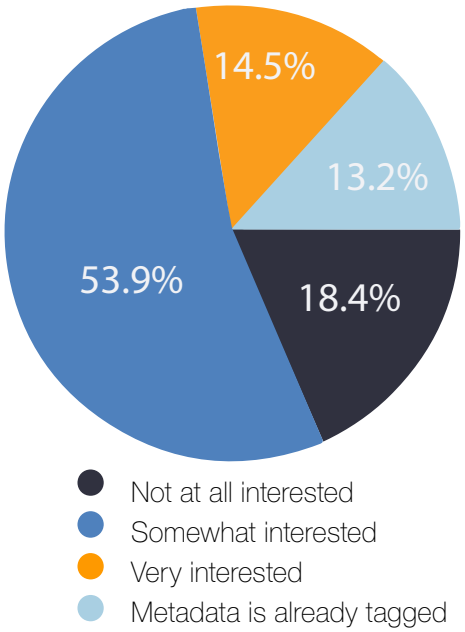

Figure 18. Is your Organization interested in having all WDS members tag their metadata as 'GEOSS Data-CORE'?

A brokering registry is a list of tools to help manage or move data and metadata between systems—such as converting data between formats or finding equivalent semantic vocabulary terms—created for ease of finding and re-use of these tools.

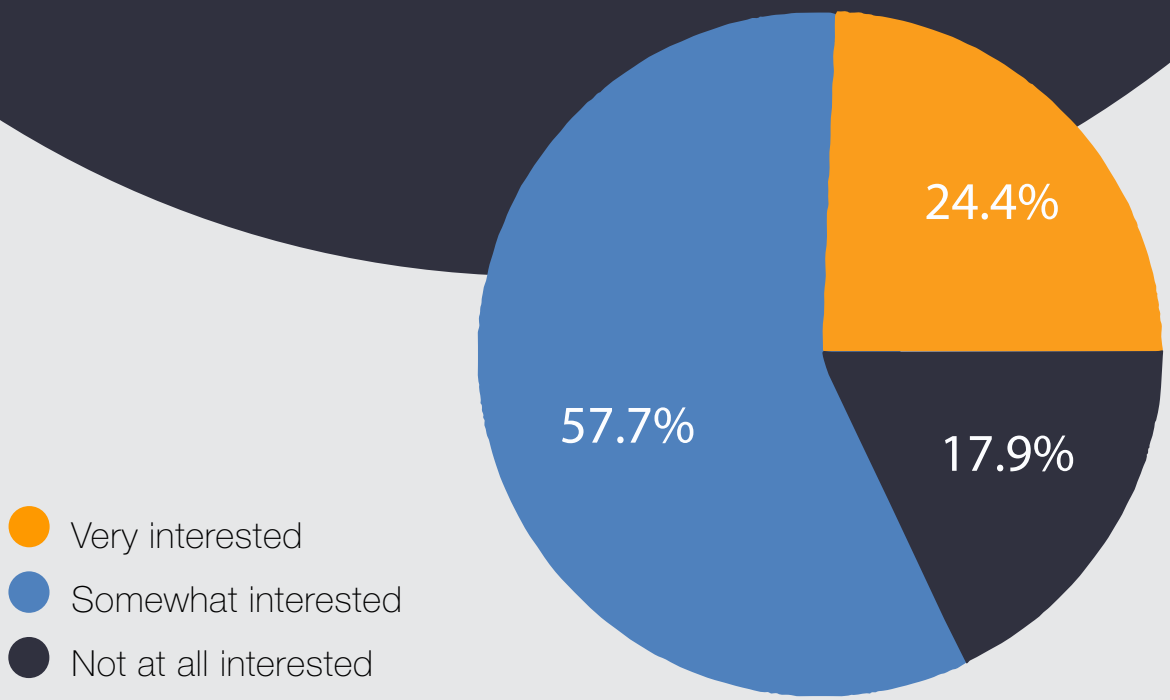

Figure 19. How interested are you in accessing or utilizing a brokering registry? We are not asking you to be responsible for maintaining a registry, we are curious if it be helpful to you to be able to search a registry in order to find and re-use existing tools.

