





## Are Institutional Research Data Policies in the US Supporting the FAIR Principles? A Content Analysis

Clara Llebot, Oregon State University, Corvallis, OR, US, [clara.llebot@oregonstate.edu](mailto:clara.llebot@oregonstate.edu)   
Diana J. Castillo, Oregon State University, Corvallis, OR, US, [diana.castillo@oregonstate.edu](mailto:diana.castillo@oregonstate.edu) 

### Abstract

**Objective:** The FAIR principles were created with the goal of enhancing the reusability of research data and to give guidance on how to make data Findable, Accessible, Interoperable and Reusable. In this article we explore the role of institutional research data policies in enabling and encouraging researchers at their institutions to generate FAIR data.

**Methods:** We identified the research data policies in place for “very high research activity” institutions (as defined by Carnegie classification) in the United States. We created a list of 31 criteria, based on previous work by Davidson et al. (2019) and Briney et al. (2015), and evaluated the 40 policies using a content analysis methodology.

**Results:** The guiding principles and the definitions for research data in the policies support the idea that institutional policies are a potential tool for the implementation of the FAIR principles. However, our analysis indicates that they are not generally used for that purpose. Only one policy mentions FAIR. Data sharing is mentioned in half of the policies, but 11 of these only note this concept in the context of funder requirements. Access and retention sections are mostly written without considering publicly available data. Twenty-nine policies do not mention data documentation.

**Received:** July 29, 2022 **Accepted:** November 4, 2022 **Published:** February 16, 2023

**Competing Interests:** The authors declare that they have no competing interests.

**Data Availability:** Data associated with this article are available from ScholarsArchive@OSU: Llebot, Clara, and Diana Castillo. 2021. “Research Data Policies at Doctoral Universities with Very High Research Activity (R1 Institutions).” Oregon State University. <https://doi.org/10.7267/4J03D6263>. Llebot, Clara, and Diana Castillo. 2022. “Content Analysis of Institutional Research Data Policies in the US.” Oregon State University. <https://doi.org/10.7267/2N49T893T>.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

See <https://creativecommons.org/licenses/by/4.0>.

∞ OPEN ACCESS

**Conclusions:** We discuss ways in which these institutional policies represent a missed opportunity to implement the FAIR principles and suggest ways policies could be modified to encourage researchers to follow them. We also discuss future research opportunities to examine how policy implementation may affect what institutional support researchers receive.

---

## Introduction

The FAIR principles were described by Wilkinson and collaborators in 2016 (Wilkinson et al. 2016), with the goal of enhancing the reusability of research data. The principles—Findability, Accessibility, Interoperability, and Reusability—are meant to guide data creators and data publishers, and apply not only to data, but to other digital objects such as algorithms, tools, and workflows that exist to create the data. The principles focus on the need for computers (in addition to humans) to be able to find, access, operate, and reuse data. Since the announcement of the FAIR principles in 2016, they have received attention from many stakeholders, including research funders, universities, publishers, libraries, and research infrastructure organizations (Budroni, Claude-Burgelman, and Schouppe 2019).

The scientific community is diligently working to implement the FAIR principles, but there is still a lot of work to do. Part of the challenge is due to the fact that the FAIR principles are guiding principles and, as such, they are aspirational and need to be interpreted by the community (Jacobsen et al. 2020; Wilkinson et al. 2019). There are also varied rates of implementation across disciplines, with a recent study finding that in a pool of 100 randomly selected articles citing the original article detailing the principles, 95% of them were in the natural science fields (van Reisen et al. 2020).

In order for the FAIR principles to be implemented within a community, incentives must be provided. These incentives range from “carrot” incentives, which reward adoption, to “stick” incentives, which impose a threat or obligation. An example of a reward incentive is making funding available to implement the principles in a particular research project. Reward incentives are generally difficult to establish, and as of 2022 the incentive structure is still misaligned, with very few rewards distributed for sharing good quality content or for developing infrastructure that will enable and facilitate this sharing (Peer et al. 2021). “Stick” incentives, represented mainly by policies, have been more clear and have achieved some success. For example, many journals have enabled policies either requiring data access statements or mandating data sharing (Lee 2022). These policies have resulted in an increase in the number of articles being published with available data, although the increase is viewed by many as small and insufficient (Federer et al. 2018; Gorman 2020).

