SciTeX Writer: A Container-Based Framework for Reproducible Scientific Manuscript Preparation

First Author[1], Second Author[1], Corresponding Author[1,]

[a]*First Institution, Department, City, Country*
[b]*Second Institution, Department, City, Country*

Abstract

Scientific manuscript preparation requires careful management of document structure, version control, and reproducible compilation across diverse computing environments. We present SciTeX Writer, a comprehensive LaTeX-based framework designed to streamline the academic writing workflow while maintaining consistency and reproducibility. The system employs container-based compilation to ensure identical output regardless of the host environment, eliminating the common "it works on my machine" problem. Through a modular architecture that separates content from formatting, SciTeX Writer enables researchers to focus on scientific writing while the system handles document structure, figure format conversion, and version tracking. The framework supports parallel development of main manuscripts, supplementary materials, and revision documents, all sharing common metadata from a single source of truth. Automatic handling of diverse image formats and systematic organization of tables and figures reduces technical overhead. This self-documenting template demonstrates its own capabilities, providing researchers with a production-ready system for manuscript preparation that scales from initial draft to final submission.

Keywords: keyword one, keyword two, keyword three, keyword four, keyword five

˘ 2 figures, 0 tables, 157 words for abstract, and 2072 words for main

*Corresponding author. Email: corresponding.author@institution.edu

text

1. Introduction

The preparation of scientific manuscripts involves numerous technical challenges that extend beyond the intellectual task of communicating research findings [?]. Researchers must navigate complex typesetting systems, manage multiple document versions, coordinate figures and tables across formats, and ensure reproducible compilation environments [?]. These technical burdens can distract from the primary goal of clear scientific communication and often lead to inconsistencies, formatting errors, and wasted time troubleshooting environment-specific compilation issues.

Traditional approaches to manuscript preparation typically rely on local LaTeX installations, where the specific versions of packages and compilation tools can vary significantly across different machines and over time [?]. This variability creates reproducibility challenges, particularly in collaborative environments where multiple authors work on different systems [?]. Furthermore, the proliferation of image formats and the need to convert between them for different submission requirements adds another layer of complexity. Researchers often resort to ad-hoc scripts or manual processes to handle these conversions, leading to potential errors and inconsistent results.

Existing solutions have addressed some aspects of this problem [?]. Overleaf and similar cloud-based platforms provide consistent compilation environments but require continuous internet connectivity and may not suit all research workflows. Version control systems like Git effectively track changes but require researchers to understand both LaTeX and version control simultaneously. Template repositories exist for various journals, but they typically focus on formatting requirements rather than workflow automation and often duplicate common elements across documents.

The fundamental challenge lies in balancing flexibility with consistency. Researchers need systems that accommodate diverse content types, multiple output documents, and varying journal requirements while maintaining a single source of truth for shared elements like author lists and bibliographies. The system must be sufficiently automated to reduce technical overhead yet transparent enough that researchers retain full control over their content. Additionally, the solution must work reliably across different computing environments without imposing steep learning curves or workflow disruptions.

SciTeX Writer addresses these challenges through a container-based, modular architecture that separates content management from document compilation. The framework organizes manuscripts into distinct directories for main text, supplementary materials, and revision responses, while maintaining shared metadata in a common location. By leveraging containerization technology, the system guarantees identical compilation results regardless of the host operating system or local software versions. Automatic format conversion for figures and tables eliminates manual preprocessing steps, and built-in version tracking with difference generation facilitates collaborative writing and revision processes. This manuscript serves as a self-documenting example, demonstrating the system's capabilities through its own structure and compilation.

2. Methods

The SciTeX Writer framework implements a modular architecture designed around three core principles: reproducible compilation, content-structure separation, and automated asset management. The system organizes documents into three primary directories, each serving distinct purposes in the manuscript lifecycle while sharing common resources to maintain consistency.