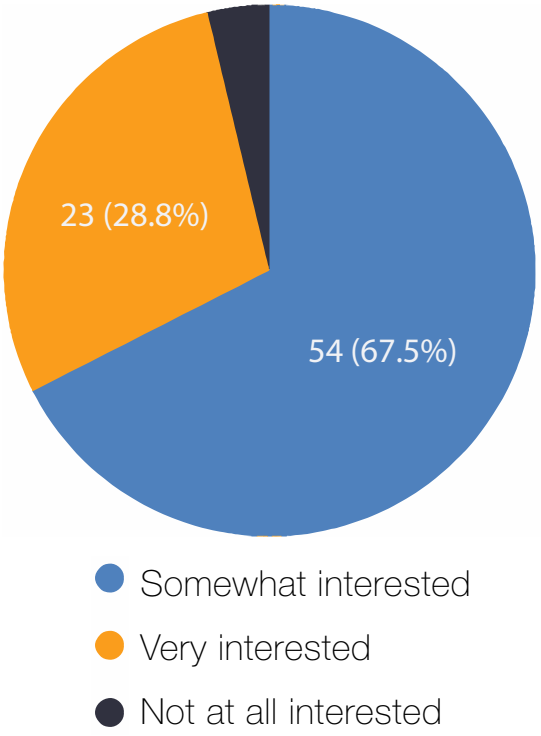

Figure 20. Interest in a collective WDS Metadata Catalogue.

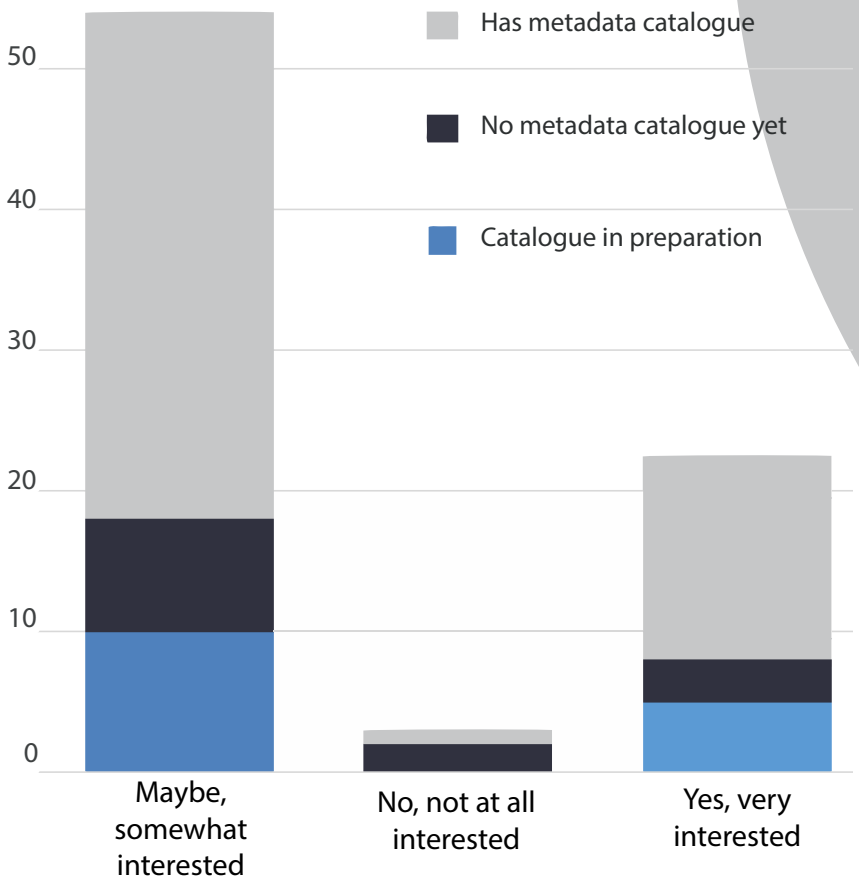

Figure 21. Interest in collective WDS Metadata Catalogue (horizontal axis) by current state of institutional repository of respondent.

