Selecting a Data Repository: Ensuring FAIRness and Compliance with the NIH DMS Policy

Now that the National Institutes of Health (NIH) Data Management and Sharing (DMS) policy has been in effect since January 25, 2023, one of the requirements is to select a repository to share that data. The NIH emphasizes the importance of using high-quality data repositories to uphold the Findable, Accessible, Interoperable, and Re-usable (FAIR) principles of data management and sharing, and the NIH provides guidelines and considerations for researchers to select appropriate repositories, which helps to promote responsible data sharing practices.

The NIH encourages researchers to prioritize established repositories when preserving and sharing scientific data. While the NIH supports numerous repositories, it does not endorse or require the use of specific repositories. However, for certain programs and data types, the NIH or other institutions may require the use of specific repositories for data preservation and sharing. When no specific repository is designated, the NIH states the primary consideration should be given to repositories that are discipline or data-type specific to facilitate effective data discovery and reuse. The NIH provides a list of discipline-specific repositories, and researchers are also encouraged to explore the list of NIH-supported scientific data repositories for guidance. If a discipline-specific repository is not available, researchers should explore other options such as generalist repositories or institutional repositories, and for large datasets, consider using a cloud-based repository. The NIH provides a list of generalist repositories that accept all data types, which offers researchers flexibility in choosing repositories based on their unique project requirements. The overarching goal is to ensure that data remain accessible and reusable by the broader research community.
Here are several desirable characteristics you can use to evaluate a data repository:

- **Unique Persistent Identifiers:** The repository should assign citable, unique persistent identifiers (e.g., DOIs or accession numbers) to datasets, enabling proper data discovery and citation.
- **Long-Term Sustainability:** A robust plan for long-term data management, including maintaining integrity and availability, is crucial for ensuring the sustained impact of the data.
- **Metadata and Curation:** Repositories should provide detailed metadata to enable effective discovery, reuse, and citation of datasets, with domain-specific repositories offering more in-depth metadata.
- **Free and Easy Access:** Repositories should offer broad, equitable, and open access to datasets and metadata, adhering to legal and ethical considerations.
- **Security and Confidentiality:** Documented security measures should be in place to prevent unauthorized access or release of sensitive data, with confidentiality safeguards for protecting participant information.
- **Provenance and Retention Policy:** Repositories should record the origin and modifications of datasets and clearly document policies for data retention.

For human data, additional considerations include fidelity to consent, compliance with use restrictions, privacy measures, breach response plans, download controls, and a transparent process for reviewing data access requests. The Office of Sponsored Programs can help you comply with MD Anderson policies that govern the protection of human data.

Selecting a data repository can be challenging because it involves careful consideration of disciplinary relevance, adherence to FAIR principles, and compliance with NIH guidelines. However, the Research Medical Library can help to ensure the repositories you are considering adhere to the FAIR principles to promote responsible and impactful data sharing practices in the scientific community.
How to Shorten Your Abstract

One of the biggest challenges in writing abstracts is meeting the word limit of the journal or conference. We’ve compiled some tips to help with shortening a too-long abstract.

1. Consider key abstract components and their proportions. An abstract is basically a short version of your article or planned presentation. Abstracts reporting study results should have 4 components, even if they aren’t labeled with subheadings. First, provide brief background, including the hypothesis or purpose of the study; aim for 2–3 sentences for this introductory information. Next, describe your methods in 1–3 sentences. The longest part of the abstract—perhaps 3–4 sentences—should be the results. Finally, state your conclusion and its implications; 1–2 sentences are usually enough. Your specific project may require a little more or less space for each of these components—but if half of your abstract is background information, you know where to start cutting.

2. Cut information that doesn’t contribute to the primary message. The purpose statement and the conclusion should work together to provide the primary message of the abstract. Keep this message in mind as you decide what information to cut. You may not need to include the results of every analysis you did—focus on those relating to or supporting the conclusion. Then focus on the methods used to obtain those included results.

3. Delete excess detail. With so little space, the hardest part of writing an abstract is deciding what amount of detail to include. Don’t waste space in the abstract on incidence and mortality statistics. Keep procedural details brief. Summarize raw data. Save standard limitations, such as a small sample size, for the main article or presentation.