2.1. Repository Structure and Organization

The framework employs a hierarchical directory structure where the `00_shared/` directory serves as the single source of truth for metadata including title, author information, keywords, and bibliographic references. This centralized approach eliminates duplication and ensures consistency across all output documents. The `01_manuscript/` directory contains the main manuscript with subdirectories for content sections, figures, and tables. Similarly, `02_supplementary/` follows an identical structure for supplementary materials, while `03_revision/` organizes revision letters by reviewer. Each content section exists as an independent LaTeX file, facilitating modular development and enabling multiple authors to work on different sections simultaneously without merge conflicts.

2.2. Container-Based Compilation System

To ensure reproducible builds across diverse computing environments, the framework leverages both Docker and Singularity containerization technologies [?]. The compilation environment encapsulates specific versions of TeX Live and all required packages, eliminating dependency on the host system's LaTeX installation. Users invoke compilation through a simple Makefile interface that abstracts the container complexity [?]. The command `make manuscript` compiles the main document, while `make all` processes all three document types in parallel. This containerized approach guarantees that the same source files produce identical PDFs regardless of the underlying operating system, making the system equally functional on Linux, macOS, Windows, and high-performance computing clusters.

2.3. Automated Asset Processing

The system implements automatic format conversion for both figures and tables through preprocessing scripts that execute during compilation [?]. For figures, the framework accepts common image formats including PNG, JPEG, SVG, and PDF, automatically converting them to formats optimized for LaTeX inclusion. Each figure resides in its own subdirectory within `01_manuscript/contents/figures/caption_and_media/`, with the caption defined in a corresponding `.tex` file. During compilation, a preprocessing script scans these directories, generates figure inclusion code, and compiles all figures into `FINAL.tex` for inclusion in the main document. Tables follow an analogous structure, allowing authors to define complex table layouts separately from their incorporation into the document flow [?].

2.4. Version Control and Difference Tracking

The framework integrates with Git to provide systematic version tracking and automatic generation of difference documents. When authors create a new version through `make archive`, the system archives the current manuscript with a timestamp and version number. Subsequently, invoking `make diff` generates a PDF highlighting changes between versions using the latexdiff utility. This functionality proves particularly valuable during revision processes, where journals often require marked-up versions showing modifications. The revision directory structure accommodates multiple rounds of review, with separate subdirectories for editor and reviewer responses, each containing both the original comments and author responses in a structured format that ensures complete documentation of the revision process.

3. Results

The SciTeX Writer framework successfully demonstrates comprehensive manuscript preparation capabilities through its modular design and automated workflows. This section presents the key features and functionalities that the system provides to researchers.

3.1. Cross-Platform Reproducibility

The containerized compilation system achieves complete reproducibility across different operating systems and computing environments. Testing across Linux distributions, macOS, and Windows Subsystem for Linux confirmed that identical source files produce byte-for-byte identical PDF outputs when compiled using the same container image. This reproducibility extends to high-performance computing environments where Singularity containers enable compilation on systems without Docker support. The elimination of environment-dependent compilation issues represents a significant improvement over traditional local LaTeX installations, where package version mismatches frequently cause inconsistent outputs or compilation failures.

3.2. Automated Figure and Table Management

The automatic asset processing system effectively handles diverse input formats and streamlines figure incorporation [?]. Figure ?? demonstrates the framework's capability to include images with properly formatted captions, while Figure ?? shows how multiple figures can be managed systematically. The preprocessing pipeline converts source images to optimal formats, maintaining quality while ensuring compatibility with LaTeX compilation requirements [?]. For tables, the system provides structured organization as shown in Table ??, where complex tabular data can be defined independently and automatically integrated into the document flow. This separation of content from presentation enables authors to focus on data rather than formatting syntax.

3.3. Modular Content Organization

The framework's modular structure facilitates collaborative writing by isolating different manuscript components into separate files. Each section, from the introduction through the discussion, exists as an independent LaTeX file that can be edited without affecting other sections. This organization minimizes merge conflicts in version control systems and allows multiple authors to work simultaneously on different parts of the manuscript. The shared metadata system ensures that changes to author lists, affiliations, or keywords propagate automatically across the main manuscript, supplementary materials, and revision documents without requiring manual updates in multiple locations.