Interest in a collective WDS metadata catalogue

18 members indicated that they are not interested in participating in a WDS catalogue, while 54 are somewhat interested. Members stated the fact that a WDS catalogue would not add value beyond existing platforms like Datacite, the European Open Science Cloud and Google Dataset Search. Moreover, they felt it would be confusing for end users, it would require more staff time and resources and it would not be robust enough to meet user expectations. This quote is a good summary of the concerns:

> "There are already a variety of options [...] available for this function. They seem to come and go with funding cycles. I think it would be more strategic to help members register with currently available catalogs rather than inventing another option."

Figure 22. How interested would you be in sharing aggregated, summary statistics about your holdings, contributors and consumers with the ITO with the intent to create reports on the magnitude and reach of the WDS system?

On question 22, which asked members about their interest in sharing aggregated, summary statistics with ITO, 88.7% of members reported being at least 'somewhat' or 'very' interested in sharing statistics with WDS (Figure 22).

Figure 23 compares the percentage of responses obtained for 4 questions: asking for a technical point of contact at member organization, contacts provided for participation in a WDS Knowledge Network initiative, the preferred platform to be contacted by WDS-ITO, and gauging interest in participation in a Technical Advisory Committee).

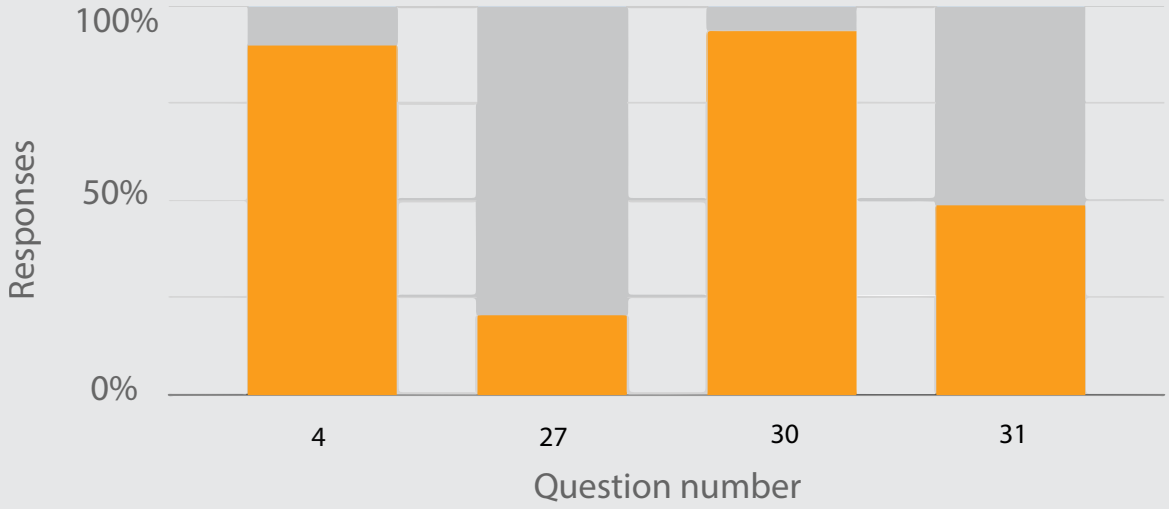

Figure 23. Survey participants (percentage of n = 77) responding to questions gauging interest in ITO initiatives or asking for contact information.

Interest in a WDS Knowledge Network Working Group

A **Knowledge Network** is a web-based, interlinked repository of relationships between the actors and entities that make up our research landscape: people, institutions, projects, research disciplines and topics, funding sources, and the like.

Over 3/4 of respondents expressed being either 'somewhat' or 'very interested' in a potential WDS working group to further develop the consortium's Knowledge Network. The respondents supplied contact information for a total of 38 names of potential participants, from 15 different organizations.

Respondent's interest in serving on the ITO Technical Advisory Committee was reflected in a list of 40 names of potential participants, from 37 different organizations, provided by survey participants.