4. Eliminate extra words. Once you’ve reduced the content to just the essential information, look for wordiness that can be eliminated. For example,
“was determined to be significantly longer” can be shortened to “was significantly longer,” and “very effective” can be shortened to just “effective.” (See our webinar on this topic for lots more ideas.) Also, it’s okay in conference abstracts and some journals to use widely understood abbreviations, such as PCR or MRI, without definition.

5. **Take advantage of punctuation and symbols.** In conference abstracts, you can use mathematical symbols to replace words; for example, “more than 20 mice” (4 words) can be shortened to “>20 mice” (2 words). Don’t put a space between mathematical symbols and numerals: Microsoft Word counts “p < 0.05” as 3 words, but “p<0.05” as 1 word. Another tip: Microsoft Word considers words connected by a hyphen to be 1 word. For example, “aged 5 to 13 years” (5 words) can be shortened to “aged 5-13 years” (counts as 3 words). Note: Be sure to use the regular hyphen (−) on your keyboard, because dash symbols (– or —) do count as a word.

6. **Use ChatGPT (with caution).** You might be tempted to use generative AI software to shorten your abstract, which can be done with a prompt such as “Shorten this abstract by about 15%” or “Shorten this abstract to 250 words.” However, there are several important issues and guidelines for using such software. First, unpublished research data and manuscripts are considered confidential, so don’t ever input actual unpublished research data into a generative AI program; make the text generic first. Also, be mindful that many publishers require disclosure if you used generative AI software. Finally, we can’t emphasize enough that the best way to use generative AI results is as inspiration for thoughtfully shortening the abstract yourself. AI can introduce errors, drop key details, or change the intended focus and emphasis of your abstract.

7. **Ask an editor.** Scientific editors in the Research Medical Library can edit and suggest ways to shorten your abstract, usually the same day, free of charge. We humans are better than AI at making sure an abstract is clear and readable for humans! Just email your abstract to RML-Editing@mdanderson.org.
Although "Although" and "Though" Mean the Same Thing, Only One Is Preferred in Scientific Writing

*Although* and *though* are both conjunctions meaning “in spite of the fact that”. The two words are considered interchangeable.

**Examples:**

Although it was raining, we enjoyed our visit to the park.

Though it was raining, we enjoyed our visit to the park.

*Though* is considered more informal than *although*, as it is a shortened form of the word. Therefore, *although* is generally preferred in scientific writing.

**Acceptable:** Though external-beam radiotherapy is a standard treatment approach in patients with prostate cancer, acute side effects can limit its use.

**Preferred:** Although external-beam radiotherapy is a standard treatment approach in patients with prostate cancer, acute side effects can limit its use.

**Acceptable:** Chemotherapy-induced neuropathy, though common in cancer patients, is not well understood.

**Preferred:** Chemotherapy-induced neuropathy, although common in cancer patients, is not well understood.
Though is also an adverb, meaning “however” or “nevertheless”.\textsuperscript{2} When used as an adverb, though is most commonly placed at the end of a clause or sentence, although it can also appear in the middle of the sentence, separated by commas. This use is more common in speech or informal communication. In scientific writing, it is acceptable to use though in an informal opinion or communication piece, but however and nevertheless are better choices in a research article or grant proposal.

Examples:

What the average American is being told about COVID-19 prevention today is quite different from what they were told a few years ago, though.

In our experience, though, the treatment was effective.

**Acceptable:** The patient’s condition improved after surgical excision; further treatment may be required, though.

**Preferred:** The patient’s condition improved after surgical excision; however, further treatment may be required.

Even though is a more emphatic form of although or though that is acceptable in scientific writing.

Some patients continue to use tobacco products, even though the health risks of doing so are well established.

References

A Guide to Using EndNote 20

EndNote is a reference management software package used to manage references. It lets users save references from databases, websites, and files and create citations and bibliographies for documents. The EndNote 20 guide from the Research Medical Library is available with step-by-step instructions as well as recorded class sessions.

EndNote 20 is available in the institution's Software Center. If you have a problem with installation, please contact 4INFO (713-794-4636) and IT will install the software for you.

If you have questions about using EndNote 20, contact us at RML-Help@mdanderson.org and a librarian will be happy to help you.

Native American Heritage Collection

November is Native American Heritage month. We have a curated collection of eBooks and audiobooks available in Libby. If you need help accessing the collection, contact us at RML-Help@mdanderson.org

Visit the library's Education Hub to register for upcoming classes, view helpful videos, or enroll in self-paced courses on scientific writing and literature searching.