IJDC | *Research Paper*

Research Roles and Responsibilities at United States Universities, as Seen in Institutional Data Management Policies

Diana J. Castillo
Oregon State University

Zach Welhouse
Oregon State University

Clara Llebot
Oregon State University

Abstract

Research data management's goal is to make the process of collecting research data more organised, as well as balance the needs and requirements of a variety of stakeholders. These stakeholders include the organisations funding the research, the institutions employing the researchers, and the publishers. Typically, research has focused on funder and publisher data policies, but there has been relatively little work done on the content of institutional research data policies. They are worth examining because they cover all researchers working at a particular institution and can help identify institutional priorities. This paper examines how research data policies at R1 (doctoral, very high research output) universities in the United States assign roles and responsibilities to researchers and institutions, as well as what support is given to researchers. The analysis found discrepancies in how responsibilities were assigned, as well as a lack of enshrined institutional support structures for researchers. The paper concludes with recommendations on how universities can contextualise their policies to be more usable by both researchers and institutions, as well as how to incorporate specific support structures.

Submitted 1 February 2025 ~ *Accepted* 28 November 2025

Correspondence should be addressed to Diana J. Castillo, email: diana.castillo@oregonstate.edu.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Data management is a complex process with many steps and numerous stakeholders. Stakeholders—including researchers, funders, research organisations, and publishers—have different priorities and expectations regarding who is responsible for each aspect of the research data lifecycle (Briney, Goben, & Zilinski, 2017; Cox et al., 2017). Because of stakeholders' different expectations and priorities, the distribution of responsibilities (such as whose job it is to organise the metadata or steward the completed project's data) is often unclear and confusing. This article will explore the data management responsibilities of researchers and institutions in the United States.

Policies can be helpful to clarify the data management roles of the different stakeholders and their responsibilities in the research lifecycle. During the last couple of decades, countries around the world have created policies with the goal of emphasising the importance of data management, and the tasks that have to be accomplished to achieve reproducibility and transparency of research, as well as enable the reuse of data. For example, at a funder level, the European Commission requires that all funded projects through the Horizon programme submit data management plans within the first six months of the award. These plans are structured around the FAIR principles (Wilkinson et al., 2016), and guided by the concept "As open as possible, as closed as necessary". In the United Kingdom, UK Research and Innovation (UKRI) requires that research data resulting from its funding will be made as open as possible, and that good data management practices will be followed during the project. Many funders require data management plans (UK Research and Innovation, 2024a). In Australia, the Australian Research Council has required researchers to outline a plan for how they manage research data since 2014, encouraging sharing of research data (n.d.). In the United States, a 2022 update of the Office of Science and Technology Policy memorandum about Increasing Access to the Results of Federally Funded Scientific Research required all federal agencies to develop or update their public access plans to ensure that scientific data underlying peer-reviewed scholarly publication will be made freely available and publicly accessible by default at the time of publication (OSTP, 2013; 2022).

The extent to which these policies clarify the roles of stakeholders varies, but in general there is an understanding that the ownership of the data resides with the institution and the researcher, and that they are responsible for managing the research data, while funders and journals have an oversight role and, in the case of funders, also the responsibility of providing funding, infrastructure, and guidance. For example, UKRI, in their "Publishing your research findings" guidelines, say that "Ownership of the data generated from the research funded by UKRI or our councils resides with the researchers or their institutions. They should maintain and manage copyright and intellectual property ownership of data so that underlying research materials remain as open as possible" (2024a).

The distribution of responsibilities between researcher and institution isn't usually well documented. In the United States, federal agencies give funding to universities, which are responsible for disbursing the money to researchers. It is the researchers, however, who are expected to comply with large portions of the funders' policies. Clearly defined roles and responsibilities are important to ensure that researchers understand their responsibilities, as well as ensure that both the researchers and institution are protected should the need arise. A solid understanding of these responsibilities can also facilitate the establishment of institutional support, such as library data curation experts, institutional repository access, policies that regulate how to work safely with sensitive data, or adequate and safe data storage.

Institutional research data management policies are our primary form of documentation for the distribution of roles and responsibilities between researchers and institutions. These policies describe the relationship between the institution and the researcher, while potentially encouraging valued actions and attitudes (Llebot & Castillo, 2023). These policies may indirectly affect other participants in the data lifecycle, such as repositories and data curators.

The goal of this paper is to use institutional data management policies to:

1. Explore the roles and responsibilities that universities assign to their researchers and themselves, specifically focusing on data access.

2. Highlight inconsistencies or imbalances (if any) in the roles that universities assign themselves and their researchers.

3. Evaluate whether institutions commit to providing support in their policies, and how much of this expressed support relies on library services.

Literature Review

Effective institutional data management and research support presents many challenges. As in any system with distributed rights and responsibilities, maintaining accurate and accessible documentation is a key challenge. However, the literature suggests that policies can be an efficient way of providing this support. For example, Fecher, Friesike, and Hebing (2015) highlight researcher interest in data sharing far outstrips actual practice due to the strong motivations for rapid publication within academia, and suggest that institutions can get the most from their data ecosystems by providing data management support (such as policies, training, and guidance) to clarify responsibilities and make the task as easy as possible. Similarly, Lefebvre and Spruit (2019) found that the regulatory, procedural, and technological challenges faced by research teams can be alleviated through procedural clarity and institutional support for staff responsible for maintaining research data, such as through creating clear policies. They also observed that stakeholders who receive institutional support for engaging with clear, manageable policies are more likely to engage in positive practices such as data sharing and preservation.

