

Article

Not peer-reviewed version

JMDSFCv1.0: an Interactive R/Shiny Application for Dataset Format Conversion with Real-Time Progress Monitoring.

[January Msemakweli](#)^{*} and Khalid Mzuka

Posted Date: 9 September 2024

doi: 10.20944/preprints202409.0665.v1

Keywords: Dataset Conversion; Shiny Application; Data Formats; Excel to Stata; Data Interoperability; Excel to SPSS; JMDSFCv1.0; Data Science Tools; Cross-platform Data Handling



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

JMDSFCv1.0: An Interactive R/Shiny Application for Dataset Format Conversion with Real-Time Progress Monitoring

January G. Msemakweli * and Khalid Juma Mzuka

Muhimbili University of Health and Allied Sciences, School of Public Health and Social Sciences, Department of Environmental and Occupational Health; khalidmzuka02@gmail.com

* Correspondence: msemakwelijanaury@gmail.com

Abstract: In modern data analysis, working with diverse dataset formats is essential for ensuring compatibility across different software tools and platforms. However, manual conversion between formats can be time-consuming and error-prone. To address this challenge, we developed an interactive Shiny application that facilitates the seamless conversion of datasets between popular software formats such as CSV, Excel, SAS, SPSS, Stata, and RData. This user-friendly app enables users to upload datasets, select an output format, and download the converted file with a single click. The application also incorporates a real-time progress indicator, enhancing the user experience by showing conversion progress. Built with simplicity and efficiency in mind, the app is designed for researchers, data analysts, and practitioners who regularly handle multiple file formats in their workflows. Additionally, the app features a clean, intuitive interface with clear guidance, making it accessible even to users with minimal technical background. The development process, key features, and potential applications of this tool are discussed in this article.

Keywords: dataset conversion; shiny application; data formats; Excel to Stata; data interoperability; Excel to SPSS; JMDSFCv1.0; data science tools; cross-platform data handling

Rationale for Developing JMDSFCv1.0: Dataset Format Converter

The increasing diversity of data formats used across various fields such as data science, research, and industry pose significant challenges for interoperability and seamless data analysis [1–6]. Analysts often work with datasets in different formats such as CSV, Excel, SAS, SPSS, Stata, and RData, each requiring specific software tools for manipulation and analysis [7,8]. This fragmentation not only complicates the workflow but also increases the likelihood of errors during manual file conversions [3,9]. Manual conversion methods are often tedious, error-prone, and time-consuming, leading to inefficiencies, especially in high-stakes environments where data integrity is critical [1]. Furthermore, some formats require specialized software knowledge, which limits accessibility for individuals without the technical expertise [10–17]. To address these challenges, JMDSFCv1.0 was developed as a user-friendly solution that automates dataset format conversion. It enables users to convert datasets between various formats effortlessly through a simple, interactive interface. Additionally, the app includes a real-time progress indicator, ensuring users are informed about the conversion process, improving transparency and user experience.

This application is especially beneficial for researchers, data analysts, and practitioners who handle diverse datasets, providing a streamlined, efficient, and reliable tool to facilitate smooth data interoperability. The development of JMDSFCv1.0 addresses the need for an accessible, automated solution, empowering users to focus more on data analysis rather than format compatibility.

Software's Approach

The development of **JMDSFCv1.0** follows a structured, user-centered approach to ensure seamless dataset format conversion with minimal technical overhead. The software is built using the **R Shiny** framework [18], which provides an interactive web interface for user engagement while leveraging the power of R for data manipulation and conversion [18–25]. This approach allows the application to function in real-time, responding to user inputs and providing immediate feedback on the progress of conversions.

Key Features and Workflow:

1. **User Interface Design:** The interface of JMDSFCv1.0 is designed with simplicity and clarity in mind, ensuring accessibility for both novice and experienced users. The app provides a file input field where users can upload datasets in various formats, such as CSV, Excel, SAS, SPSS, Stata, and RData. A drop-down menu allows users to select the desired output format, simplifying the conversion process [22–25].
2. **File Format Support:** JMDSFCv1.0 supports a wide range of data formats, including CSV, Excel (.xlsx), SAS (.sas7bdat), SPSS (.sav), Stata (.dta), and RData (.rdata/.RDATA) [19–21]. This broad format compatibility ensures that the application can cater to a variety of user needs, regardless of their preferred data analysis software.
3. **Real-Time Progress Indicator:** The app features a progress bar that updates in real-time during the conversion process. This feature is particularly valuable for larger datasets, as it provides users with a clear indication of how far along the conversion process is, preventing unnecessary delays or confusion.
4. **Automated Conversion Logic:** The core functionality of JMDSFCv1.0 is driven by automated conversion logic. Once a dataset is uploaded and the target format is selected, the app automatically handles the conversion using appropriate R libraries. For instance, readr is used for CSV files, readxl for Excel, haven for SAS, SPSS, and Stata, and native R functions for RData. After conversion, the app enables users to download the newly formatted file directly.
5. **Download Feature:** The app's download button becomes visible only after the conversion process is complete, ensuring that users can confidently retrieve their dataset in the desired format. This feature minimizes errors and improves the overall user experience.
6. **Transparency and Feedback:** JMDSFCv1.0 is designed to provide users with continuous feedback, both through the progress bar and the file preview functionality. This transparency ensures that users are always aware of the current status of their data and the conversion process.