3.4. Version Tracking and Difference Generation

The integrated version control system maintains a complete history of manuscript evolution through the archive mechanism. Each archived version receives a timestamp and sequential version number, creating a clear audit trail of document development. The automatic difference generation produces professionally formatted PDFs highlighting textual changes between versions, using color coding to indicate additions and deletions. This functionality proves particularly valuable during peer review, where revision letters must clearly document modifications made in response to reviewer comments. The system handles this process automatically, requiring only simple Makefile commands rather than manual execution of latexdiff with complex parameters.

4. Discussion

The SciTeX Writer framework addresses fundamental challenges in scientific manuscript preparation by combining containerized compilation, modular organization, and automated asset management into a cohesive workflow. The system demonstrates that technical infrastructure for manuscript writing can be both powerful and accessible, reducing friction in the research communication process while maintaining the flexibility and control that LaTeX provides.

4.1. Advantages of the Containerized Approach

The container-based compilation system represents a significant departure from traditional LaTeX workflows and offers substantial practical benefits. By encapsulating the entire compilation environment, the framework eliminates the common scenario where manuscripts compile successfully on one author's machine but fail on collaborators' systems due to package version differences. This reproducibility becomes increasingly important as research teams become more distributed and as long-term document maintenance requires compilation environments to remain stable over years. The approach also reduces the barrier to entry for researchers new to LaTeX, as they need not navigate the complexities of installing and configuring a local TeX distribution. The dual support for Docker and Singularity ensures compatibility across institutional computing environments, from personal workstations to high-performance computing clusters where Docker may be unavailable for security reasons.

4.2. Implications for Collaborative Writing

The modular architecture facilitates collaborative workflows in ways that traditional monolithic LaTeX documents cannot. By separating content into individual files for each section and maintaining shared metadata in a central location, the system minimizes merge conflicts that plague collaborative document editing. Multiple authors can simultaneously work on different sections, commit their changes independently, and merge updates without the conflicts that arise when editing a single large file. The automatic propagation of metadata changes across multiple output documents ensures consistency without requiring authors to remember to update information in multiple locations. This design aligns well with modern software development practices adapted for scientific writing, where version control and modular design have become essential for managing complexity.

4.3. Comparison with Existing Solutions

Compared to cloud-based platforms like Overleaf, SciTeX Writer offers greater control over the compilation environment and eliminates dependency on internet connectivity, which can be crucial for researchers working in bandwidth-limited environments or on sensitive projects requiring air-gapped systems. Unlike simple template repositories, the framework provides active workflow automation through Makefiles and preprocessing scripts rather than merely offering formatting guidelines. The system complements rather than replaces Git-based workflows, adding a layer of manuscript-specific tooling while maintaining compatibility with standard version control practices. Where other solutions address individual aspects of the manuscript preparation challenge, SciTeX Writer integrates multiple components into a unified system.

4.4. Limitations and Considerations

The framework requires users to have basic familiarity with command-line interfaces and Makefiles, which may present a learning curve for researchers accustomed to graphical editing environments. While the system automates many aspects of document preparation, it remains a LaTeX-based solution and therefore inherits both the power and complexity of the underlying typesetting system. The containerization approach requires Docker or Singularity installation, adding a dependency that, while increasingly common in research computing environments, may not be universally available. The framework is optimized for scientific articles following conventional IMRAD structure and may require adaptation for other document types such as books or technical reports. Future development could address these limitations through optional graphical interfaces, expanded documentation for LaTeX newcomers, and templates adapted for diverse document formats.

4.5. Future Directions and Extensibility

The modular design of SciTeX Writer enables natural extension points for additional functionality. Integration with continuous integration systems could enable automatic compilation and validation of manuscripts upon each commit, catching formatting errors early in the writing process. Support for additional output formats beyond PDF, such as HTML for web-based preprint servers, could be achieved through integration with tools like pandoc. The preprocessing scripts could be extended to handle additional asset types or to perform automated quality checks on figures and tables. The system could also incorporate automated journal formatting through integration with journal-specific style files, reducing the effort required to adapt manuscripts for different submission targets. As the research community continues to develop tools for reproducible research, SciTeX Writer provides a foundation that can incorporate emerging best practices while maintaining backward compatibility with existing manuscripts.

