TRX: A community-oriented tractography file format

Francois Rheault[a], Valérie Hayot-Sasson[b], Robert E. Smith[c1,c2], Christopher Rorden[d], Jacques-Donald Tournier[e], Eleftherios Garyfallidis[f], Fang-Cheng Yeh[h], Christopher J. Markiewicz[i], Matthew Brett[j], Ben Jeurissen[k], Paul A. Taylor[l], D. Baran Aydogan[m1,m2], Derek A. Pisner[n], Serge Koudoro[f], Soichi Hayashi[o], Daniel Haehn[p], Steve Pieper[q], Daniel Bullock[r], Emanuele Olivetti[s], Jean-Christophe Houde[u], Marc-Alexandre Côté[i], Flavio Dell'Acqua[x], Alexander Leemans[y], Maxime Descoteaux[u,v], Bennett Landman[a,w], Franco Pestilli[g], and Ariel Rokem[z]

Affiliations:

[a]*Electrical and Computer Engineering, Vanderbilt University, Nashville, USA;*
[b]*Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada*
[c1]*Florey Institute of Neuroscience and Mental Health, Melbourne, Australia*
[c2] *Florey Department of Neuroscience and Mental Health, The University of Melbourne, Melbourne, Australia*
[d]*Department of Psychology, University of South Carolina, Columbia, USA*
[e]*Biomedical Engineering Department, School of Biomedical Engineering and Imaging Sciences, King's College London,UK*
[f]*Department of Intelligent Systems Engineering, Indiana University, Bloomington, USA*
[g]*Department of Psychology, The University of Texas at Austin, Austin, USA.*
[h]*Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, USA*
[i]*Department of Psychology, Stanford University, Stanford, CA, USA*
[j]*Brain Imaging Center, University of California, Berkeley, Berkeley, CA, USA*
[k]*Department of Physics, University of Antwerp, Antwerp, Belgium*
[l]*Scientific and Statistical Computing Core, National Institute of Mental Health, USA*
[m1]*A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland*
[m2]*Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland*
[n]*University of Texas at Austin, Austin, USA*
[o]*Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA*
[p]*Machine Psychology Lab, Department of Computer Science, University of Massachusetts, Boston, USA*
[q]*Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston, United States*
[r]*Center for Magnetic Resonance Research, University of Minnesota*
[s]*Neuroinformatics Laboratory, Fondazione Bruno Kessler, Trento, Italy*
[t]*Microsoft Research Lab, Montréal, Canada*
[u]*Imeka Solutions, Sherbrooke, Canada*
[v]*Sherbrooke Connectivity Imaging Laboratory (SCIL), Département d'Informatique, Université de Sherbrooke, Canada;*
[w]*Vanderbilt University Institute of Imaging, Nashville, USA;*
[x]*NatBrainLab, Dept. of Forensics and Neurodevelopmental Sciences, King's College London, London, UK*
[y]*PROVIDI Lab, Image Sciences Institute, UMC Utrecht, Utrecht, The Netherlands*
[z]*Department of Psychology and eScience Institute, University of Washington, Seattle, USA*

Introduction

File formats that store the results of computational tractography were typically developed within specific software packages. This approach has facilitated a myriad of applications, but this development approach has generated insularity within software packages, and has limited standardization. Moreover, because tractography file formats were developed to solve immediate challenges, only a limited breadth of applications within a single software package was envisioned, sometimes also neglecting computational performance. Given the growing interest in tractography methods and applications, and the increasing size and complexity of datasets, a community-driven standardization of tractography have become a priority. To address these challenges, our community initiated a discussion to design a new file format and agreed to participate in its conception, development, and, if successful, its adoption. Here, we propose **TRX,** a tractography file format designed to facilitate dataset exchange, interoperability, and state-of-the-art analyses, acting as a community-driven replacement for **TCK** [1], **TRK** [2], **VTK**, **TT** [3], **TRAKO** [4], **qFIB** [5], **zFIB** [6], **NIML.TRACT** [10] and **DPY** [8].