Policies have been used as incentives for the adoption of good data practices and FAIR principles at different levels, with many journals, funders, and institutions implementing data policies. The goals of the journal and funder policies have been generally aligned with the goals of the FAIR principles: making data accessible, documented, and formatted in a way that maximizes its usability to others. In the case of funders, these

practices are designed to maximize the impact of their investment. In the case of journals they indicate that the journal's content reflects reproducible, high quality research. The goals of institutional policies, however, particularly in the United States, generally appear to be disconnected from the FAIR principles. As currently written, these policies reflect the need of research institutions to assert their ownership over research data; to protect the institution's intellectual property; to ensure that the institution will be compliant with different rules, laws, and contracts, such as human subjects research, or funder requirements; and to ensure that the institution retains enough information to be able to act in cases of allegations of research misconduct and the like. However, these institutional policies also attempt to encourage and support researchers to engage in robust data management practices that produce good quality research, and therefore could potentially be a tool for the implementation of the FAIR principles.

The goal of this article is to evaluate whether and to what extent institutional policies are used as a tool to support and encourage FAIR data in research institutions in the United States of America. We defined a series of evaluative criteria, reached out to research institutions to collect their data policies, and performed a content analysis of the research data policies institutions published on or before 2020. A second goal of this article is to describe the content of these institutional data policies, especially the elements of the policies that are related to the FAIR principles. We do not evaluate whether the policies themselves are FAIR, only if they encourage practices that support FAIR data.

## Literature review

Our review of the literature indicates that over the last several years, most of the policy work done to implement data sharing and the FAIR principles in the US has been carried out by journals and funders. In the case of journals, publishers tend to set broad guidelines organized in tiers, and then allow individual journals to implement specific requirements based on which tier they fall into. For example, Taylor & Francis and Springer Nature have created data sharing tiers for their journals to be implemented over a period of time, with Taylor & Francis specifically including “open and FAIR” at its most robust tier (Jones, Grant, and Hrynaszkiewicz 2019). Springer Nature, on the other hand, classified its journals as Types 1-4, with each subsequent policy having more rigorous data sharing requirements. While individual journals might include the FAIR standards in their version of the policy—the journal *Nature*, for example, endorsed FAIR principles in the earth sciences and began requiring data deposits (“Announcement: FAIR Data in Earth Science” 2019)—the principles are not usually specified by name at the publisher level. Funding organizations, as the entities that provide money for research, have the capacity to provide strong incentives to ensure alignment with FAIR principles by including data sharing in their policies. A 2020 research article examining the move towards health research funding organizations requiring FAIR data found that their main strategies have been to integrate data management plans in the research funding cycle and provide guidance and financial support for implementation, but that there needs to be more education by funders about what FAIR data entails, as well as engagement with stakeholders (Bloemers and Montesanti 2020).

In contrast with the extensive work focusing on funder and journal policies, there is very little research on the role of institutional policies to implement the FAIR principles. The most relevant work on the potential that institutional policies have as an instrument for implementation has been the “Turning FAIR into reality” report (European Commission, Directorate-General for Research and Innovation 2018), which examined the current state of the FAIR guidelines and provided recommendations for moving forward. The report talks about the need for a “FAIR ecosystem” and identifies the following essential components of a FAIR ecosystem: data management plans, identifiers, standards, repositories, and policies. Following this report, another document was published by Davidson et al. (2019) analyzing the policy landscape in the European Union, including national, funder, publisher, and institutional policies. The authors created a series of “features” to evaluate policies across three stakeholder groups (higher education institutions, funders, and publishers) and used them to evaluate eleven institutional policies from the perspective of the FAIR principles. Throughout this paper we will refer to this report as the Davidson report. We have used this report extensively as a foundation for the criteria defined in this article and to put the results we find for US policies in context. However, it is important to note that the policies used in the Davidson report represent a relatively high level of engagement with the FAIR principles and may not be representative of the policies across Europe.

The authors found no evidence of any work reviewing how FAIR principles have impacted institutional data policies in the United States. In 2015, Briney, Goben, and Zilinski reviewed the landscape of institutional policies within the United States and provided an overview of their general content (referred to from now on as the Briney study). The Briney study found that 90 out of the 206 universities they analyzed had a data policy (44%). When taking into account only R1 institutions, the percentage grows to 52%. However, the Briney study used a different definition of research data policy than what we use in this work. Their criteria is broader, and their dataset includes intellectual property policies, policies about patents and inventions, and policies that only affect specific types of data, like human subjects data. If we revise the policies identified by the Briney study and include only the policies that fit the goals of the current work (see the methods section), we find that in 2014 26% of the R1 institutions had a research data policy.