Unambiguous data policies protect institutional investments in easy-to-parse documents highlighting the rights and responsibilities of the people and resources involved in the research process. A clear policy informs the research community of the legal aspects of their work and how the material is authorised for commercial use (Donner, 2023; Erway & OCLC Research, 2013). Erway and OCLC Research (2013) suggest the key elements covered in an effective data policy include data preservation, retention, maintenance, ownership, access, and openness. Situating these topics within a clear organisational structure allows institutions and their researchers to have more coherent workflows and protect their reputations by upholding moral and legal principles. As a result, a clear data policy promotes confidence among researchers and those impacted by their research. Moreover, a strong data policy can encourage data-sharing and help educate researchers about institutional support (Higman & Pinfield, 2015).

Conversely, unclear data policies can create governance challenges and exacerbate tensions between aspirational goals and the practicalities of research (Erway & OCLC Research, 2013; Higman & Pinfield, 2015). These issues are especially concerning when coupled with research funded outside of the institution, which bears the reputational and legal weight of contractual obligations. York, Gutmann, and Berman (2018) warn against policies that create stewardship gaps: lacunae where no party is legally or practically responsible for important steps in the preservation, curation, or publication of institutionally-owned data.

Differences within academic traditions and cultural structures account for a variety of data management roles, responsibilities, and—importantly—rights across institutions; however, the most common roles within research and data management infrastructures provide a starting point for comparison. Common participants in the research process include representatives of institutional interests and data centres, funders, scientists and other researchers, information users, and publishers (Lyon, 2007). Under this schema, researchers have many rights, including the right of first use to data and to receive data training and advice from their institutions, as well as the responsibilities to manage data, meet standards for good practice, and comply with funder and institutional data policies. Pryor (2014) identifies the roles and responsibilities of a typical data management team: researchers, managers, legal authorities, data management,

specialist advisers (e.g., university representatives), archivists, repositories staff, data centre employees, and other custodians.

These data management challenges are not unique to research within the United States. In the United States, funders provide research grants to universities rather than individuals; therefore, the university is responsible for compliance requirements (Briney, Goben, & Zilinski, 2017). This explains US institutional policies' focus on compliance and ownership. However, European Open Science plans show comparable ambiguities regarding group policies and individual responsibilities. For example, the European Open Science Cloud (EOSC) facilitates FAIR data principles across a multitude of European data infrastructures, including domestic, national, and international organisations, where unclear policies introduce friction to communication. Bridging existing data silos to provide free, federated access and analysis to 1.7 million European researchers is a monumental task, whose ultimate success depends on attracting a critical mass of compliant organisations and individuals (Budroni, Claude–Burgelman, & Schouppe, 2019). Davidson et al. (2019) likewise identify unclear, non-standardised policy language as contributing to less effective policies. Their report analyses 11 European data management policies across three stakeholder groups (higher education institutions, funders, and publishers), employing the Turning FAIR into Reality (TFIR) action plan to determine funding body policies are a key driver in stakeholders developing FAIR policies. Similarly, the Digital Preservation Coalition's FAIR Forever analysis of the EOSC recommends the community clarify data management terminology, roles and responsibilities, and assumptions in order to facilitate understanding of nuanced problems and make its implementation more appealing (Currie & Kilbride, 2021). UK Research and Innovation (UKRI) likewise support open research through data sharing as part of their 2025 research data management framework (2024b). The previous findings on data management terminology among European institutions suggests similar analysis of large American institutions of higher education will provide a more complete understanding of effective data management policies. Moreover, our focus on the distribution of roles and responsibilities among those policies provides a new lens for analysing policies' value.

Methods

The collection of policies used for this project developed from work done in Llebot and Castillo (2023), which collected and analysed institutional research data policies based on how they supported the FAIR principles. The same criteria are used when identifying institutional research data policies for this analysis.

The criteria are:

1. The policy must be standalone and focused on research data. While there can be mentions of intellectual property (IP) within the policy, it cannot be the main focus.

2. The policy is university-wide, issued and administered by the university's administration. This eliminates policies that only apply to certain segments of the university (e.g., a medical school) or non-administration policies originating from faculty senate, departments, or the like.

This study focuses solely on data management policies for research data. Policies governing roles and responsibilities that affect both business and research data are excluded. For example, the Colorado School of Mines' (n.d.) Data Code of Conduct outlines the responsibilities of people working with university data including personal information. Washington State University's (2022) Data Administration Policy covers institutional data and information systems that the university keeps as an employer. These policies fall outside the scope of this study.

The list of policies used for this study is an updated version of the dataset Llebot and Castillo (2021). In our original study, we collected the policies published by universities classified as "Doctoral University - Very High Research Output (R1)" by the Carnegie Classification of

Institutions of Higher Education. We identified their institutional data policies by contacting the data librarians at each institution, searching online, and revising similar lists that had been collected by other authors in previous studies. This present article reflects the updated list of universities using the Carnegie Classification of Institutions of Higher Education (2022) list. The revised dataset includes 13 new R1 universities, bringing the total number of R1 universities to 148. We conducted a manual search of the websites of each new university and found 14 new policies. In September 2022, all the university websites were reviewed to ensure that we had the most recent versions of all policies. This study does not include policies published or updated after that date. Overall, for this work, we have considered 50 policies.

The policies were evaluated using a content analysis methodology (Sage Research Methods, 2017). The three authors created a set of criteria (see Appendix) focused on how policies describe the roles and responsibilities for researchers and the university, including whether the policy describes data management as a shared responsibility, and how the policy indicates support for researchers with their data management responsibilities. The criteria were developed by brainstorming a list of initial topics of interest, reading some policies, and revisiting the criteria so that they reflected the information we wanted to capture from the policies. We iterated the process several times to account for specific terminology and ways of expressing responsibility within the dataset. For example, Criteria 6 ("The policy describes responsibilities of a data steward") also includes policies that refer to a "researcher", "PI", or "principal investigator", as the terms are used interchangeably in practice. When developing the criteria, we made an effort to minimise overlap with the criteria developed for our previous study (Llebot & Castillo, 2023), with the goal of avoiding reporting similar information in different publications.