Background

Following discussions on GitHub (https://github.com/nipy/nibabel/issues/942), during a videoconference between the developers of several software libraries for tractography, and at the DIPY online group meeting, we identified key features that were desired by the community (both developers and users):

1. Maintain functionality currently implemented by other file formats;
2. Computational efficiency (both in speed and file size);
3. Extensibility so the file format could evolve more easily.

Feedback from different groups showed that developers desired features such as being based on open-source, well-maintained libraries, and robustness, while users prioritized new functionalities, simplicity, and low memory usage.

Results

A TRX file format candidate specification was designed to favor simplicity and human readability. Its data are organized following an intuitive naming convention that conveys dimensionality and data type. The specification allows storage of both data per vertex and per streamline (similar to TRK), streamlines-group, with both sparse and dense group indices representation. RASMM (world/scanner coordinates) for vertex coordinates is proposed (similar to TCK). The file format supports random-access with minimal upfront RAM requirements, and other low-level operations with virtually no RAM requirements, by using a data and offsets approach (similar to VTK9) and raw binary arrays that allow memory mapping. The file format supports choice over the data type for each raw data array (see Figure 1) and, optional, ZIP compression. Just like other file formats, efficient streamline compression [6] via linearization method is possible without modifying the current specification.

An initial benchmark for operations like partial file loading (Figure 2), concatenation (not shown), affine transformation (not shown), and file saving (Figure 2) was performed using the Python implementation and compared to Nibabel. A C++ prototype was also developed and proposed for community feedback.

Conclusion

The goal of TRX is to become the first, community-driven, standard amongst tractography file formats. As with other file formats like NiFTI [9], we believe that TRX will serve the community well and the growing computational needs of our field. We encourage community members to consider early contributions to our proposal so as to ensure the new standard will cover the needs of the wider audience of software developers, toolboxes, and scientists. Our long-term plan is to integrate TRX within the Brain Imaging Data Structure (BIDS) ecosystem [8]. Community contributions and related discussions are encouraged using: https://github.com/tee-ar-ex/trx-python/issues.