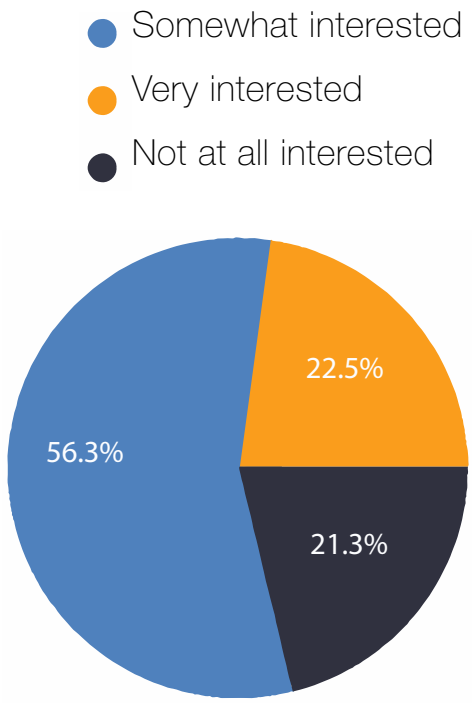

Figure 24. Interest in a WDS Knowledge Network WG.

Interest in a WDS Technical Advisory Committee

The ITO is currently creating a **Technical Advisory Committee** drawn largely from the WDS membership. The primary purpose of the TAC is to advise the ITO on infrastructure strategies and technical road maps. We are creating a list of WDS members who may at some point in the future be interested in serving on the TAC.

If you could have free support from the new International Technology Office for any infrastructure project of your choice, what would you like to do?