We tagged each policy as not meeting the criteria, meeting the criteria with "clear language", or meeting the criteria with "vague language". In general, we coded "vague language" when a responsibility was described, but it was not clear who was the responsible party. For example, let's take a look at the excerpts from Harvard and Old Dominion University that we considered when coding for the criteria "The responsibilities of the data steward/university include complying with applicable state and federal laws and regulations" (Criteria 6.6 and 7.6). Harvard's University Ownership Policy says that "The PI's responsibilities with respect to research data include, but are not limited to: (...) complying with applicable federal, state, and local laws and regulations". This was coded as "clear language" because it is clear that the PI is the responsible party, and it refers unambiguously to all applicable laws and regulations. On the other hand, Old Dominion University's policy says "Research data should be: (...) Kept in a manner that is compliant with legal obligations and, where applicable, the requirements of funding bodies and project-specific protocols approved through Office of Research". We qualified the text as "vague language" because the passive voice and the mention of the Office of Research make it confusing to understand who is responsible, and "legal obligations" is less specific than "applicable federal, state, and local laws and regulations".

All policies were tagged by at least two researchers, and differing answers were reconciled through discussion. We kept notes of decisions made during these discussions to ensure that the criteria were applied consistently to all policies.

The dataset used for this analysis, which includes the universities included, the links to the publicly available data policies that we have analysed, the coding of each of the criteria, and the paragraphs that justify the coding, is publicly available as Castillo, Welhouse, and Llebot (2025).

Results

Table 1. Number of policies that assigned a certain responsibility to either the researcher or the institution, of a total of 50. The criteria number is indicated as an X.number when the criteria was coded for both researchers and institutions. The table shows how many policies fulfilled each criteria with clear language (Clear), or vague

language (Vague), and the total of Clear+Vague (Total). The full list of criteria can be found in the Appendix.

	Criteria	Policies that assign responsibility to a researcher			Policies that assign responsibility to a university		
		Clear	Vague	Total	Clear	Vague	Total
X.3	Data management in general, overseeing research data	48 (96%)	2 (4%)	50 (100%)	29 (58%)	6 (12%)	35 (70%)
X.4	Data retention and preservation	49 (98%)	1 (2%)	50 (100%)	23 (46%)	5 (10%)	28 (56%)
X.5	Complying with terms of programmes/sponsored project agreements/contracts	37 (74%)	6 (12%)	43 (86%)	37 (74%)	6 (12%)	43 (86%)
X.6	Complying with applicable state and federal laws and regulations	35 (70%)	4 (8%)	39 (78%)	29 (58%)	10 (20%)	39 (78%)
X.7	Data protection (human subjects, backups, data security and access, etc.)	38 (76%)	4 (8%)	42 (84%)	30 (60%)	8 (16%)	38 (76%)
X.8	Protection of rights of people or entities (e.g., intellectual property, misuse of scientific record, etc.)	26 (52%)	11 (22%)	37 (74%)	34 (68%)	5 (10%)	39 (78%)
X.9	Making decisions regarding data access	28 (56%)	5 (10%)	33 (66%)	25 (50%)	4 (8%)	29 (58%)
X.10	Giving access to appropriate parties	35 (70%)	10 (20%)	45 (90%)	31 (62%)	4 (8%)	35 (70%)
X.11	Educating researchers and/or participants regarding research data management	32 (64%)	3 (6%)	35 (70%)	13 (26%)	1 (2%)	14 (28%)
6.12	Ensuring accuracy and authenticity of research data	18 (36%)	9 (18%)	27 (54%)	-	-	-

6.13	Assisting university in either investigation of charges (e.g., misconduct or conflict of interest) or resolving disputes arising from research	4 (8%)	1 (2%)	5 (10%)	-	-	-
7.12	Providing infrastructure or financial support (e.g., storage or repositories)	-	-	-	12 (24%)	4 (8%)	16 (32%)
7.13	Providing support to researchers different than infrastructure or financial	-	-	-	34 (68%)	0 (0%)	34 (68%)
7.14	Investigating charges (e.g., misconduct or conflict of interest) or resolving disputes	-	-	-	34 (68%)	2 (4%)	36 (72%)

The content analysis performed in this project focuses on three main sections, to answer our research questions. First, we explore what the policies say in terms of data management responsibilities in general (Criteria 6.3 and 7.3), and whether the responsibility should be shared (Criteria 8). Second, we look at the distribution of the responsibilities between researcher and institution, taking a closer look at data access in particular (Criteria 6.X and 7.X, but special emphasis in 6.9, 6.10, 7.9, 7.10). Finally, we explore support for data management outlined in the policies, highlighting the role of the library in this (Criteria 7.13). The findings from the first two sections inform our first and second goals of exploring roles and responsibilities assigned by universities and evaluating possible imbalances. The third section is directly related to the third goal, which focuses on institutional commitment to providing support.

The numbers reported throughout this paper are total numbers (Table 1), except when we explicitly mention clear language or vague language.

Data as a Shared Responsibility

A little more than half of the policies analysed for this project described data management as a shared responsibility between researchers and the universities (Criteria 8). Of the 30 policies that met this criteria, 13 (26% of the total) used the same language: "Both the University and the Principal Investigator (PI) have responsibilities and rights concerning access to, use of, and maintenance of research data." The remaining 17 universities used different wording to explain the shared responsibility and some added additional context.

The universities that provide additional context on data management as a shared responsibility give two main explanations. The first centres on university ownership of research data, with four universities (Northwestern, Wayne State, Virginia Polytechnic, and University of Nebraska–Lincoln) explicitly mentioning university ownership. While numerous universities include research data ownership in their overall policies, only those four use it as a framework for shared responsibility.