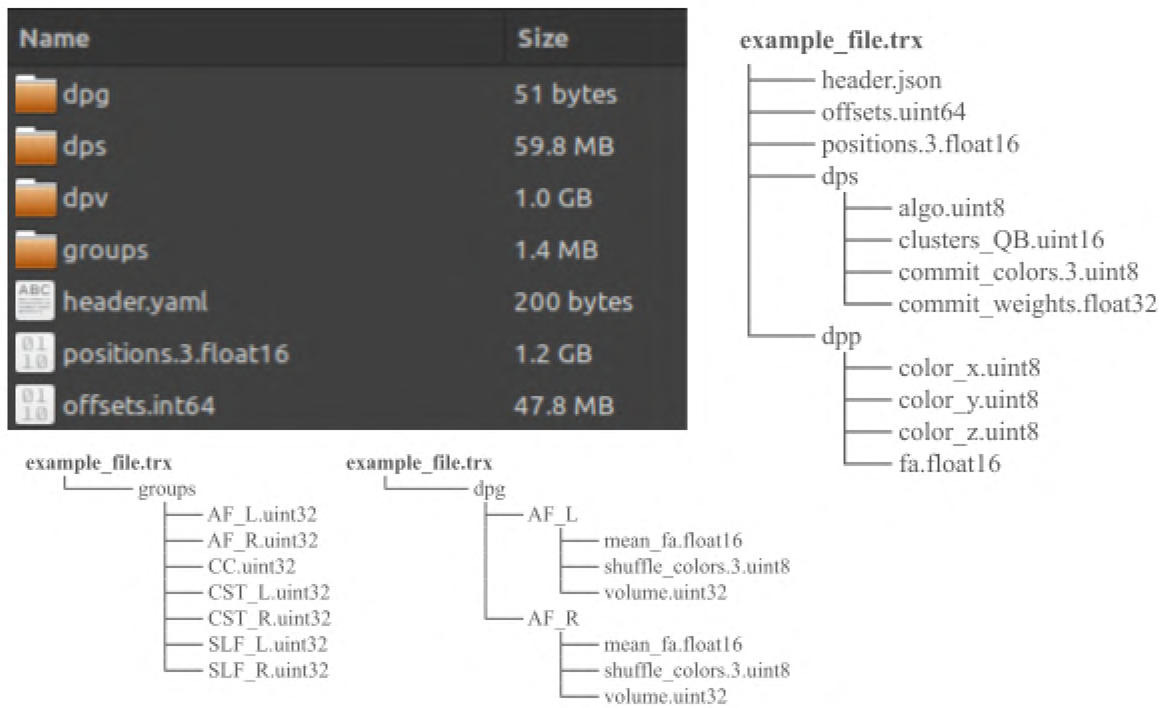
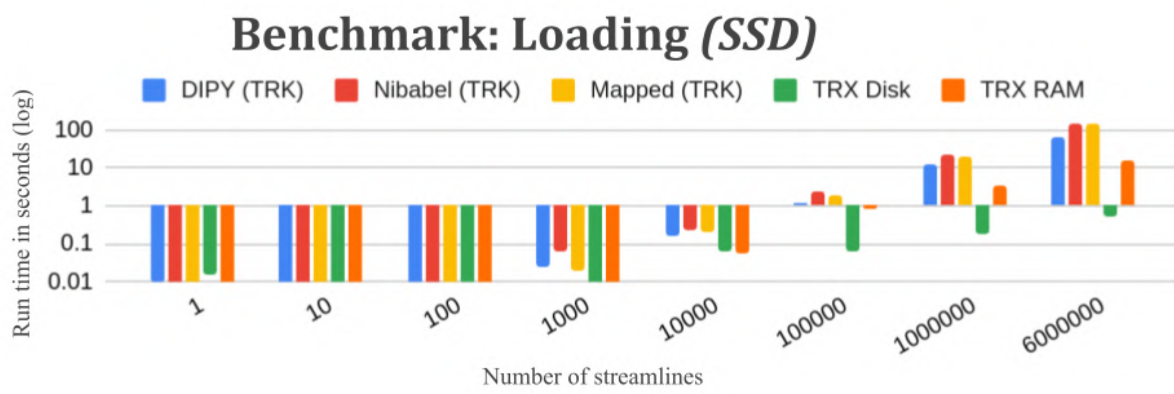


Figure 1: File structure of the TRX file format. A filename convention conveys the meaning, dimensionality and data type for each data array. The header, positions (coordinates of streamlines) and offsets (starting position of streamlines in the array) are mandatory files. Optional data include the data per streamline (dps), data per vertex (dpv), groups and data per group (dpg). Specifications can be found at https://github.com/tee-ar-ex/trx-spec/blob/master/specifications.md.