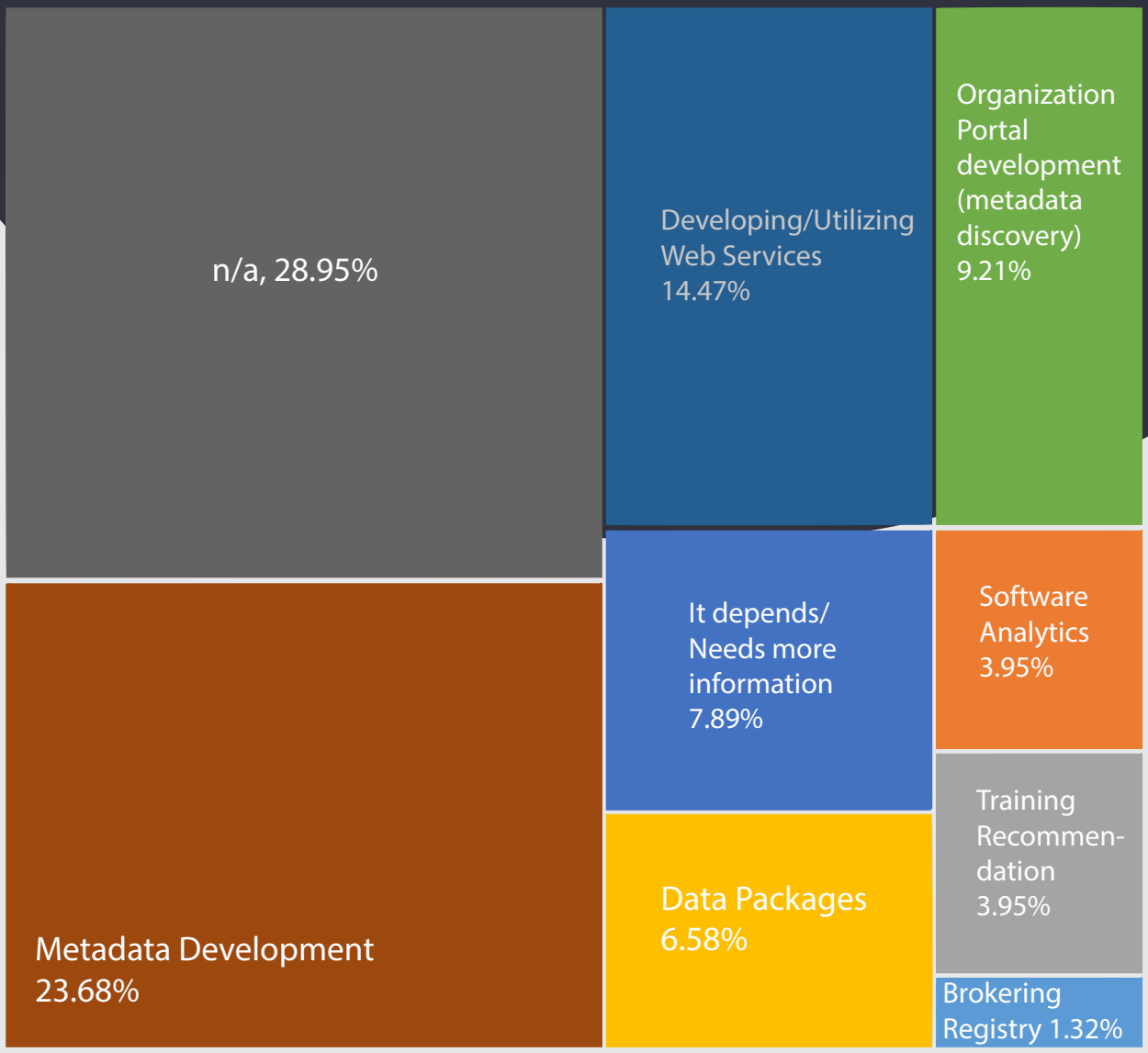

Figure 25. Projects of interest among WDS members, if they could have free support. Each square represents the percentage of responses in each category. All responses of "Needs are unclear to respondent, No need, Unclear or nonspecific response, n/a or cannot give answer" were collectively grouped into "n/a."

A member-support role for WDS-ITO

Through an open-ended question, the survey sought to give WDS members an opportunity to highlight any technical issues the ITO was not aware of. 76 survey participants responded to this question, and their responses grouped into 8 broad areas (Figure 25), with an additional category that grouped surveys with missing, insufficient or invalid data (29% of responses).

Of those, 6 organizations indicated they could not talk about support without knowing the capacity of the ITO. These organizations may already have a lot of technical expertise on staff, and/or because the ITO does not yet have a track record of work with the community, they did not want to speculate on potential projects.

Of the named potential technical projects, the largest group of responses was associated with metadata development. Among them, most referenced adding semantic markup to metadata, either manually or via some AI automation process, and 8 responses mentioned either ontologies, controlled vocabularies, structured metadata, semantics, structures for automated metadata harvesting, linked open data markup or schema.org explicitly. Six responses referenced developing metadata to make it more standardized, findable, visible, modern, interoperable or to increase it's exposure to other portals, all of which likely includes adding semantic markup, but also could include changing metadata formats, adding unique identifiers or adding new endpoints to metadata services.

Three responses explicitly mentioned adding persistent unique identifiers to metadata, and one response noted a need to handle metadata for scanned microfilms (an issue within the data rescue or non-standard data management community). One further response noted a need for "compliance checkers," which we speculate could be a reference to enhancing either a metadata tool or data tool so that it has the ability to use defined domains, sensible data ranges, and checks for completeness, but that would need to be clarified by the respondent.

Some organizations said they would appreciate support for the development of the portal they use to serve data and metadata, in order to enhance distribution and awareness of their content (11); these responses included the need for building portals/platforms, new viewers, new servers, standards for infrastructure, and a metadata management system. Most of these members also said they were "very interested" in participating in a collective WDS metadata catalogue, and all of them were either "very" or "somewhat interested" in a collective catalogue.

WDS members also expressed a need for help developing and/or utilizing web services (10). 3 of these explicitly named metadata services, but also data services, services associated with the knowledge graph such as linking publications with related datasets, generic "cloud" services, and monitoring functions (presumably NAGIOS type software that pings and checks on the availability of their own services but this would need to be clarified). WDS sees a need for training in the areas of:

- New technologies and their adoption issues
- Standards, especially for data and metadata quality
- Licenses, access policies and legal issues

3 of the responses reported in the section on the need for metadata development also noted a need to develop the full dataset that the metadata describes, including full semantic markup and support for associated formats like JsonLD. An additional response mentioned "Frictionless data packages" while another response noted the need for support for establishing data quality evaluation methods; this may be similar to the request for "compliance checkers" reported above.

Projects associated with data analytics or interoperability were raised with (3) responses associated with support for software to add value to the data, including visualization and (1) response that requested the development of a brokering and mediation registry. There were approximately 20 additional comments that are worth investigating in the context of how the ITO can be of service to the WDS community. We have broadly classified them under two headings: Technical projects, and Strategic focus areas.