This user-friendly and efficient approach ensures that JMDSFCv1.0 caters to a wide range of data conversion needs, reducing the time and effort typically involved in switching between dataset formats while improving overall data accessibility and analysis efficiency.

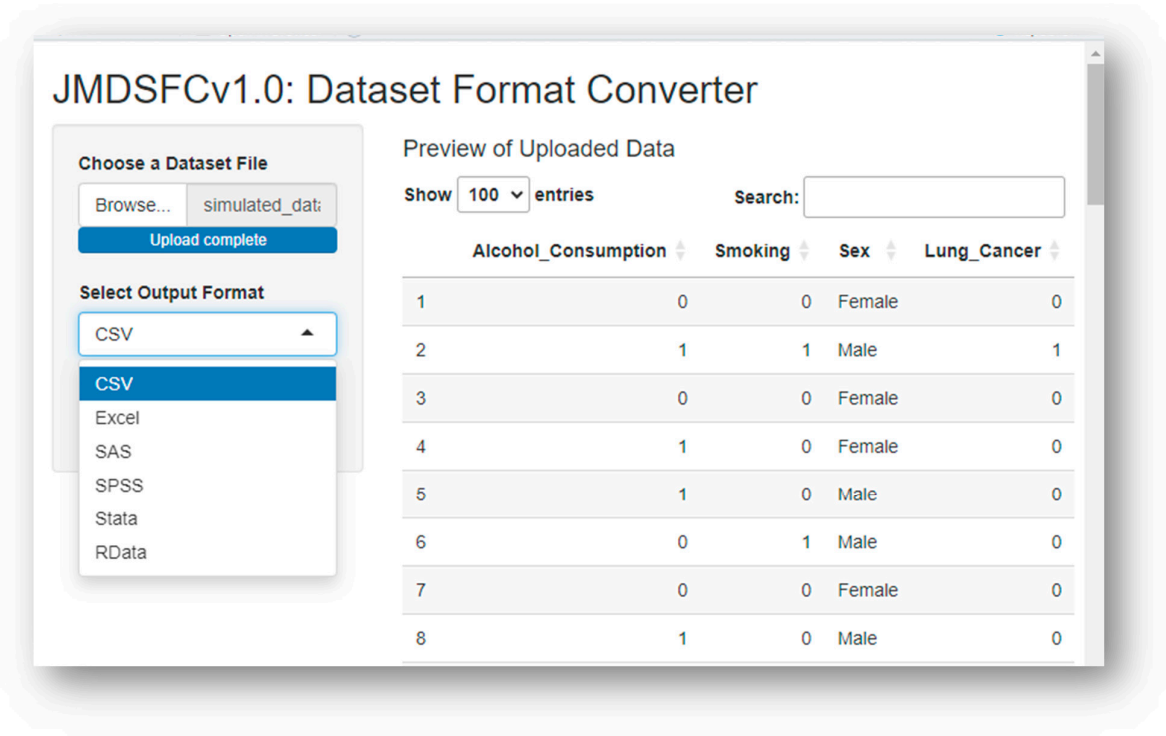


Figure 1. Screenshot of JMDSFCv1.0’s UI.

Conclusions

The development of JMDSFCv1.0 addresses a critical need for an easy-to-use, reliable, and efficient solution for converting datasets between various formats. By simplifying the process of data format conversion and integrating real-time progress monitoring, the application significantly reduces the complexities faced by researchers, data analysts, and practitioners when dealing with multiple formats. JMDSFCv1.0’s intuitive interface, automated logic, and wide format compatibility make it a powerful tool for streamlining data interoperability and improving workflow efficiency. This application ensures that users can focus on analysis and decision-making rather than the technicalities of data conversion, providing a valuable resource in the realm of data management.

Acknowledgments: None.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: Data sharing is not applicable to this article as no new data were created in this study.