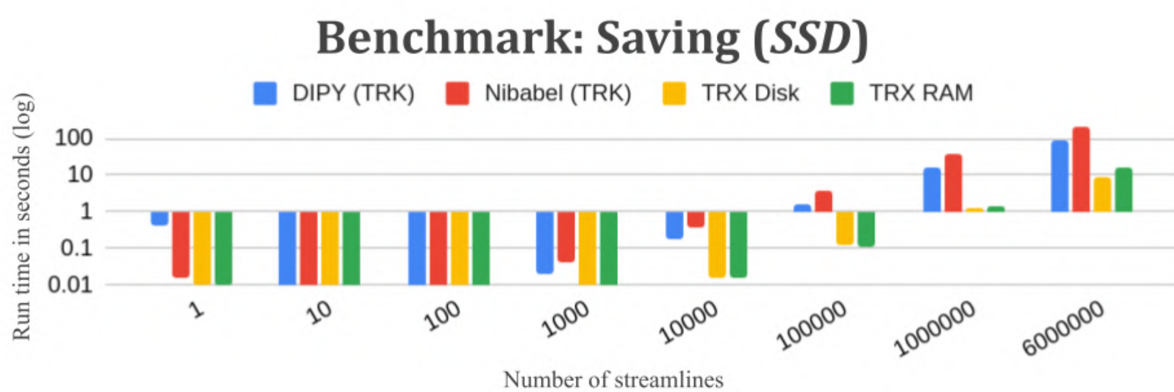


Figure 2: Load times (in seconds) for loading and saving a small (1 streamline) to a large file (6M streamlines). To facilitate visualization and comparison, the graph is shown on a logarithmic scale. *DIPY* is a Python implementation from DIPY [7] that does not support data per streamline / data per vertex (deprecated, TRK). *Nibabel* is the Nibabel API (version 3.0, TRK). *Mapped* is an implementation that allows random access (available on https://github.com/emanuele/load_trk, TRK). The loading operation of *TRX Disk* is in fact, simply the memory mapping initialization, there is no "loading". For the saving operation, *TRX Disk* is simply a copy-paste operation, which showcases the baseline disk speed. *TRX RAM* is the memory mapping initialization and transfer of all data from disk to RAM. Python implementations are likely slower than the implementations of TrackVis or MI-Brain, but this nevertheless showcases the advantage of memory mapping even for slower languages. The Python prototype used for this benchmark can be found at https://github.com/tee-ar-ex/trx-python/.

References

[1] Tournier, J-Donald, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun-Hung Yeh, and Alan Connelly. "MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation." *Neuroimage* 202 (2019): 116137.

[2] Wang, Ruopeng, Thomas Benner, Alma Gregory Sorensen, and Van Jay Wedeen. "Diffusion toolkit: a software package for diffusion imaging data processing and tractography." In *Proc Intl Soc Mag Reson Med*, vol. 15, no. 3720. 2007.

[3] Yeh, Fang-Cheng, Timothy D. Verstynen, Yibao Wang, Juan C. Fernández-Miranda, and Wen-Yih Isaac Tseng. "Deterministic diffusion fiber tracking improved by quantitative anisotropy." *PloS one* 8, no. 11 (2013): e80713.

[4] Mercier, Corentin, Sylvain Rousseau, Pietro Gori, Isabelle Bloch, and Tamy Boubekeur. "QFib: fast and efficient brain tractogram compression." *Neuroinformatics* 18 (2020): 627-640.

[5] Haehn, Daniel, Loraine Franke, Fan Zhang, Suheyla Cetin-Karayumak, Steve Pieper, Lauren J. O'Donnell, and Yogesh Rathi. "Trako: Efficient transmission of tractography data for visualization." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 322-332. Springer, Cham, 2020.

[6] Rheault, Francois, Jean-Christophe Houde, and Maxime Descoteaux. "Visualization, interaction and tractometry: dealing with millions of streamlines from diffusion MRI tractography." *Frontiers in neuroinformatics* 11 (2017): 42.

[7] Garyfallidis, Eleftherios, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. "Dipy, a library for the analysis of diffusion MRI data." *Frontiers in neuroinformatics* 8 (2014): 8.

[8] Pestilli, Franco, Russ Poldrack, Ariel Rokem, Theodore Satterthwaite, Franklin Feingold, Eugene Duff, Cyril Pernet, Robert Smith, Oscar Esteban, and Matt Cieslak. "A community-driven development of the brain imaging data standard (BIDS) to describe macroscopic brain connections."

[9] Gorgolewski, Krzysztof J., Tibor Auer, Vince D. Calhoun, R. Cameron Craddock, Samir Das, Eugene P. Duff, Guillaume Flandin et al. "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments." Scientific data 3, no. 1 (2016): 1-9.

[10] Cox, Robert W. "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages." Computers and Biomedical Research 29, no. 3 (1996): 162-173.