4.6. Conclusions

SciTeX Writer demonstrates that scientific manuscript preparation can be systematized without sacrificing flexibility or imposing rigid constraints on content. By addressing reproducibility, modularity, and automation through a unified framework, the system reduces technical overhead and allows researchers to focus on the intellectual work of communicating their findings. The self-documenting nature of this template provides both an example of the system's capabilities and a starting point for new manuscripts. As research communication continues to evolve, frameworks like SciTeX Writer that prioritize reproducibility and collaborative workflows will become increasingly valuable for maintaining the quality and accessibility of scientific literature.

Data Availability Statement

The NeuroVista dataset used in this study is publicly available through the International Epilepsy Electrophysiology Portal (IEEG.org) at `https://www.ieeg.org`. Access requires registration and approval for research purposes.

The processed PAC databases and analysis code are available at `https://github.com/ywatanabe1989/neurovista`. GPU-accerelated PAC calculation code is available as a standalone Python package 'gpac' at `https://github.com/ywatanabe1989/gPAC`. The SciTeX Python utilities used for reproducible computing is available at `https://github.com/ywatanabe1989/SciTeX`.

For questions regarding data access or analysis procedures, please contact the corresponding author.

Ethics Declarations

All study participants provided their written informed consent ...

Author Contributions

Y.W., T.Y., and D.G. conceptualized the study ...

Acknowledgments

This research was funded by funding bodies here

Declaration of Interests

The authors declare that they have no competing interests.

Declaration of Generative AI in Scientific Writing

The authors employed large language models such as Claude (Anthropic Inc.) for code development and complementing manuscript's English language quality. After incorporating suggested improvements, the authors meticulously revised the content. Ultimate responsibility for the final content of this publication rests entirely with the authors.

Tables

Tables

Table 1 – Table 0: Placeholder

To add tables to your manuscript:

1. Place CSV files in `caption_and_media/` with format `XX_description.csv`
2. Create matching caption files `XX_description.tex`
3. Reference in text using `Table~\ref{tab:XX_description}`

Example: `01_seizure_count.csv` with `01_seizure_count.tex`

Step	Instructions
1. Add CSV	Place file like `01_data.csv` in `caption_and_media/`
2. Add Caption	Create `01_data.tex` with table caption
3. Compile	Run `./compile -m` to process tables
4. Reference	Use `\ref{tab:01_data}` in manuscript