The other main policy explanation for shared data management is meeting sponsor and publisher data requirements. Four universities (Brown, Northwestern, University of New Hampshire, and University of Rochester) include language around outside data requirements. This is particularly interesting in the context of how many policies assign complying with

sponsor requirements as individual responsibilities for both researchers and universities (Criteria 7.5). These four universities make that compliance an explicit reason why researchers and universities need to work together to manage research data.

One final item of note when looking at the universities whose policies list data management as a shared responsibility: only two universities (University of Minnesota Twin Cities and University of Kentucky) name specific administrative positions and how they are expected to work with researchers on data management responsibilities in this section of their policies.

Assigned Responsibilities and Data Access

All the policies in our study assign at least one responsibility to a researcher (Criteria 6), and all except for two (Binghamton University and Duke University) assign at least one responsibility to the university with clear language (Criteria 7). Of the 50 policies included in this study, 48 (96%) have explicit language indicating that data management is a key responsibility for researchers (Criteria 6.3); the other two have vague language around the issue. In contrast, only 29 (58%) clearly state that the university is responsible for data management or overseeing research data (Criteria 7.3), while six (12%) include ambiguous language, and 15 (30%) do not mention it. Despite this frequency, only 21 (42%) universities' research data policies provide responsibilities formatted as a list for both researchers/PIs and universities (Criteria 6.1 and 7.1 respectively). Of the remaining universities, three universities (6%) only have detailed lists for researchers/PIs, 13 (26%) only have detailed lists for universities, and 13 (26%) have no detailed lists of roles and responsibilities for either researchers or universities.

Table 1 shows the frequencies of the responsibilities that are assigned to researchers and institutions, with clear or vague language. These results suggest that, generally, policies tend to assign more responsibilities to researchers than to institutions. Most institutions expect researchers to be responsible for data retention and preservation (Criteria 6.4); complying with laws, regulations, and the terms of contracts and sponsored project agreements when managing their data (Criteria 6.5, 6.6); protecting data as needed (Criteria 6.7, 6.8); and educating other members of the research team on data management practices (Criteria 6.11). Many policies indicate that these are also institutional responsibilities, but at significantly lower frequencies. There are only three exceptions: (1) the responsibility of complying with the terms of programmes, sponsored project agreements, and contracts is assigned to institutions with the same frequency as it is assigned to researchers (Criteria 6.5 and 7.5); (2) the responsibility of protecting the rights of people or entities is assigned more frequently to institutions than researchers (Criteria 6.8 and 7.8) (when applied to researchers, this usually means that the researcher is responsible for protecting the university intellectual property, by keeping the data long enough, or protecting university proprietary information and trade secrets by not sharing the data with unauthorised individuals; in the case of institutions, protecting the rights of people or entities usually means that the university commits to protecting the rights of researchers, such as their right to access data from research in which they participated); (3) responsibilities regarding data access (Criteria 6.9, 6.10, 7.9, and 7.10), which are assigned at very similar frequencies to universities and researchers. Because of the complexity of data access responsibilities, and this similarity in the number of policies that assign it to institutions and researchers, we have explored these criteria in more detail.

The data access criteria differentiate between the ability to make decisions regarding data access (Criteria 6.9 and 7.9) and the responsibility of actually giving access to appropriate parties (Criteria 6.10 and 7.10), but these will be combined in this discussion. We did not include in these criteria statements regarding data access that were directly related to data transfer (when a researcher leaves the university) or data ownership, to keep the discussion concise and focused. It is, however, worth mentioning that most policies include these. For readers interested in learning more about what institutional policies say about data ownership and transfer, see the dataset Llebot and Castillo (2022). Text analysis of these criteria leads to several observations, described below until the end of this section.

Twenty policies (40%) state that it is the responsibility of the PI to ensure that the data is available to at least one of the following entities: research group and co-investigators (11 policies,

in text analysis), the university (13, in text analysis), sponsors (ten, in text analysis), journals (four, in text analysis), or others, which sometimes includes the public (11, in text analysis).

Access to co-investigators language varies throughout the seven policies where it is mentioned. While some policies emphasise the right of the PI to have access to the data, and others stress that co-investigators have the right to access research data that they have worked on, most policies focus on defining the role of the PI as the person who decides who gets access to what data, when these decisions must be made (i.e., when a new researcher joins the project), and what documentation is needed (i.e., notifications in writing of any limitations to the use of data), as well as the importance of communicating these decisions to the team. For example, the Colorado State University Fort Collins policy says that PIs are responsible for "Making reasonable and equitable decisions about the use of and access to Research Data and Materials by those involved in projects from which they arise including, without limitation, students".

Thirteen (26%) policies clearly state that it is the responsibility of the PI to provide data to the university. The language that states how and when this access must be provided varies, but there are three clarifications that appear recurrently: that the access must be timely and prompt; that access must be provided under the appropriate circumstances (e.g., when necessary to conduct or respond to audits and investigations), and upon request; and that access must be provided regardless of the location of the responsible investigator (e.g., even after the researcher has left the university). One university also specifies that the data must be provided in readable form.

Thirteen universities (26%) mention public access in some shape or form, and several of these specifically discuss that it is the responsibility of the PI to give access to the public when appropriate. A common way of expressing it is "Principal investigators share research data, including placing the research data in public repositories, unless specific terms of sponsorship or other agreements supersede this right."

The roles assigned by university policies to the institutions cover mainly two aspects of data access: the right of the university to access research data when needed (Criteria 7.9, 25 policies, 50%), and the responsibility of the university to secure data access to others (Criteria 7.10, 31 policies, 62%). The right to access data is usually framed in the context that the university must have access to data so that they can give access to others (e.g., funders, FOIA requests, etc.). Access is also justified for circumstances in which the institution must facilitate a response to research misconduct allegations. The majority of these policies explain that when necessary, the university will take custody of the data. The responsibility of giving access to others is often expressed as "Protecting the rights of researchers, including, but not limited to, their rights to access data associated with research in which they participated". Fourteen of these (in text analysis) policies make clear that "researchers" also include staff and students.