This work builds on the criteria developed in the Davidson report and in the Briney study to analyze the content of institutional data policies in the US. Even though the “Turning FAIR into reality” report points to the importance of policies for the implementation of the FAIR principles, the only analysis about this topic on institutional policies is a partial analysis conducted in the Davidson report in the European Union. This article constitutes the first exhaustive analysis of this type for US institutional policies.

## **Methods**

### *Collection of policies*

To create a pool of research data policies for this research project, we created a list of standards the policies had to meet in order to be included. This article only reflects policies that fit these standards:

1. The policy is a standalone policy that exclusively refers to research data. This excludes policies that focus on other types of university data, intellectual property (IP) policies, or data classification policies.
2. The policy is a university-wide official policy, created by university administration. This excludes faculty senate policies, faculty workbooks, etc., and policies from individual departments and colleges.

Data collection occurred in spring and summer of 2020 with the help of two interns. Using the Carnegie Classifications of Institutions for Higher Education, we started with an initial pool of 131 institutions classified as “Doctoral University - Very High Research Output (R1)” in 2020. Using publicly available contact information, the interns emailed research data librarians at each of the R1 institutions asking about their research data policies. If they did not receive a response, they searched the university website for a publicly available policy.

As a quality control mechanism, we compared our list of policies with the list of policies published in 2015 (Briney, Goben, and Zilinski 2015). We use their published data (Briney, Goben, and Zilinski 2015a), as well as the PDF copies of the policies that the authors collected at that time, and that they graciously shared with us. As mentioned in the literature review, the Briney study used different criteria when classifying the policies than the ones used here. We revised the policies identified by the Briney study for the universities that we include in the present work and found 27 instances of policies that they included in their study, but that we are not including here. Throughout this article, when we refer to results from the Briney article, these will reflect the revised dataset, not the original.

The complete list of universities considered in the present article, with links to their policies can be found in Llebot and Castillo (2021).

### *Content analysis*

Data analysis was performed using a content analysis methodology—applying a set of criteria, and then counting how frequently the criteria occur in the policies (Byrne 2016). Content analysis was done through an iterative process. Both authors read each policy and marked in the project spreadsheet whether or not the policy fit the criteria, and if it did, what kind of language was used. Coding was then compared, with discussions and reconciliation when it differed between authors. Finally, results were reviewed for each variable and inconsistent coding was fixed. Once the coding was finalized, a content analysis was conducted on the relevant selections to find variations on themes within criteria and to determine if there were significant differences within them.

## *Criteria*

The criteria set used for the content analysis came from three different sources: (1) the FAIR Policy Landscape Analysis conducted by Davidson et al. in 2019; (2) the analysis conducted by Briney, Goben, and Zilinski in 2015; and (3) several criteria that the authors created for this project. A full list of which criteria were included and excluded from each source, along with the rationale for the decision is published in Llebot and Castillo (2022).

The FAIR Policy Landscape Analysis developed a set of 42 criteria. For this project, 20 of the Davidson et al (2019) criteria were excluded from the final analysis for a variety of reasons ranging from not being within the scope of the project (e.g., criteria 31, relating to the intellectual property of the researchers) to the criteria not being relevant to institutional data policies (e.g., criteria 29, which focuses on data availability statements for journals).

The criteria chosen from the Briney, Goben, and Zilinsky article focused on the general types of content that is included in policies, such as ownership, access, transfer, and description of responsibilities (Briney, Goben, and Zilinsky 2015).

The final criteria were chosen to add key details, like capturing the definition of data used in each policy, or whether the policies mentioned open data principles. This last criteria was chosen specifically because a number of policies were implemented before the creation of the FAIR principles. In this paper, we will only be discussing the criteria that are the most relevant to the promotion of the FAIR principles (see Table 1). The full list of criteria is included in [Appendix A](#). The whole dataset with the results for each criteria can be found in Llebot and Castillo (2022).

## **Results and Discussion**

### *Policy landscape in 2020*

Forty (approximately 30%) of the 131 R1 universities had research data management policies that fit the scope of this project. Of the 40 policies identified in the current study, 30 correspond to universities that already had a research data policy in 2014 according to the Briney study; nine have new policies, and one university was not included in the Briney study. Ninety-one universities have never had a research data management policy as defined in this study during the period 2014-2020. This suggests a growth of new research data policies of around 30% in six years. Of the 30 universities that had a policy in 2014 and in 2020, 18 have the exact same policy and 12 (40%) were updated, ranging from major rewrites to minor changes.