Ethics Statement: Not applicable.

Author Contributions: J.M. and K.M.: Conceptualization, Investigation, Project administration, Validation, Coding and Visualization, Writing – original draft, Writing – review & editing.

References

1. Pereira M, Velosa N, Pereira L. dsCleaner: A python library to clean, preprocess and convert non-intrusive load monitoring datasets. Data. 2019;4:123.
2. Tuli JK. EVALUATED NUCLEAR STRUCTURE DATA FILE--A MANUAL FOR PREPARATION OF DATA SETS. Brookhaven National Lab.(BNL), Upton, NY (United States); 2001.
3. Reddy KB, Reddy CL, Pulluri S, Akash MDN, Gopal SV. VMEG Mini Tool Kit--An Intelligent Approach For File Conversion. IJIRT. 2022.

4. Hulstaert N, Shofstahl J, Sachsenberg T, Walzer M, Barsnes H, Martens L, et al. ThermoRawFileParser: modular, scalable, and cross-platform RAW file conversion. *J Proteome Res.* 2019;19:537–42.
5. Chen Q, Liao Q, Jiang ZL, Fang J, Yiu S, Xi G, et al. File fragment classification using grayscale image conversion and deep learning in digital forensics. *2018 IEEE Secur Priv Work. IEEE;* 2018. p. 140–7.
6. Sriramakrishnan P, Kalaiselvi T, Padmapriya ST, Shanthi N, Ramkumar S, Kalaichelvi N. An medical image file formats and digital image conversion. *Int J Eng Adv Technol.* 2019;9:74–8.
7. Han HJ, Yoon S-H, Oh H-J, Yang D. Empirical Verification of Conversion and Restoration of Preservation Format for Dataset: Application of Dataset with Disaster Safety Information to SIARD. *정보관리학회지.* 2020;37:251–84.
8. Chernov S, Minack E, Serdyukov P. Converting desktop into a personal activity dataset. *Proc 9th Russ Natl Res Conf Digit Libr. Citeseer;* 2007. p. 280–3.
9. Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics.* 2012;28:298–9.
10. Ermilov I, Auer S, Stadler C. Csv2rdf: User-driven csv to rdf mass conversion framework. *Proc ISEM.* 2013. p. 4–6.
11. Salheb N, Arroyo Oho K, Stoter J. Automatic conversion of CityGML to IFC. *Int Arch Photogramm Remote Sens Spat Inf Sci.* 2020;44:127–34.
12. Kumar BH, Babu MSP. Performance Analysis of New Conversion Tool from Relational Datasets to XML Datasets on Selected Websites.
13. Francart T, Clavaud F, Charbonnier P. RiC-O converter: a software to convert EAC-CPF and EAD 2002 XML files to RDF datasets conforming to records in contexts ontology. *Linked Arch 2021 Proc Linked Arch Int Work 2021 co-located with 25th Int Conf Theory Pract Digit Libr (TPDL 2021).* 2021. p. p-30.
14. Bigi B. Annotation representation and file conversion tool. *Contrib del Cent Linceo Interdiscip 'Beniamino Segre.'* 2018;137:99–116.
15. Butler MH, Gilbert J, Seaborne A, Smathers K. Data conversion, extraction and record linkage using XML and RDF tools in Project SIMILE. *HP Labs, Bristol, UK.* 2004;
16. Andrews SJ. Data conversion and interoperability for FCA. 2009;
17. Kashani AA, Noori AM, Mahriyar H. CONVERSION LOG DATASET INTO UNDERSTANDABLE FORMAT FOR DATA MINING. 2014;
18. Core R. Team. R a Lang Environ Stat Comput. 2015;2021.
19. Wickham H, Bryan J. R packages. " O'Reilly Media, Inc."; 2023.
20. Team RC, Bivand R, Carey VJ, DebRoy S, Eglen S, Guha R, et al. Package 'foreign.' 2020.
21. Wickham H, Bryan J, Kalicinski M, Valery K, Leittenne C, Colbert B, et al. Package 'readxl.' Version. 2019;13:1.
22. Warnes MGR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A. Package 'gplots.' Var R Program tools plotting data. 2016;112–9.
23. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Package 'shiny.' See <http://citeseerx.ist.psu.edu/viewdoc/download>. 2015;
24. Jia L, Yao W, Jiang Y, Li Y, Wang Z, Li H, et al. Development of interactive biological web applications with R/Shiny. *Brief Bioinform.* 2022;23:bbab415.
25. Resnizky HG. Learning Shiny. Packt Publishing Ltd.; 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.