Support as a Responsibility

Four of our criteria assess the types of support that are captured in the policies. Responsible researchers are in charge of educating other researchers or study participants regarding data management (Criteria 6.11) in 32 policies (64%), while institutions only are assigned this responsibility (Criteria 7.11) in 13 cases (26%). Twelve policies (24%) include clear language that expresses the responsibility of the institution to provide some sort of infrastructure or financial support (Criteria 7.12). One of these universities (University of Massachusetts Amherst) specifically mentions financial support, while nine mention infrastructure such as storage. Other types of support (Criteria 7.12) are mentioned in two thirds of the policies (34 policies, 68%). This type of support can be varied, but some of the elements mentioned are assistance with development of data management plans, assistance to store, retain, or generally manage data, maintaining policies related to data management, resources to ask questions about the policy, and how to contact them, or library services.

The supportive role of the library is not thoroughly captured in any of the criteria, but since it is of special interest for the audience of this paper, and it can range from infrastructure to educational support, we include a summary below.

Eleven of the 50 policies (22%) mention the roles and responsibilities of libraries or librarians in data management. Three of these research data policies position the libraries as responsible for long-term preservation, access, or curation of research data. An additional two policies, Texas A&M University and Utah State University, highlight their libraries' connection to the university's digital research repository.

Other typical duties undertaken by campus librarians as seen in these policies include consulting on practical and ethical documenting, preserving, and disposing of data; consulting on sponsor requirement compliance; determining future research data management requirements; teaching researchers about data management; and serving on scholarly data governance committees. The policies of Iowa State and Minnesota Twin Cities are notable for explaining most of these responsibilities.

Some policies position university libraries and librarians as resources for their researchers, while others use stronger language to indicate one of the library's primary duties is to support researchers. The University of California takes the former approach: "University Researchers are encouraged to consult with the California Digital Library, campus libraries, or other campus or systemwide resources". More direct language includes Iowa State University's charge that the university library "Ensures the long-term preservation, access, and curation of research data entrusted into its care to meet federal, state, sponsor, and university requirements."

Discussion

This discussion is structured in three parts: first, we address the division of roles and responsibilities (first research question). Second, we highlight imbalances in how these policies distribute responsibilities (second research question) compared to the support they offer, and focus on the role of the library as a support structure (third research question). Third, we contextualise these observations within the larger landscape of data management policy. Finally, we close with recommendations for the inclusion of roles and responsibilities in policies.

Defining Roles and Responsibilities

This analysis shows that data policies can help define roles and responsibilities of researchers at universities, which can impact further efforts to make data accessible, as well as ensure its quality.

Overall, policies clearly show that data management is considered a shared responsibility. More than half of the policies explicitly state this. Others simply announce that general data management is the responsibility of researchers and then state a similar responsibility for the institution. All of the policies, except for two, at the very least outline one or more responsibilities for both.

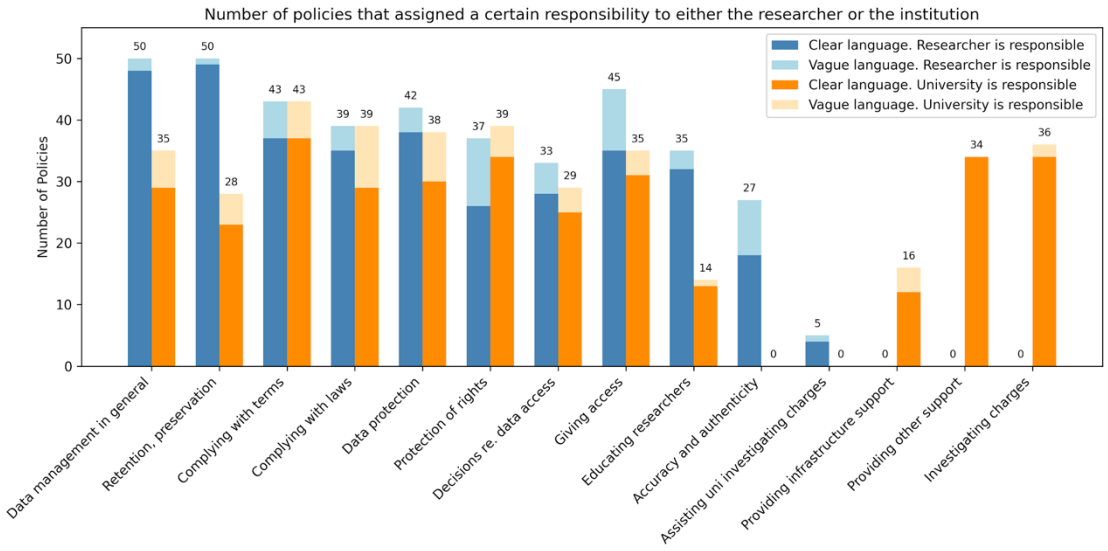

Figure 1. Responsibilities assigned to either the researcher or the institution by the 50 policies reviewed in this article. See Table 1 for percentages.

Several trends and gaps emerge when looking at how roles and responsibilities are defined within the policies (Figure 1). It is worth noting that these trends emerge when looking at all the policies as a body, but it does not mean that every policy individually exhibits these. First, basic responsibilities for data management, as well as retention and preservation, are more likely to be given to researchers than universities. The gap in how policies assign retention and preservation duties is particularly notable. With more funders requiring data retention and sharing, this places a potentially unequal burden on researchers. This is particularly concerning since universities own research data provided at their institution. While several policies specifically mention institutional repositories as a potential data storage location, or support in terms of library services or storage infrastructure, it is not uniform. York, Gutmann, and Berman (2018) warned of policies that include no mention of retention or preservation, and so the fact that responsibilities regarding retention and preservation are mentioned in all but two policies is definitely a positive trend. The uneven distribution of these responsibilities, however, is an area that could create meaningful stewardship gaps.