Technical projects

- Can WDS-ITO support linking the Global Change Research Data Publishing & Repository (China) metadata to Google?

- I think the most important thing is to build an open source data portal to be reused by many data centres.

- We do already serve semantic data in the headers of our landing pages. However, we would like to find better ontologies, standards etc.

- Although we are not harvested, we do regularly upload [our metadata].

- We are 'very interested' in Google Dataset, even though we are already harvested. So this answer does not refer to technical issues but to the following: We would like to have influence on the way the harvested files are presented, and tagged with keywords. The way Google DS does this is not always optimal.

- Create a landing page that links to ISO 19115 profile definitions used by member organizations (or others), in order to increase the possibility for shared interpretation and better interoperability.

- Metadata catalogue question: PSMSL does not have its own metadata catalogue, as we have one dataset, but contribute our discovery metadata to the European Directory of Marine Environmental Datasets (EDMED), under the auspices of the EU SeaDataNet projects (edmed.seadatanet.org).

- Metadata service protocol: I think there may be a standard protocol - can be investigated and information provided later on.

- Persistent identifiers question: A DOI planned for PSMSL dataset in the very near future.

Strategic focus areas

- Could you define good/best practices within the WDS members (based on the CTS assessments)? It would be great for us to know how we are performing with respect to the other members (e.g. using the compliance levels).

- Provide a collective avenue of input from WDS members to entities like DataCite and ORCID to influence governance and improvements. Organize webinars for WDS members to share solutions and implementations.

- It will be great if ITO could provide some technical documents about data and metadata such as tools, protocols, and so on.

- Help us to foster/enforce data citation in the scientific community and data management plans, including corresponding budgets at the level of science foundations.

- Maybe ITO can fund some technical projects which are completed by the data centers of the same discipline.

- This is a huge area, how does/will the ITO keep abreast of similar initiatives and developments

- The questions in the survey are quite surprising. I would have anticipated the focus of activity to be more closely related to repository data management rather than broader data integration. I think of the former as the area of specialization of this organization while the latter competes with a host of existing efforts. I am very puzzled that this new effort is not focusing on strengthening the certification process, embracing new technologies, and developing creative ways to improve repository sustainability,

- We would welcome an ITO representative to deliver a presentation at our Governing Board meeting in the future.

ITO Action Items

Generally, WDS sees a need for training in the areas of new technologies and their adoption issues and standards (especially for data and metadata quality). When possible, we align our work product with the FAIR principles and support CoreTrust-Seal guidelines for trusted repositories.

Item #1: Harvestable metadata and semantic markup

Given the communities vested interest in metadata development, the ITO has focused on 2 action arising from this survey, the first of which is already underway:

A WDS Working Group for harvestable metadata services (HMetS)

While, strictly speaking, data served by members does not have to be harvestable in order to be findable (a person can search a catalogue that is not a harvestable catalogue), harvestable metadata does extend data accessibility and reuse, as other portals can aggregate available datasets. Harvestable metadata is addressed in FAIR Principle A1.1: "the [retrievable (meta)data service] protocol is open, free, and universally implementable."[5,6] Similarly, CoreTrustSeal requirement R13 states that a trustworthy data repository "enables users to discover the data and refer to them in a persistent way through proper citation."[7] Acceptable evidence of R13 includes a "repository [that] facilitate[s] machine harvesting of the metadata," though it is not a requirement; metadata services can provide retrievable (meta)data without being harvestable in bulk.

The WDS HMetS-WG Charter has been approved and, to date, there are 10 representatives from WDS member institutions interested in participating. Two parallel information sessions, one on the European/Atlantic time zone, and another on the Asia/Pacific region were held in November 2019.

Supporting semantic enrichment of WDS member repositories' metadata

The WDS HMetS-WG Charter (referred to above) supports one type of data syndication, specifically, harvestable metadata services. Adding schema.org to metadata is an alternative syndication pathway that leverages existing web technologies developed to take advantage of search engine indexing and discovery. In response, the ITO has scoped and joined the relevant RDA working groups associated with these tasks. Specifically relevant for the repository managers interested in semantically enriched metadata is the RDA Research Metadata Schemas Working Group.

