



# The Anatomy of Data Sharing Locations

## Objective

Learn the benefits of sharing data and when, how, and where to share your data.

## Introduction

Data location is the place where either raw or processed data can be accessed. Ideally, your data location is one that can be accessed without permissions and without the threat of broken links (e.g., using a PID instead of a URL leading to GitHub).

Even when articles have data availability statements (DAS), their data are not always easily accessible. Where and how authors make their data available strongly influences how easy they are to access. Furthermore, research shows that data location can also serve as a proxy for the completeness of the data; for instance, full data sets are more likely to be available when they are shared in external repositories or upon request rather than when they are made available in the article or supplemental files. Therefore, Ripeta believes it is very important to track how authors make their data available, not just whether or not authors have included data availability statements. Particularly, we see if authors have stored their data in external repositories, as data tend to be both easier to access and more complete when they are stored in this way. Currently, we have ranked locations in the following order: 1) *external repository*, 2) *in article and supporting files*, 3) *within paper*, 4) *not publicly available*, 5) *upon request*, and 6) *not applicable*.

We have gone in-depth about data availability statements (DAS) and data citations -- both of which require publicly accessible data, but what does sharing your data mean?

## Why/When share your data?


At ripeta, our mission is to build trust in science by making better science easier. Trust in science is in part related to the transparency of data collection -- both the process and the results. Data sharing also promotes accountability and collaboration among scientists which, in turn, can yield a stronger study overall.

The FAIR guiding principle for data sharing is that your data should be: **findable, accessible, interoperable, and reusable**. For more information on these principles, [visit the FAIR website](#).

## When Not to Share Your Data

Data is a powerful resource, and there are cases where sharing data would be inappropriate and potentially dangerous to individuals and communities. Data that should not be shared includes: sensitive, commercial, and proprietary information, as well as datasets which include personal identifiers.

One crucial mechanism to protect data is that of the [CARE Principles for Indigenous Data Governance](#) created by the [Global Indigenous Data Alliance](#). The Global Indigenous Data



Alliance developed these principles as a way to control data about their peoples, lands, and resources. Many of the current open data movements do not take into account the effect that data sharing can have on indigenous sovereignty, and ignore the potential for carelessness and misuse of sensitive data. Read more about the CARE Principles, as well as their alliance with FAIR to create “Be FAIR and CARE,” here:

- <http://doi.org/10.5334/dsj-2020-043>
- <https://doi.org/10.1038/s41597-021-00892-0>
- <https://www.gida-global.org/care>

Ripeta recognizes that there are legitimate reasons for authors to restrict access to certain data, particularly protected health information. We are therefore training our software to identify when authors have stated reasons for their data restrictions. This improvement will allow us to differentiate between legitimate and arbitrary data restrictions, and we will be able to rank papers more highly when they have provided a reason for their restrictions.

## How & Where to Share Your Data?

There are options for choosing how and where to share your data. While there are some basic questions you can ask yourself each dataset is different and each location type should be taken into consideration.

The Qualitative Data Repository and the Social Science Research Council has compiled a set of use questions for researchers to consider when determining where to share their data. For additional information, please see: <https://managing-qualitative-data.org/modules/3/d/>.

- How much does it cost to deposit your data in the different venue choices?
- Are you going to want direction and assistance with depositing data – is it important to you to be able to interact face-to-face with the people who are handling your data? Or would you rather be able to post your data quickly and efficiently without interacting much with repository personnel, answering questions about curation, etc.?
- How findable and accessible do you want your data to be for other scholars?
- How much do you want and need your data to be carefully curated?
- How sensitive are your data?
- Will you need to place some kinds of access controls on your data?
- How well-described and automated do you want the methods for accessing your data to be?
- To what types of scholars do the different venue choices seem to cater, and does that match the audience with whom you would like to share your data?
- Will other scholars who wish to access your data be charged?
- How easily and effectively do you want other scholars to be able to reuse your data?
- What kinds of reputations do the different venue choices have?
- How much do you trust the different venue choices to keep your data safe?
- Does the venue have Core Trust Seal certification?
- Can the venue issue DOIs? (Social Science Research Council, 2021)

## Types of Data Location

### External Repository

A repository is a location that stores, organizes, allows access to, and preserves data. Common repositories are Dryad Digital Repository, Figshare, Harvard Dataverse, and Zenodo.