Another notable trend is whether policies are designed with usability in mind. Some policies are structured as basic legal frameworks, just in case something happens. Others are designed as documents that are meant to be helpful and consulted often. Llebot and Castillo (2023) discussed how most policies in the US are not designed to encourage or support researchers in the implementation of the FAIR principles, and argued that policies could be a good tool for that purpose. In this case, we also observe this dichotomy in how policies structure the responsibilities of researchers vs universities. Overall, policies are much more likely to outline responsibilities in clear and well-structured lists for institutions than for researchers (24 (48%) for researchers compared to 34 (68%) for universities), which is surprising given that generally these policies tend to include more researcher responsibilities than university responsibilities. Only 21 policies (42%) contain well-defined lists for both. Responsibilities assigned to researchers are more likely to be in an unstructured format, which could make it difficult to discern what responsibilities are explicitly assigned to them, and make these policies less usable.

Policies also vary in how specific they are to the institution that wrote them. The use of shared language from institutional peers is a common practice, and often we see the same expressions in several policies. But we can also see text that provides context or is tailored to a university's specific needs. For example, in the case of these well-defined lists of responsibilities, we find that some of them are generic lists that outline the researcher's responsibilities with regards to management, retention, communicating with team members, etc. These policies are short and to the point, and they may provide flexibility for institutions and researchers in a

rapidly changing research landscape. The lack of specificity however, can risk stewardship gaps. More detailed lists, such as the University of California system's, reference specific university offices like the Vice-Chancellor of Research, or expand upon the scope of the responsibilities: "Determine use of the Research Data by other University Researchers and collaborators on the project in accordance with relevant agreements and their scholarly discipline's practices, and taking into account the need for academic progress of academic appointees, post-doctoral scholars, degree candidates, and other students." While the more generic lists may be useful for acquainting researchers and institutions to their roles and responsibilities, more detailed lists that reflect the specific circumstances of the institutions may be more useful in the long run and are a good opportunity to emphasise the support that researchers can expect from institutions and where to seek it.

Enshrining Specific Support Structures and the Role of Libraries

Our analysis reveals that institutional policies can be a good place to define supporting structures, but it also shows that they are not consistently used for that. Strong support like infrastructure and financial support is only mentioned in twelve policies (24%). Soft support like guidance, development and maintenance of policies, and assistance is included more often, in two thirds of the policies (34, 68%). Education regarding data management is twice as likely to be assigned as a responsibility of the researcher than as a responsibility of the institution. Its omission does not unequivocally mean the universities do not provide these supports, of course, but the inclusion of support systems in a policy provides accountability to the institution and clarity. In our opinion, all institutional research data management policies should include descriptions of support systems for managing data.

Libraries are a specific support structure that provide many forms of support beneficial to research data management. However, this unique expertise is not evenly represented among our surveyed policies: only 22% of the policies include library roles. Evidence supports the importance of librarians within institutional research data environments. Briney, Goben, and Zilinski (2015) examine the presence of university-level data policies, data librarians, and data services within academic libraries to conclude librarians contribute to researcher awareness of and compliance with positive data practices. Tammaro et al. (2019) highlight data curation librarians' unique research expertise, including information organisation, metadata, and archiving. They likewise support explicit recognition of librarians in data policies, additionally noting that data curation librarianship would be further served with additional support. Long-term data management presents several organisational and technological challenges that are best shared by multiple individuals or groups (Pryor, 2014). With the clear support of a policy, librarians are well-suited to being part of these groups.

Contextualisation

Successful policies will be policies adapted to their context. This includes institutional context, and the context of the specific country where the policy has been developed. It is possible that some of the recommendations outlined in this text will not make sense outside of the United States. For example, the Canadian Tri-Agencies (the government's three main funding bodies) published the "Tri-Agencies Research Data Management Policy" (Government of Canada, 2021). The policy asked all postsecondary institutions and research hospitals eligible to administer Tri-Agency (CIHR, NSERC or SSHRC) funds to post an institutional research data management strategy by 2023, and to outline what steps the institution is taking or plans on taking to support researchers. The intent of the policy is to ensure that the affected institutions will outline how they support data management, but it is very flexible in how these institutions document their strategy. In Canada, then, documentation related to institutional data management support may be in a policy, but it could also be in a different document.

The use of policy templates, when existing, can be useful. Adapting material that has been already tested allows institutions to focus on the aspects that make their research environment unique. Guidelines and templates exist in some communities. For example, the DCC in the UK

(Donnelly, 2014), the Australian Research Data Commons in Australia (Australian Research Data Commons, 2021), or the Consortium of University Services in Catalonia (Consorci de Serveis Universitaris de Catalunya, 2018) have developed templates to facilitate the adoption of policies by the institutions in their communities. When templates or guidance do not exist, it can be useful to look at examples. For institutions in the United States, the dataset published with this manuscript (Castillo, Welhouse, & Llebot, 2025) and Llebot and Castillo (2022) include examples of language for recent policies, organised according to diverse topics. It is important to stress, however, that every template, guidance, or example should be tailored to the institution for the policy to be effective and meaningful.

The research landscape can change over time. New national regulations, funder policies, or laws can affect institutional policies. These policies are usually written in such a way that policy updates are not required following each of these changes. For example, expressions like "Research Data must be maintained, managed, archived, and disseminated in accordance with all applicable requirements of research sponsors" ensure the applicability of the policy will remain. Regular updates, however, will guarantee that policies remain responsive to a changing data landscape, and policies should show clearly when they were approved and updated.