### *Content analysis*

In this section we outline the content of the 40 policies, using the criteria described in section 3.3. The “guiding principles” subsection describes the values the policies claim guided the policy development when explicitly stated. The “definitions” subsection analyzes the definition of research data, and discusses whether

**Table 1:** Subset of criteria reported in this article. See the complete list of criteria in [Appendix A](#).

<b>Guiding principles</b>
<p>Criteria 7: Policy references Open Access to research data</p> <p>Criteria 8: The policy mentions principles that guide the policy. For example, open principles, reproducibility, transparency, or research integrity.</p> <p>Criteria 8.1: Which principles?</p> <p>Criteria 20: Specifically references FAIR</p>
<b>Definitions</b>
<p>Criteria 6: Definition of data provided</p> <p>Criteria 6.1: If definition of data is provided, text of the definition.</p>
<b>Data sharing</b>
<p>Criteria 9: Policy requires data sharing</p> <p>Criteria 10: Policy requires metadata sharing</p> <p>Criteria 11: Exceptions to data sharing are allowed</p> <p>Criteria 12: If exceptions are allowed, justifications are required</p> <p>Criteria 26: The Principal Investigator (PI) and/or data steward decide whether or not to share research data (e.g., in a repository). Policy may list exceptions, like when agreements/terms of sponsorship supersede this right.</p>
<b>Retention and preservation</b>
<p>Criteria 18: Includes minimum length of data availability</p> <p>Criteria 18.1: Years of data availability that are required</p> <p>Criteria 19: Includes specific reference to preservation (mid to longer term)</p> <p>Criteria 21: References specific data repositories or scientific databases for deposit</p>
<b>Documentation</b>
<p>Criteria 29: Requires data documentation.</p>
<b>Access</b>
<p>Criteria 30: Defines who has access to data.</p>

the definitions are consistent with the scope of the FAIR principles. In “data sharing” we outline the different models presented in some of the policies recommending or mandating data sharing, and discuss whether these support FAIR data. In “retention and preservation” we summarize the mandates to retain research data included in these policies, and in “documentation” we look at whether the policies give any details about how metadata about research data should be recorded. Finally, the “access” subsection discusses the overlaps with the “accessible” principle, and summarizes how these policies do or do not dictate research data access.

### *Guiding principles*

Institutional data policies in the US usually focus on issues like data ownership, retention, and access to the data by university officials (Briney, Goben, and Zilinski 2017). In order to confirm that data policies are also concerned with best data management practices, we examine how these policies describe the principles that govern and guide them. Twelve of the 40 policies include these guiding principles. Six mention research integrity and soundness of the research process, with reproducibility specifically mentioned in two of those six. Openness appears in four policies either explicitly or in the context of making the data publicly available. One university mentions the free dissemination of ideas as guiding their policy. While there are other principles mentioned in the policies, these are the ones most relevant to FAIR.

Regardless of the principles included in the policies, we also looked at explicit mentions of open access or openness in the policies. Only five universities meet this criteria, and only one of the policies, from Brown University, explicitly mentions FAIR, saying “The deposits in the [Brown Digital Repository] endeavor to align with FAIR Principles (Findable, Accessible, Interoperable, Reusable) for making data machine-actionable.” These are low numbers compared with the 10 out of 11 policies that mention openness and the five that mention FAIR in the Davidson report.

### *Definitions*

To make sure that institutional data policies are a good instrument for the implementation of the FAIR principles, it is important to check that both refer to the same type of data. Shared vocabulary is key for establishing a shared culture and enforceable action across an institution.

Most policies (38 out of 40) include a definition of “research data.” Many of the policies borrow or modify language from other sources. Twelve policies use, directly or with variations, language from The Code of Federal Regulations in CFR 200.315 (e) (3), which defines data as “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.” Nineteen of the policies define data as “records that are necessary for the reconstruction and evaluation of reported results of research and the events and processes leading to those results.” We have not been able to track the origin of this definition; it may be from a source that we have not found, or just a definition originated in a data policy that has been copied in others. Two policies use the exact definition from the NIH Grants Policy Statement, which says that “data” is “recorded information, regardless of the form or media on which it may be recorded,



and includes writings, films, sound recordings, pictorial reproductions, drawings, designs, or other graphic representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing or computer programs (software), statistical records, and other technical research data.” Although only two policies in our sample use this exact definition, 28 contain the concept that research data can be recorded in any form or medium.

These broad definitions of research data fit with the intent of the FAIR principles that, according to Wilkinson et al. 2016 “apply not only to ‘data’ in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles.”