What does a DAS look like when it is referencing a repository or online location?

For cases in which data is shared through one or more repositories, the data location will include keywords of [deposited](#) and [database](#). Generally, the [accession number](#) and/or [link for said repository](#) should be stated as well.

Note: Though a paper may state that further information/data is located in supplementary files online, it is important to check what information is shared (unless it specifically states which datasets are available).

#### Repository

All raw sequencing data and ancillary analyses are [deposited](#) in the GEO [database](#) under the [accession number](#) GSE94518.


#### Online

All statistical data are deposited in Supplementary Table [1](#), which is available at the [journal's website](#).

Repositories are the most common and reliable way to store your data. If you are able, it is recommended you use whatever repository fits your data best. Reference the chart below for the different types of publicly accessible repositories:

Repository types for public access data sharing			
Repository	Example	Pro	Con
Domain/discipline-specific data repository, data center or 'scientific database' (DR)	<a href="#">BioFresh</a> , <a href="#">ArcGIS Online</a>	Most likely to <b>offer both the specialist domain knowledge and data management expertise</b> needed to ensure your data collection is properly kept and used (Whyte 2015).	Most likely to be <b>selective</b> , requiring advance planning of the effort needed to meet high standards for metadata and documentation ( <i>ibid</i> ).
General-purpose or Open Research (OR) data repository	<a href="#">Dryad</a> , <a href="#">Figshare</a> , <a href="#">Zenodo</a>	Most likely to offer <b>useful search, navigation, and visualization</b> functionality ( <i>ibid</i> ).	Requires <b>scrutiny</b> of terms and conditions to ensure consistency with your funder, journal, or institution's policies on cost recovery, copyright/IP, long-term preservation ( <i>ibid</i> ).
Institutional data repository (IR)	<a href="#">Texas Data Repository</a>	Most likely to <b>accept any data of value</b> , especially if no suitable home can be found for it elsewhere, and to ensure that <b>policy requirements for long-term</b> access are met ( <i>ibid</i> ).	<b>Unlikely to be as well-resourced</b> as either general-purpose or domain repositories ( <i>ibid</i> ).

Source: [Digital Curation Center](#), [Stanford Libraries](#)



Tip: [re3data.org](https://re3data.org) is a great resource for figuring out which repository fits your research best!

## **Article and Supporting Files**

This is when the paper states that data is accessible in the article itself and its supporting files.

What does a DAS look like when it is referencing a repository or online location?

Contains a statement along the lines of “all relevant data can be found...”

- “...in published article (and its Supplementary Information files)”
- “...within the paper and its Supporting Information files”

## **Within paper**

The data location type in which data is stated to be accessible within the paper itself. This would not include situations in which a paper states that data is accessible within the paper and supporting files.

What does a DAS look like when it is referencing a repository or online location?

Similar to “In article & supplementary files” but qualifies as a separate data location type as it does not include supplementary files and is thus less robust.

## **Not Publicly Available**

This is when the authors claim that data is not publicly available due to circumstances such as patient confidentiality, or may simply state data is not publicly available with no further explanation.

What does a DAS look like when it is referencing a repository or online location?

This is consistently stated as in the two examples above following the pattern of “datasets (generated and/or) analyzed during/in the current study are not publicly available due to...”.

## **Upon Request**

This location type is the infamous, ‘data is available upon request to the authors.’ Occasionally, a paper may state the data is available upon request to an organization, such as the hospital from which data was acquired; either way, the data is available “upon request.”