FAIR principle I2 requires that (meta) data use vocabularies that follow FAIR principles.[5] While, strictly speaking, a simple tagging system with a controlled vocabulary would satisfy I2 without the need for semantic structures, more mature repositories will have machine readable (automatically resolvable) markup for their metadata. CoreTrustSeal requirements do not reference controlled vocabularies directly, but they are within the scope of R11 as part of an "automated assessment of metadata adherence to relevant schema" for applicants to describe their data curation expertise. They are also relevant to R14, in the applicant's description of how their repository "ensure[s] understandability of the data[.]"[8].

In response to this interest by the WDS membership we have secured funds to extend the outputs from the relevant RDA working groups and more fully investigate the appropriate use of schema.org in scientific data management.

Item #2: Reusable brokering components

While the ITO sees metadata development as a first important need of the membership, we are looking at more support for the knowledge network and brokering in the future. The large amount of customization reflected in this survey would indicate that members have already developed a rich set of brokering components to support their workflows. However, it remains unclear if these components could be reusable in other custom sites that do not share the same underlying technology. Hence, if ITO develops a registry of brokering components in the future, we can identify, at a minimum, 2 activities to move the question of reusable brokering components forward:

1. Reach back to all members who specified custom websites or metadata tools and identify any common underlying technologies.
2. Reach out to the wider RDM community outside of WDS (likely through RDA) to determine up and coming trends.

If there are common threads in tool and platform development among WDS members, these could be harnessed into a fruitful community of practice. For sites like Drupal, CKAN, GeoNetwork and ESRI, there already exists a large group of engaged users and developers that are likely being tapped by WDS members who are customizing these platforms.

In preparation for the development of this work package in the future we are engaging in relevant RDA and federal working groups to learn more about common practices in the global knowledge network and existing brokering services. It is likely that this work will also extend to support for PID services.

Item #3: Customer Relationship Management with Insightly

The members indicated a preference for e-mail as the primary form of communication. As a result, the ITO has stood up an instance of Insightly, a customer relationship management platform used to track and reach out to members. The ITO has begun training on how to use the platform and will investigate the functionality associated with email blasts and newsletters.

Item #4: Future surveys

For future WDS-ITO surveys:

- In response to participants' feedback, some questions will be reformulated for clarity. Some members found the questions difficult to answer, especially respondents who represent more than one repository or institution. Associations with multiple members, with differing assets, systems, objectives and constituencies, and those who work through technical commissions or network affiliates may not be a practical partner for technical implementations. These institutions may be better served by qualitative one-on-one interviews to learn more about their strategic plans and road maps.
- In light of their very low response rates, ITO will develop targeted approaches to reach Associate (32%) and Partner members (40%) to gather data that allows us to draw more meaningful insights about their needs and interests.

Notes

1. At the time of the survey: 123 active members (regular, network, associate and partner) and 19 membership candidates.
2. An equivalent PDF was sent to members who could not access Google services.
3. As an example, the Federated Identity Management group can be thought of as both a group that deals with Platform Development and Interoperability, as well as a Security enhancement. Similarly, publishing IGs touch on issues with Interoperability, Metadata enhancement and Cultural Issues surrounding data management. Likewise, the "Publishing" category may be slightly inflated, as it includes responses such as "publishing related topics," RDA Data Publishing IG, Scholix, and Data Citations.
4. https://www.icsu-wds.org/files/geoss-data-core-tagging-metadata-instructions.pdf/view
5. https://www.force11.org/group/fairgroup/fairprinciples
6. https://www.go-fair.org/fair-principles/
7. https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf
8. https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/

