



Data Submission Guidelines

A Guide for Elements and their Principal Investigators

<https://nlsp.nasa.gov/explore/lisahome/datasubmit>



This information sheet serves as a practical guide for investigators preparing to archive data with NASA's Life Sciences Data Archive (LSDA). It provides a concise overview of key principles for organizing information, naming files, and structuring datasets to ensure clarity and consistency.



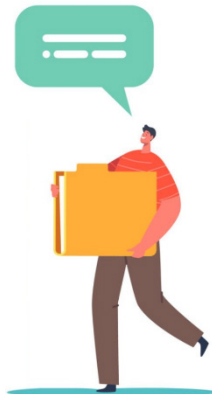
By following these recommendations, investigators can contribute to creating high-quality, standardized, and computable datasets that align with NASA's commitment to making data findable, accessible, interoperable, and reusable (FAIR).



Tidy Data

Well-organized data ensures easy understanding, reproducibility, and analysis. Poorly organized data makes it difficult for Archivists and researchers to understand and use, delaying research and analysis.

- Tabular data should be CSV or other non-proprietary format.
- Organize your dataset with one observation per row and one variable per column.
- Avoid merged cells, empty fields, empty rows, or unnecessary formatting (e.g., colors, borders).
- Ensure all values are labeled consistently and accurately (e.g., use 'N/A' for missing data).
- Double-check that no duplicate rows or redundant data entries are present.
- Ensure consistent units and formats for numeric data (e.g., mm/dd/yyyy for dates).



(Continued next column)



Checksums

Checksums verify the integrity of your files during submission and storage. Without checksums, files may become corrupted or altered without detection, causing issues during archival.

- Generate a checksum for each file to ensure it hasn't been corrupted during transfer.
- Recommended tools: MD5summer or MD5-Checker, R, Python, PowerShell, other computing languages or built-in operating system utilities.
- Include MD5 checksums in the manifest.

Without checksums, issues with files will go undetected.



File Manifest

A file manifest helps others navigate your dataset and ensures nothing is overlooked. An incomplete or unclear file manifest leaves Archivists guessing, increasing processing time and errors.

- Include a detailed list of all files in the dataset, including their names and locations.
- Specify details for each file: name, description, format, size, and purpose.
- Submit the manifest in a plain text format (e.g., CSV or TXT) for readability.
- Include version numbers or dates to identify file updates easily.

(Continued back page)



README

The README file serves as a roadmap to your dataset, explaining its purpose and structure. Without a README file, Archivists may misunderstand the dataset's purpose or structure, leading to delays.

- Suggested sections:
 - **Overview:** Provide a high-level summary of the dataset and its purpose.
 - **Methodology:** Describe data collection, processing, and analysis techniques.
 - **File List:** Detail each file's purpose and contents.
 - **Variables/Metadata:** Explain variable names, units, and formats.
 - **Contact Information:** Provide a way for users to reach you with questions.
- Use plain text or Markdown formats for easy access.
- Highlight any dependencies between files (e.g., scripts and raw data).
- More than one README file may be submitted per submission but only one README should accompany any one type of data.



File Organization / Naming Conventions

Good organization and naming make it easy for others to find and use your files. Disorganized files and inconsistent names confuse Archivists and increase the risk of misfiled data.

- Use clear and consistent folder hierarchies (e.g., 'ProjectName/Data/Raw').
- File names should describe their content and purpose while being concise.
- Avoid spaces, special characters, or overly long names in file naming.
- Examples of good file names:
 - - 'study1_results_2024.csv, - 'metadata_experimentA.txt'
 - Maintain consistent naming patterns across all files and folders.
 - Group related files into clearly labeled folders to avoid confusion.



Avoiding nested zipped files ensures easy access to your data. Nested zipped files slow down Archivists and may cause critical data to be missed during review.

- Nested zipped files (e.g., ZIP within ZIP) complicate access and review.
- Compress datasets only when necessary and provide clear instructions for extraction.
- Always include uncompressed metadata (README, manifest) alongside compressed data.
- Ensure the compression does not corrupt or alter the original data.



Contacts and Resources

- Have questions? jsc-lsda@mail.nasa.gov
- Web: <https://nlspl.nasa.gov/explore/lisahome/datasubmit>
- **Additional Resources** are available to ensure a smooth submission process for you and Archivists.
 - Check the FAQ section on the LSDA submission page for common issues and solutions.
 - Example templates for manifests and README files can be downloaded.
 - Data management tools, tutorials, and videos are also available for free download.