Recommendations

Our recommendations are the following:

- Include responsibilities of researchers and institutions in a policy. This study argues that institutional policies are used for this purpose, and it is valuable to do so.
- Structure these responsibilities in a well-defined list to make the policies actionable and easy to use for both the university and researchers.
- Include language specific to the institution. Institutional context matters. Even though the use of shared language from institutional peers and templates (when available) is helpful, policies should focus on the aspects that make their research environment unique.
- Include information about institutional support. Researchers, as university employees, are the ones in charge of performing most of the tasks that are necessary to manage data properly, even if the institution is ultimately responsible for most of them. This ultimate responsibility often means that the university will take action only if there are problems (for example, take custody of the data if there are allegations of research misconduct) or if the researcher cannot perform these tasks (for example, retain data from a researcher who left the institution), but it also means that the institution must also provide resources to support researchers with these tasks. Since providing support is intrinsically linked to fulfilling a responsibility, our recommendation is that all policies document the resources that the institution invests in data management (e.g., committing to providing infrastructure for storage, backup and security of the data, computing, repositories, guidance, etc.).
- University libraries can be a strong partner to support and provide policy training during implementation. They should be included in policies when appropriate.

Conclusion

This study looks at how institutional research policies in the United States outline the roles and responsibilities of researchers and institutions. Even though the documents generally make clear that data management is a shared responsibility, either textually or by describing responsibilities for both parties, there is an imbalance between the responsibilities that are assigned to

researchers and the support that institutions commit to in these documents. Although undocumented support does not mean that it does not exist, we recommend that institutional policies describe these supports, including the support provided by libraries, to strengthen and clarify their responsibilities. Policies are documents that are context-dependent, so we recommend that they are tailored to the specific institution, country, and research landscape where they will be implemented. The dataset accompanying this document collects language for 50 policies, reflecting the roles and responsibilities outlined by their institutions, and can be used to that effect.

References

Australian Research Council. (n.d.). *Research Data Management | Australian Research Council.* Retrieved January 30, 2025, from https://www.arc.gov.au/about-arc/arc-strategies-and-policies/international/research-data-management

Australian Research Data Commons. (2021). *Institutional Research Data Management Policies and Procedures* (Zenodo). doi:10.5281/ZENODO.5784765

Briney, K., Goben, A., & Zilinski, L. (2015). Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies. *Journal of Librarianship and Scholarly Communication* 3(2), Article 2. doi:10.7710/2162-3309.1232

Briney, K., Goben, A., & Zilinski, L. (2017). Institutional, Funder, and Journal Data Policies. In L. R. Johnston (Ed.), *Curating research data volume one: Practical strategies for your digital repository* (pp. 61–78).

Budroni, P., Claude–Burgelman, J., & Schouppe, M. (2019). Architectures of Knowledge: The European Open Science Cloud. *ABI Technik* 39(2), 130–141. doi:10.1515/abitech-2019-2006

Carnegie Classification of Institutions of Higher Education. (2022). *Basic Classification Description.* Retrieved from https://carnegieclassifications.acenet.edu/

Castillo, D., Welhouse, Z., & Llebot, C. (2025). *Dataset for "Research roles and responsibilities at United States universities, as seen in institutional data management policies" (Version 1, p. 16 KB)* [Data set]. Corvallis, OR: Oregon State University. https://doi.org/10.7267/xg94j0250

Colorado School of Mines. (n.d.). *Mines Data Code of Conduct. Policy Library.* Retrieved January 29, 2025, from https://www.mines.edu/policy-library/data-code-of-conduct/

Consorci de Serveis Universitaris de Catalunya. (2018). *Model de política de gestió de dades de recerca per a una universitat.* Retrieved from https://www.csuc.cat/sites/default/files/2021-06/1807ModelPolitica_RDM.pdf

Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology* 68(9), 2182–2200. doi:10.1002/asi.23781

Currie, A., & Kilbride, W. (2021). *FAIR Forever? Long Term Data Preservation Roles and Responsibilities, Final Report* (Zenodo). Retrieved from doi:10.5281/zenodo.4574234

Davidson, J., Engelhardt, C., Proudman, V., Stoy, L., & Whyte, A. (2019). *D3.1 FAIR Policy Landscape Analysis.* doi:10.5281/ZENODO.3558172

Donnelly, M. (2014). *Five Steps to Developing a Research Data Policy | DCC.* Retrieved from https://www.dcc.ac.uk/guidance/policy/five-steps-developing-research-data-policy

Donner, E. K. (2023). Research data management systems and the organization of universities and research institutes: A systematic literature review. *Journal of Librarianship and Information Science* 55(2), 261–281. doi:10.1177/09610006211070282

Erway, R. & OCLC Research. (2013). *Starting the conversation: University-wide research data management policy* (OCLC Research). Retrieved from http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf

Fecher, B., Friesike, S., & Hebing, M. (2015). What Drives Academic Data Sharing? *PLOS ONE* 10(2), e0118053. doi:10.1371/journal.pone.0118053h

Government of Canada. (2021). Tri-Agency Research Data Management Policy. Innovation, Science and Economic Development Canada. Retrieved from https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97610.html

Higman, R., & Pinfield, S. (2015). Research data management and openness: The role of data sharing in developing institutional policies and practices. *Program: Electronic Library and Information Systems* 49(4), 364–381. doi:10.1108/PROG-01-2015-0005

Lefebvre, A., & Spruit, M. (2019). A socio-technical perspective on reproducibility in research data management. In *Proceedings of The 13th Mediterranean Conference on Information Systems (MCIS) (Naples, Italy)* (10). Retrieved from https://aisel.aisnet.org/mcis2019/10

Llebot, C., & Castillo, D. (2021). *Research Data Policies at Doctoral Universities with very high research activity (R1 institutions)* [Data set]. Corvallis, OR: Oregon State University. doi:10.7267/4J03D6263

Llebot, C., & Castillo, D. (2022). *Content Analysis of Institutional Research Data Policies in the US (Version 1)* [Data set]. Corvallis, OR: Oregon State University. doi:10.7267/2N49T893T