Several of the data definitions in our policies make clear that these policies apply not only to digital objects, but also to physical objects. Seven universities explicitly include both tangible and intangible research data. Even though the FAIR principles are generally applied to digital objects, there are examples of application of the FAIR principles to physical objects, like the work by Lannom, Koureas, and Hardisty (2020) to make natural science collections FAIR.

### *Data sharing*

While FAIR data is not necessarily open data, sharing data and/or metadata is an important element to meet the FAIR principles for many datasets. Researchers sharing their data, either by choice or due to journal, funder, or institutional mandates, create an ecosystem where data can be findable, accessible and reusable by others, provided that the sharing is done in the right formats, repositories, with the right documentation, etc.

How institutional research data policies include and discuss data sharing can provide additional insight into whether or not they encourage practices that lead to FAIR data. Of the 40 policies analyzed, about half (21) regulate data sharing in some way. In this group, three of these policies make data sharing mandatory, four encourage researchers to share data but don’t require it, and 14 explicitly state that data sharing is allowed. These policies talk about the sharing of research data, but only once mention metadata or data documentation in this context.

The three policies that mandate data sharing do it very differently. The University of Massachusetts Amherst makes it very clear that research data must be shared publicly as soon as possible “in a useful form,” although some delays and exceptions are acceptable in certain circumstances. The university includes details on how researchers should handle confidential data in these circumstances. Yale University’s policy has a similar message, saying that data sharing is essential to the university’s values and increasingly required by funders. It also states that “Research data and materials shall be made publicly available to the extent feasible while minimizing harm to the legitimate interests of the University, to the research subjects, and to other parties.” Finally, the University of Louisville states that all research data and protocols “shall be unrestricted as to its

public dissemination.” In addition, this policy says that the University will not accept research projects that restrict access to the data or dissemination of results and regulate situations in which the dissemination of data may be delayed. It is worth mentioning that Colorado State University does not require data sharing in general, but it mandates the sharing of research data associated with dissertations conducted at the university.

The four universities that encourage data sharing justify it by explaining that the responsible sharing of data is in accordance with the university mission (Brown University, Colorado State University, and University of Tennessee Knoxville) or by considering data sharing a “right and responsibility” of the principal investigator (Ohio State University). Most of the universities that explicitly allow data sharing do it in the context of funder requirements, explaining that these data sharing requirements exist, and that researchers should comply with all public access requirements. In this sense, these 11 universities are only allowing data sharing for data that is subject to funder requirements. Three of the policies in the pool consider data sharing a right of the researcher, and two of the policies explain that public access to research data is part of the values of the university.

Of the 18 policies that encourage or allow data sharing, 11 explicitly state that the principal investigator and/or data steward decide whether or not to share research data. Some policies list exceptions, like when agreements/terms of sponsorship supersede this right. Other universities (University of New Hampshire and Virginia Commonwealth University) allow data sharing but require the researchers to formalize this sharing. The universities that allow or encourage data sharing sometimes clarify that there may be reasons why the data cannot be shared, such as when there are privacy concerns, licensing restrictions, or the potential for intellectual property development. The three universities that mandate data sharing list similar exceptions.

Six universities mention repositories (institutional or external) as a place where researchers can choose to deposit their data. A few policies include guidance and explain good practices about where to deposit research data for sharing and retention. For example, Brown University’s policy states:

“Brown recommends that researchers submit data to an established digital archive or repository whenever possible and offers the Brown Digital Repository (BDR) as an option to meet researchers’ needs for retaining digital copies of their data and laboratory notebooks and/or as a complementary option for local access to data deposited in external or national repositories. [...] While personal or lab websites, ELNs, wikis, and similar tools may be sufficient for the short term, Brown does not recommend them as long-term data retention or sharing solutions.”

Overall, data sharing does not seem to be a key element in institutional data policies, although it is mentioned frequently, and we found different models of how this data sharing can be implemented in a policy. In general, universities are content to provide researchers with the latitude to make decisions about data sharing unless otherwise required by a funder. For the universities that either require or encourage researchers to share their data, many of them frame it as being in line with institutional values around research integrity and ensuring public access to data, which aligns better with the FAIR principles than those that allow it only when it is required by a funder. The lack of guidance on sharing metadata is also a lost opportunity of US policies to encourage the FAIR principles. Our results contrast with the Davidson report, which found that out of 11 policies, five suggested data sharing, three required it, and only three did not mention it. In addition, seven of these required the sharing of metadata, while only one university in our pool encouraged researchers to share their metadata.