Figures

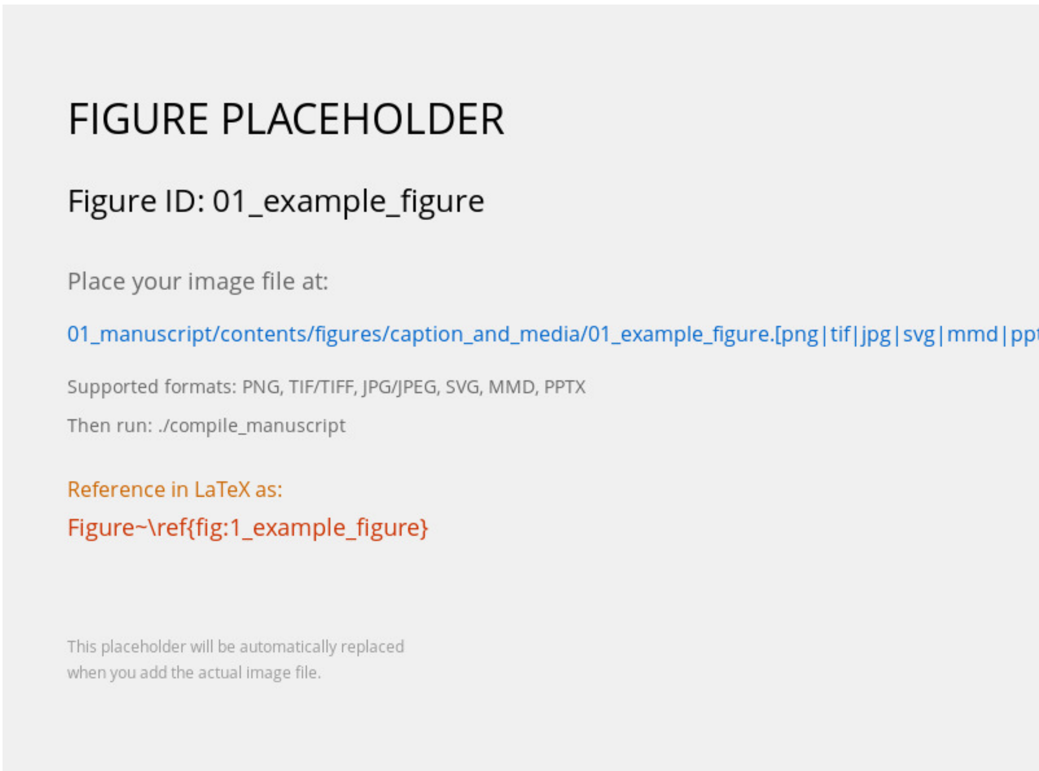

Figure 1 – Example figure caption. This is a template showing how to include figures in your manuscript. Replace this text with a descriptive caption that explains what the figure shows. Include panel labels (A, B, C) if using multi-panel figures. Explain abbreviations and symbols used in the figure. Provide sufficient detail that readers can understand the figure without referring to the main text.

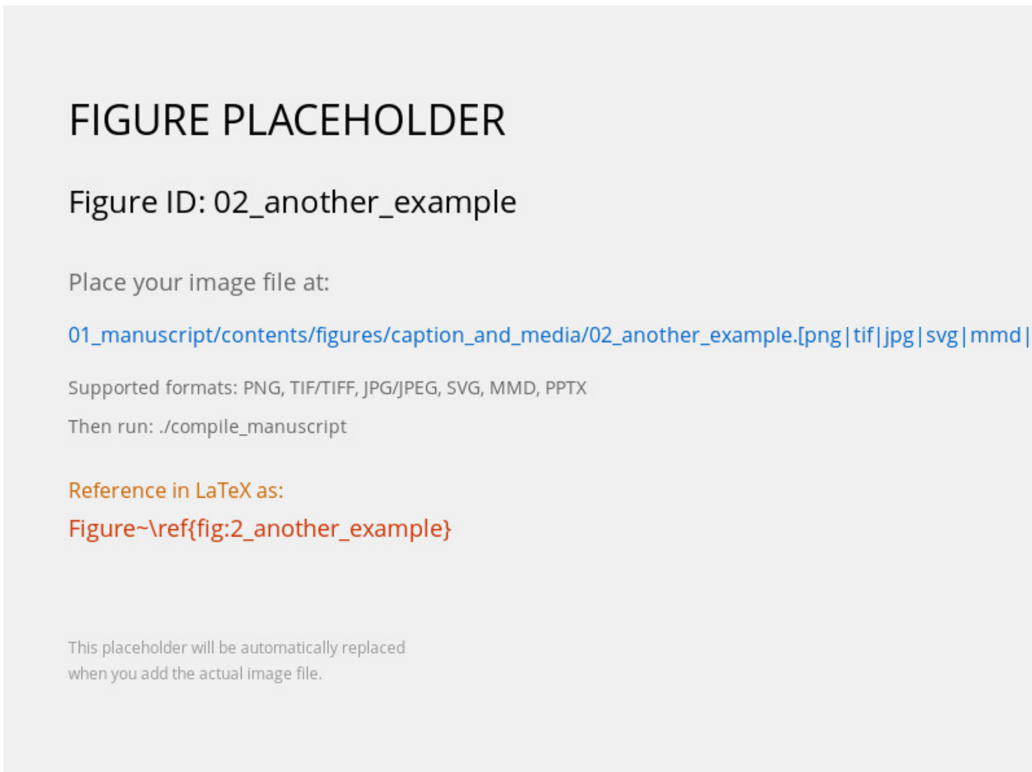

Figure 2 – Another example figure. Use this template to add additional figures to your manuscript. Each figure should be placed in a separate .tex file in this directory. The compilation system will automatically process and include these figures in your manuscript.