Llebot, C., & Castillo, D. J. (2023). Are Institutional Research Data Policies in the US Supporting the FAIR Principles? A Content Analysis. *Journal of eScience Librarianship* 12(1), Article 1. doi:10.7191/jeslib.614

Lyon, D. L. (2007). *Dealing with Data: Roles, Rights, Responsibilities and Relationships. Consultancy Report.* Retrieved from https://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

Pryor, G. (2014). Who's doing data? A spectrum of roles, responsibilities and competences. In G. Pryor, S. Jones, & A. Whyte (Eds.), *Delivering Research Data Management Services* (1st ed., pp. 41–58). doi:10.29085/9781783300242.004

Sage Research Methods. (2017). *Data Analysis and Interpretation.* doi:10.4135/9781526408570

Tammaro, A. M., Matusiak, K. K., Sposito, F. A., & Casarosa, V. (2019). Data Curator's Roles and Responsibilities: An International Perspective. *Libri* 69(2), Article 2. doi:10.1515/libri-2018-0090

UK Research and Innovation. (2024a). *Making your research data open.* Retrieved from https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/

UK Research and Innovation. (2024b). *UKRI developing new research data policy framework.* Retrieved from https://www.ukri.org/news/ukri-developing-new-research-data-policy-framework/

Washington State University. (2022). *EP08 – WSU System Data Policies | Office of Policies, Records, and Forms | Washington State University.* Retrieved from https://policies.wsu.edu/prf/index/manuals/executive-policy-manual/ep08/

White House Office of Science and Technology Policy (OSTP). (2013). *Expanding Public Access to the Results of Federally Funded Research.* Retrieved from https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

White House Office of Science and Technology Policy (OSTP). (2022). *Desirable Characteristics of Data Repositories for Federally Funded Research.* doi:10.5479/10088/113528

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), Article 1. doi:10.1038/sdata.2016.18

York, J., Gutmann, M., & Berman, F. (2018). What Do We Know about the Stewardship Gap. *Data Science Journal* 17, 19. doi:10.5334/dsj-2018-019

Appendix

Information collected about each of the policies:

1. Title of policy [Short text]
2. Link for policy [URL]
3. Organisation/body [Short text]
4. Year policy was introduced [yyyy; NA]
5. Date of current version [yyyy; NA]
6. The policy describes responsibilities of a data steward [Yes; NA]

    7.2 The responsibilities are captured in a well-defined list or section [Yes; NA]

    7.3 If 6.1 is true, copy of the text of the responsibilities [Text]

    7.4 The responsibilities of the data steward include data management in general (or overseeing research data) [Clear Language; Vague Language; NA;]

    7.5 The responsibilities of the data steward include data retention and/or preservation [Clear Language; Vague Language; NA]

    7.6 The responsibilities of the data steward include complying with terms of programmes/sponsored project agreements/contracts [Clear Language; Vague Language; NA]

    7.7 The responsibilities of the data steward include complying with applicable state and federal laws and regulations [Clear Language; Vague Language; NA]

    7.8 The responsibilities of the data steward include complying with data protection regulations (human subjects, backups, data security and access, etc.) [Clear Language; Vague Language; NA]

    7.9 The responsibilities of the data steward include protecting rights (e.g., intellectual property, misuse of scientific record, etc.) [Clear Language; Vague Language; NA]

    7.10 The responsibilities of the data steward include making decisions regarding data access [Clear Language; Vague Language; NA]

    7.11 The responsibilities of the data steward include giving access to research data to appropriate parties [Clear Language; Vague Language; NA]

    7.12 The responsibilities of the data steward include educating researchers and/or participants regarding research data management [Clear Language; Vague Language; NA]

    7.13 The responsibilities of the data steward include ensuring accuracy and authenticity of research data [Clear Language; Vague Language; NA]

    7.14 The responsibilities of the data steward include assisting the university in either investigation of charges (e.g., misconduct or conflict of interest) or resolving disputes arising from research [Clear Language; Vague Language; NA]

7. Describes data management responsibilities by the university [Yes; NA]

    7.1 The university responsibilities are captured in a well-defined list or section [Yes; NA]

    7.2 If 7.1 is true, copy of the text of the responsibilities [Text]

    7.3 The responsibilities of the university include data management in general (or overseeing research data) [Clear Language; Vague Language; NA]

    7.4 The responsibilities of the university include data retention and/or preservation [Clear Language; Vague Language; NA]

    7.5 The responsibilities of the university include complying with terms of programmes/sponsored project agreements/contracts [Clear Language; Vague Language; NA]

    7.6 The responsibilities of the university include complying with applicable state and federal laws and regulations [Clear Language; Vague Language; NA]

    7.7 The responsibilities of the university include data protection (human subjects, backups, data security and access, etc.) [Clear Language; Vague Language; NA]

    7.8 The responsibilities of the university include protecting the rights of people or entities (e.g., intellectual property, loss, misuse of scientific record, etc.) [Clear Language; Vague Language; NA]

    7.9 The responsibilities of the university include making decisions regarding data access [Clear Language; Vague Language; NA]

    7.10 The responsibilities of the university include giving access to appropriate parties [Clear Language; Vague Language; NA]

    7.11 The responsibilities of the university include educating researchers and/or participants regarding research data management [Clear Language; Vague Language; NA]

    7.12 The responsibilities of the university include providing infrastructure or financial support (e.g., storage or repositories) [Clear Language; Vague Language; NA]

    7.13 The responsibilities of the university include providing support to researchers different than the support captured in 7.11 and 7.12 [Clear Language; Vague Language; NA]

    7.14 The responsibilities of the university include investigating charges (e.g., misconduct or conflict of interest) or resolving disputes arising from research [Clear Language; Vague Language; NA]

8. Policy describes data management as a shared responsibility [Clear Language; Vague Language; NA]

9. Policy assigns responsibilities to students [Clear Language; Vague Language; NA]