### *Retention and preservation*

The FAIR principles do not explicitly mention retention and preservation, but it is implied in several of them. For example, the accessibility principle “A2: Metadata should be accessible even when the data is no longer available” implies that metadata must be preserved. Retention is a very important part of institutional research data policies because it is tightly related to many of their goals, like protecting the university’s intellectual property, or ensuring that the institution has information to be able to act in cases of allegations of research misconduct. All the policies analyzed in our sample talk about preservation in some form. Most of them (35) regulate a minimum length of data retention between three and seven years.

Many policies list exceptions that extend the retention period. The most common is an extension of the retention period when there are questions about the research, such as claims, audits, or charges against the researchers like allegation of research misconduct or conflict of interest (28), to protect intellectual property (26), when required by other contractual obligations (20), or when there are laws or regulations that govern the use of the data such as research related to human subjects (8) or FDA applications (4). Additionally, retention may be extended when students participated in the work (22), either to when the degree is awarded or when it is clear that the work has been abandoned.

Retention is always discussed in the context of keeping data accessible to the university. While this type of retention could be part of a FAIR strategy for retaining data for the purposes of reusability, it is not aligned with the FAIR principles as they are currently structured. About half of the policies describe how to retain the data, with 12 mentioning retaining original data wherever possible, and others referring to retaining data in a “durable form,” a “secure location,” or “using discipline standards.”

In general, institutional policies do not connect retaining data with data repositories as a method of preservation. The majority of policies do not mention repositories, and when they do (seven of the policies

did), it appears as a clarification of a place where researchers can choose to share their data, not to preserve it.

### *Documentation*

Metadata is a crucial part of the FAIR principles, with 13 of the 15 sub principles mentioning it. Of these, there are at least two that are directly related to research data as defined in institutional policies and that fit within this criteria: “F2. Data are described with rich metadata” and “R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.” In contrast, documentation is rarely mentioned in policies. Only 11 policies include language around documentation of research data. Ten of these universities specifically describe their expectations for documenting research, with six of those putting documentation as one of the duties of the researcher and/or principal investigator. Five policies discuss how the data should be documented in specific formats, such as bound notebooks or electronic research notebooks. This lack of focus on metadata can be explained by the fact that data documentation is in many cases included in the definition of research data. Furthermore, while institutional policies in general do not focus on data sharing, the absence of metadata and metadata sharing is worth noting. These are all missed opportunities from institutions to engage researchers in the creation of FAIR data.

### *Access*

All universities in the pool included sections discussing access, but they are primarily concerned with who at the university has access. For example, half of the universities included language about when the university can take control of the data to ensure access in the event there are research integrity questions or audits. Additionally, 11 policies contain language centered on ensuring university access to data in the event the lead researcher leaves the institution. Only two universities, the University of Massachusetts Amherst and Yale University, contain language about public access to data. In their policy, Yale frames the public access component as both complying with funder requirements, as well as tying into their data sharing components. The University of Massachusetts Amherst also ties public access to their data sharing expectations but states that data should be made available earlier if the public would benefit from it. Overall, similarly to what happens with descriptions of retention, the sections about data access show a missed opportunity for institutions to discuss how outside researchers can access research data and to point to the FAIR principles to ensure this access is robust.

## **Conclusions**

Are institutional policies used as a tool to support and encourage FAIR data in research institutions in the United States? Our results indicate that data policies are, generally, not used for this purpose. They are also not generally used as a tool to encourage other principles such as openness, even when policies often mention open access, replicability and reproducibility of research, high quality research, academic integrity, and outstanding research and integrity as guiding principles.

Our results also point to important differences between US policies and the policies represented by the Davidson report. These differences reflect both methodological differences (a systematic collection of policies in this study versus a collection biased towards institutions engaged in open practices in the Davidson study) but also contextual differences. As Briney, Goben, and Zilinski (2017) note, in the United States grants are given to the university to administer instead of to the researcher directly, and therefore the university is responsible for compliance requirements. This explains the focus of US institutional policies on compliance and ownership, which contrasts with the focus on open practices by institutional policies in Europe.

An important conclusion of this work is that these policies represent missed opportunities. There are several sections of these policies that could include concepts from the FAIR principles but that often do not. We offer the following suggestions as ways to address these missed opportunities in a way that would not require a major restructure of the policies.