Appendix

Survey Questions

1, 2	Email Address, Name of respondent	Contact
3	What is the Name of WDS Member Organization you represent?	Contact
4	What is the name and email address of the best technical point of contact for your Member Organization (if different from above)?	Contact
5	Does your Organization have a harvestable public facing metadata catalogue of its data sets?	Existence, Services
6	If you do not have a harvestable metadata catalogue, how interested would you be in getting support to create one (assume that the support is free of charge)?	Existence, Services, Support
7	How does your organization currently serve its metadata catalogue to the community? In other words, what front facing or management software do you use for your repository (Drupal? GeoNetwork? CKAN? A custom site you built? Other?)	Technologies
8	What tools or platforms are you using to write metadata (for example ArcGIS Catalog, MERMAid, Darwin Core Archive Assistant)?	Technologies
9	If you could have free support from the new International Technology Office for any infrastructure project of your choice, what would you like to do?	Support, Technologies, Services
10	Do you supply a metadata service based on a standard protocol that can be used to harvest the catalogue content?	Standards, services
11	If yes, which specific metadata service protocol or protocols do you provide (OGC-CSW, OAI-PMH, Opensearch, other)?	Standards, services
12	Do you generate unique, persistent identifiers (for example DOIs, UUID, PURL) for your data holdings?	Standards, services
13	If you do not currently generate unique, persistent identifiers (for example DOIs, UUID, PURL) how interested are you in doing so in the future?	Standards, services
14	Do you serve any linked open data/semantic data?	Standards, services

15	If you do not currently serve data semantically, how interested are you in doing so in the future?	Standards, services
16	Is your repository listed on re3data (https://www.re3data.org/)?	Availability, Network
17	If it is listed, do you agree with the information re3data has presented about your repository?	Availability, Network
18	Is your metadata harvested by DataCite (https://datacite.org/)?	Availability, Network
19	Is your metadata harvested by Google Dataset search (https://toolbox.google.com/datasetsearch)?	Availability, Network
20	How interested would you be in getting support to help to make sure your data holdings represented in Google Dataset Search or DataCite? Assume that the support is free of charge.	Availability, Network
21	Do you have an off-site backup of your repository?	Availability, Services
22	How interested would you be in sharing summary statistics about your holdings, contributors and consumers with the ITO with the intent to create reports on the magnitude and reach of the WDS system (not individuals or record level contributions or consumption, but aggregated metrics)?	Reporting/clients
23	Is your organization interested in participating in a collective WDS Metadata Catalogue?	Initiatives
24	If you are not interested in participating in a collective WDS Metadata Catalogue, what is the primary reason for your Organization not participating?	Initiatives
25	Is your Organization interested in having all ISC-WDS members tag their metadata as 'GEOSS Data-CORE'?	Initiatives, Standards
26	How interested are you, or someone in your organization, in participating in a working group to further develop the WDS Knowledge Network. described here: https://www.icsu-wds.org/services/knowledge-network (The answer to this question DOES NOT sign you up for notifications, it only gauges your interest)?	Initiatives
27	If you know of anyone else within your organization (or otherwise) that might be interested in participating in a Knowledge Network Working Group, please list their details below.	Initiatives
28	Do you, or does anyone in your office, serve on any RDA IGs or WGs? If so which ones?	Initiatives

29	How interested are you in accessing or utilizing a brokering registry? A brokering registry is a list of tools that have been developed to help manage or move data and metadata between systems, for example by converting data between formats or finding equivalent semantic vocabulary terms. We are not asking you to be responsible for maintaining a registry, we are curious if it be helpful to you to be able to search a registry in order to find and re-use existing tools	Initiatives, Technologies, Services
30	How would you like the new WDS Information Technology Office to communicate with you (email, a Skype group chat, a members web forum, a Listserv, a Slack Channel, other)?	Initiatives, Contact, Support
31	The ITO is currently creating a Technical Advisory Committee drawn largely from the WDS membership. The primary purpose of the TAC is to advise the ITO on infrastructure strategies and technical road maps. We are creating a list of WDS members who may at some point in the future be interested in serving on the TAC. We are not asking for a commitment now, but are you, or someone in your organization willing to have your name added to a list of potential ITO Technical Advisory Group members? If so, please give your name or the name of the member of your organization you believe may be interested below. Again, the answer to this question confers no obligation on anyone's part to serve on the ITO TAC, we are only gauging interest. You can learn more about the TAC remit here: http://tinyurl.com/yyskos84.	Initiatives, Contact
32	Any other ideas about how the new International Technology Office can be of service to you or any other comments about the ITO?	Initiatives, Support
33	Any comments you would like to add to this survey or anything else you want to add?	--

World Data System
International Technology Office
Ocean Networks Canada
University of Victoria
#100-2474 Arbutus Rd Victoria, BC V8N 1V8
P 250 472 4527 | wds-ito.org