GRIIDC Dataset Compression Guidance for Windows Users

When datasets contain multiple files, GRIIDC asks data providers to package these dataset files into a single compressed archive file. Most Windows users normally create zip files using File Explorer's 'Send to -> Compressed (zipped) folder', as seen in Figure 1. **GRIIDC finds that large zip files created using File Explorer's 'Send to -> Compressed (zipped) folder' cannot be opened reliably.** GRIIDC recommends using other applications to create compressed archive files, such as 7-zip.
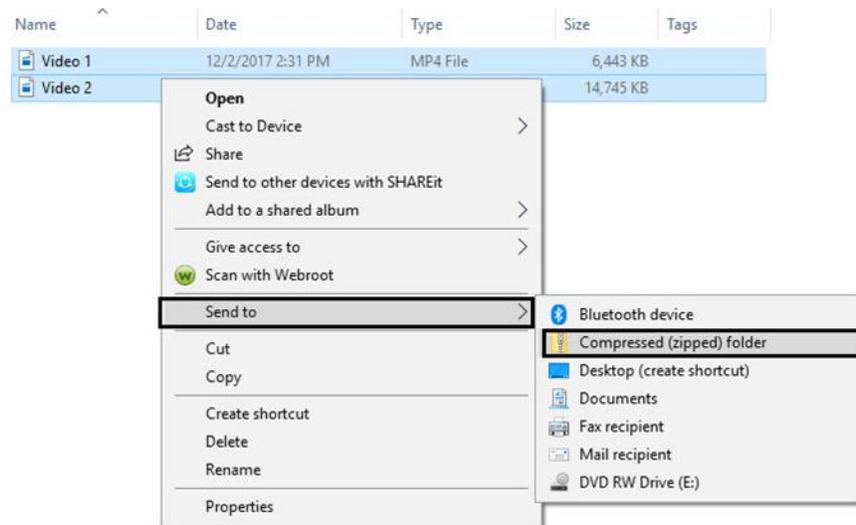
Figure 1: File Explorer's 'Send to -> Compressed (zipped) folder'

As part of the review process, GRIIDC tests all compressed files to ensure all data files can be extracted. GRIIDC has repeatedly encountered larger zip files that cannot be successfully opened with standard tools. After investigation, many of these datasets have two traits in common, they are created with File Explorer's 'Send to -> Compressed (zipped) folder' and are larger datasets (> 3GB and/or containing a large number of files).

When GRIIDC encounters errors during testing, we require that the data provider repackage the files and resubmit the dataset. To minimize these issues, **GRIIDC recommends that data providers who are submitting large dataset files do not use File Explorer's 'Send to -> Compressed (zipped) folder' to create zip files.** Alternative compression and archive tools include third party applications such as 7-zip.

How should I compress my data into a single file?

1. Give files and folders meaningful but concise names. Avoid spaces and special characters; use dashes or underscores instead.
2. Do not compress already compressed files — avoid nesting compressed archives.
3. Put the dataset files and folders inside a top-level folder.
4. Archive and compress the folder into a single file using:
   - Small Datasets < 3GB – File Explorer's 'Send to -> Compressed (zipped) folder'
   - Large Datasets > 3GB
     - A third party compression application (we recommend 7-zip)
5. Test the compressed archive file to verify the contents are intact.

What third party compression application should I use?

- There are many available third party compression applications available for Windows. GRIIDC recommends 7-zip. It is freely available on its home page - http://www.7-zip.org/.

Should I compress my files before compressing the folder?

- Please do not create nested compressed files (i.e., zip files that contain more zip files) because multiple unpacking steps are required to extract a file listing or a subset of files. In addition, multiple compression steps do not further reduce the file size.

How do I test my compressed archive?

- For smaller files, it is often easiest to extract the file to a new folder and compare contents, preferably with a different program than the one used to create the file. Larger files can be tested using 7-zip.

What should I do if my dataset is a single file?

- Please do not create a zip archive to submit a single file.