- Include goals and/or principles that emphasize why researchers should be following best practices. This opens the door to include guidance on topics like FAIR.
- Mention the FAIR principles in the policy, in particular when talking about making data and metadata openly available.
- Clearly state that researchers have the right to make their data openly available, not just when a funder mandates it. Because these policies include lengthy segments on data ownership and procedures for when a researcher leaves, a researcher not well versed in intellectual property may doubt whether they are allowed to share data unless explicitly allowed.
- Explicitly mention repositories as a place where data can be shared, retained, and preserved for the purposes of the university and for public access.
- Include public access when talking about data access. This addition emphasizes that researchers should be thinking broadly when it comes to making data accessible.
- Explicitly include data documentation under researcher responsibilities and emphasize that they should follow discipline-specific best practices and standards.
- When discussing data sharing, include the stipulation that metadata must also be shared.

Institutional policies that incorporate the FAIR principles legitimate them and may open additional adoption methods, such as funding. We think that incorporating topics like the ones suggested in these conclusions can be beneficial, although we acknowledge that there are many other ways that a university can support their researchers in implementing these principles that can be more effective. The participants

in the Davidson report focus groups noted that policies from funders and publishers are more influential than institutional policies, and that they felt that “institutional data policies may be viewed more positively by researchers if the HEI (Higher Education Institution) was not the one making demands but instead focusing on supporting researchers in adhering to external mandates.” Some examples of this support could be creating infrastructure, recognizing the creation of FAIR data as a valued activity for researchers, or having research support personnel with the expertise and capacity for giving specific, applied support and guidance on the FAIR principles. For disciplines, like the ones identified by van Reisen et al. (2020), where there is little uptake on FAIR principles implementation, institutional data policies may be helpful. These policies, because they apply to all researchers at an institution, can lay the groundwork for more widespread understanding and implementation.

Institutional research data policies represent an important starting point in examining what a university prioritizes when it comes to their research. But a policy affirming the FAIR principles or encouraging researchers to create FAIR data does little good if institutions do not create an infrastructure that supports researchers in complying with their own policies. Implementation of the FAIR principles should not fall only onto the researcher’s shoulders; research institutions should be included in this responsibility. Future research examining the implementation of the FAIR principles should focus on how these policies are being put into practice, and whether or not there are other ways universities are doing this work. A review of the distribution of responsibilities between researchers and institutions in these data policies would also provide insight into what institutions are responsible for, and what additional support could be provided to researchers.

---

## Acknowledgements

We thank the anonymous reviewers who gave their time to review this article. We thank our colleagues at Oregon State University, in particular Hannah Rempel, for their prompt and thorough review of earlier drafts of this manuscript. We thank Kristin Briney, Abigail Goblen, and Lisa Zilinski for sharing with us the policies that they found in their 2015 article, which allowed us to quality control our dataset. Thanks to Jennifer Martin and Gloria Park for their contributions collecting data for this work.

The content of this article is based on the Lightning Talk presentation at RDAP Summit 2022, available from Open Science Framework: <https://osf.io/vfajt>.

## Data Availability

Data associated with this article are available from ScholarsArchive@OSU:

Llebot, Clara, and Diana Castillo. 2021. “Research Data Policies at Doctoral Universities with Very High Research Activity (R1 Institutions).” Oregon State University. <https://doi.org/10.7267/4J03D6263>.

Llebot, Clara, and Diana Castillo. 2022. "Content Analysis of Institutional Research Data Policies in the US." Oregon State University. <https://doi.org/10.7267/2N49T893T>.

### Supplementary Files

Appendix A: Full criteria set used for the content analysis. <https://doi.org/10.7191/jeslib.614.s830>.

### Competing Interests

The authors declare that they have no competing interests.