What does a DAS look like when it is referencing a repository or online location?

This data location type will always contain the phrase “[Upon \(reasonable\) request.](#)”

## **Not applicable**

This is when the authors claim that data sharing is not applicable to their study.

What does a DAS look like when it is referencing a repository or online location?

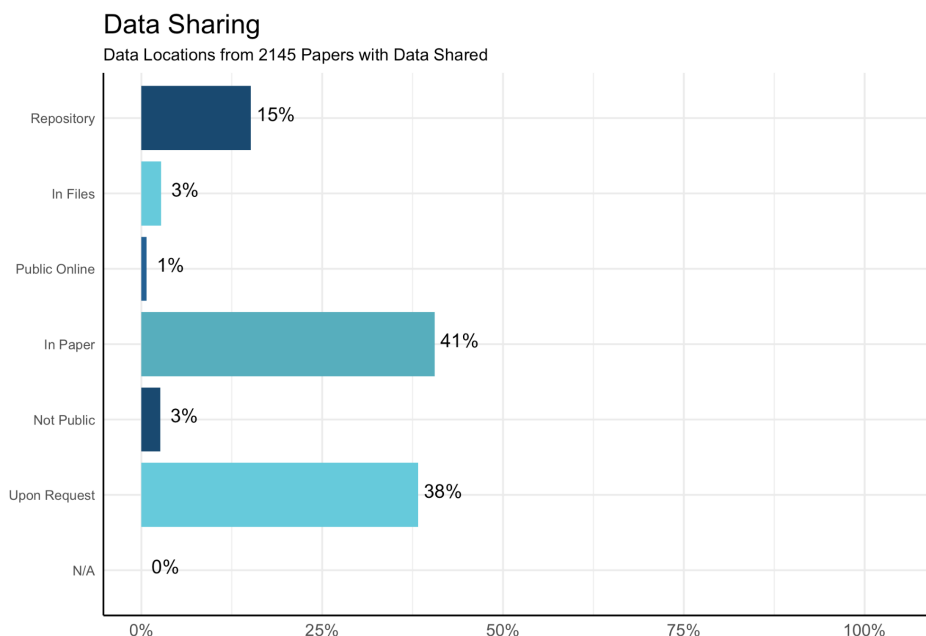
N/A is the least favorable of buckets but easy to identify. Often a paper may have a DAS heading but just put N/A in this section because the journal required a DAS. Though it is rare

that a reason is given, this data location type is divided into “reason given” and “no reason” as there are some papers in which datasets were not generated simply due to the type of study.

## State of Data Location

Our team recently sampled approximately 5000 papers, 2145 of which included a specific data location in their Data Availability Statements. The results show that the majority of data is shared within the Paper (41%) or Upon Request (38%).

We recommend, however, that researchers instead choose a repository for maximum reproducibility. Repositories are a secure method for sharing and storing data, and allow users to access data with little ease.



Funding agencies, institutions, publishers, and researchers can gain valuable insights by understanding where and how researchers share their data. We know that ethical, broad sharing of research data increases reuse, upholds scientific integrity, and accelerates the impact of scientific results. If you are interested in learning more about how Ripeta can provide you insight into the responsible reporting of research within your organization, across your portfolio, or within a corpus of manuscripts, please reach us at: [info@ripeta.com](mailto:info@ripeta.com).



## Works Cited

- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19(1), 43. DOI: <http://doi.org/10.5334/dsj-2020-043>
- FAIR Principles. (2021). GOFAIR. Accessed Jul 1: <https://www.go-fair.org/fair-principles/>.
- When, How, and Where to Share. (n.d.). *Social Science Research Council*. Accessed Jun 20: <https://managing-qualitative-data.org/modules/3/d/>.
- Whyte, A. (2015). 'Where to keep research data: DCC checklist for evaluating data repositories' v.1.1 Edinburgh: Digital Curation Centre. Available online: [www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)