### References

- "Announcement: FAIR Data in Earth Science." 2019. *Nature* 565(7738): 134–134. <https://doi.org/10.1038/d41586-019-00075-3>.
- Bloemers, Margreet, and Annalisa Montesanti. 2020. "The FAIR Funding Model: Providing a Framework for Research Funders to Drive the Transition toward FAIR Data Management and Stewardship Practices." *Data Intelligence* 2(1–2): 171–180. [https://doi.org/10.1162/dint\\_a\\_00039](https://doi.org/10.1162/dint_a_00039).
- Briney, Kristin, Abigail Goben, and Lisa Zilinski. 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): 1232. <https://doi.org/10.7710/2162-3309.1232>.
- Briney, Kristin, Abigail Goben, and Lisa Zilinski. 2015a. "Data from: Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies." [Data set and data dictionary]. Harvard Dataverse. <https://doi.org/10.7910/DVN/GAZPAJ>.
- Briney, Kristin, Abigail Goben, and Lisa Zilinski. 2017. "Institutional, Funder, and Journal Data Policies." In *Curating Research Data Volume One: Practical Strategies for Your Digital Repository*, edited by Lisa R. Johnston, 61–78. Chicago, Illinois: Association of College and Research Libraries.
- Budroni, Paolo, Jean Claude-Burgelman, and Michel Schouppe. 2019. "Architectures of Knowledge: The European Open Science Cloud." *ABI Technik* 39(2): 130–141. <https://doi.org/10.1515/abitech-2019-2006>.
- Byrne, David. 2016. "Data Analysis and Interpretation." In *Research Project Planner*, by David Byrne. United Kingdom: SAGE Publications, Inc. <https://doi.org/10.4135/9781526408570>.
- Davidson, Joy, Claudia Engelhardt, Vanessa Proudman, Lennart Stoy, and Angus Whyte. 2019. "D3.1 FAIR Policy Landscape Analysis." FAIRsFAIR. <https://doi.org/10.5281/zenodo.5537032>.
- European Commission, Directorate-General for Research and Innovation. 2018. *Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. Publications Office. <https://doi.org/10.2777/1524>.
- Federer, Lisa M., Christopher W. Belter, Douglas J. Joubert, Alicia Livinski, Ya-Ling Lu, Lissa N. Snyders, and Holly Thompson. 2018. "Data Sharing in PLOS ONE: An Analysis of Data Availability Statements." *PLoS ONE* 13(5): e0194768. <https://doi.org/10.1371/journal.pone.0194768>.
- Gorman, Dennis M. 2020. "Availability of Research Data in High-Impact Addiction Journals with Data Sharing Policies." *Science and Engineering Ethics* 26(3): 1625–3162. <https://doi.org/10.1007/s11948-020-00203-7>.

"Intangible property." Code of Federal Regulations, title 2 (2022):122-123.

<https://www.govinfo.gov/app/details/CFR-2022-title2-vol1/CFR-2022-title2-vol1-sec200-315>.

Jacobsen, Annika, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtot, et al. 2020. "FAIR Principles: Interpretations and Implementation Considerations." *Data Intelligence* 2(1–2): 10–29. [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024).

Jones, Leila, Rebecca Grant, and Iain Hrynaszkiewicz. 2019. "Implementing Publisher Policies That Inform, Support and Encourage Authors to Share Data: Two Case Studies." *Insights* 32(1): 11. <https://doi.org/10.1629/uksg.463>.

Lannom, Larry, Dimitris Koureas, and Alex R. Hardisty. 2020. "FAIR Data and Services in Biodiversity Science and Geoscience." *Data Intelligence* 2(1–2): 122–130. [https://doi.org/10.1162/dint\\_a\\_00034](https://doi.org/10.1162/dint_a_00034).

Lee, Jian-Sin. 2022. "Setting up a Checkpoint for Research on the Prevalence of Journal Data Policies: A Systematic Review." In *Information for a Better World: Shaping the Global Future*, edited by Malte Smits, 13192:100–121. Lecture Notes in Computer Science. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-96957-8\\_11](https://doi.org/10.1007/978-3-030-96957-8_11).

Llebot, Clara, and Diana Castillo. 2021. "Research Data Policies at Doctoral Universities with Very High Research Activity (R1 Institutions)." Oregon State University. <https://doi.org/10.7267/4J03D6263>.

Llebot, Clara, and Diana Castillo. 2022. "Content Analysis of Institutional Research Data Policies in the US." Oregon State University. <https://doi.org/10.7267/2N49T893T>.

"NIH Grants Policy Statement | grants.nih.gov." 2021. Accessed July 29, 2022.

<https://grants.nih.gov/policy/nihgps/index.htm>.

Peer, Limor, Florio Arguillas, Tom Honeyman, Nadica Miljković, Karsten Peters-Von Gehlen, and CURE-FAIR WG Subgroup 3. 2021. "Challenges of Curating for Reproducible and FAIR Research Output." Research Data Alliance. <https://doi.org/10.15497/RDA00063>.

van Reisen, Mirjam, Mia Stokmans, Mariam Basajja, Antony Otieno Ong'ayo, Christine Kirkpatrick, and Barend Mons. 2020. "Towards the Tipping Point for FAIR Implementation." *Data Intelligence* 2(1–2): 264–275. [https://doi.org/10.1162/dint\\_a\\_00049](https://doi.org/10.1162/dint_a_00049).

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3(1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Wilkinson, Mark D., Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, et al. 2019. "Evaluating FAIR Maturity through a Scalable, Automated, Community-Governed Framework." *Scientific Data* 6(1): 174. <https://doi.org/10.1038/s41597-019-0184-5>.