# **Figshare Tool Documentation**

## **Purpose**

We have manually done the Figshare process, and we have taken 10-20 minutes to complete a single article. This Perl script is designed to process ZIP archives containing XML metadata files and associated supplementary data. It performs the following tasks:

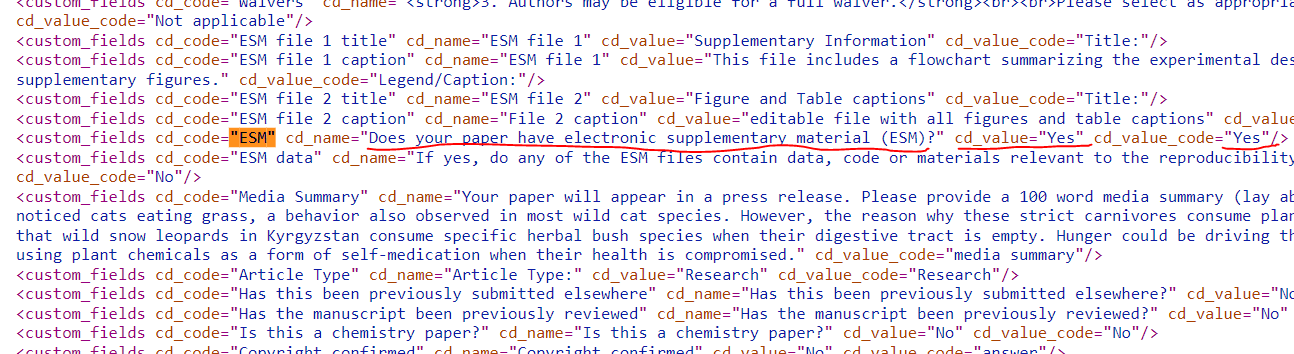
1. **Extracts ZIP Files:** Unzips files from a specified directory and extracts their contents.
2. **Processes XML Metadata:** Reads and parses XML files to extract metadata, including custom fields, author information, and dates.
3. **Copies Supplementary Data:** Searches for and copies files from specific subdirectories (e.g., suppl\_data) if their names match entries in the metadata.
4. **Generates XML and Renames Files:** Renames extracted files according to metadata information and create corresponding XML files with updated details.

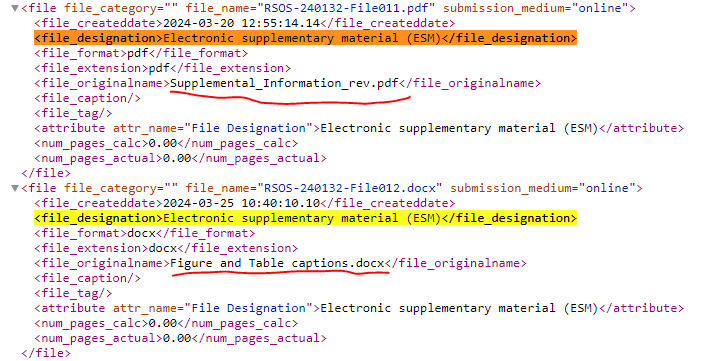
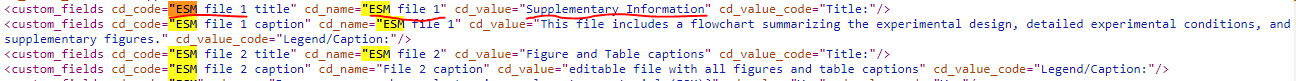
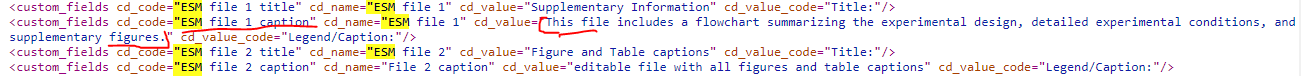
Now, this script helps to complete the Figshare process in 30 seconds.

Github URL(Source code): <https://github.com/nagaraj665/Figshare>

**Manual Figshare Process**

1. The ESM process is confirmed from the ‘\*\*\*-metadata.xml’. If the XML indicates 'yes', proceed with the file; if it indicates 'no', skip the article.



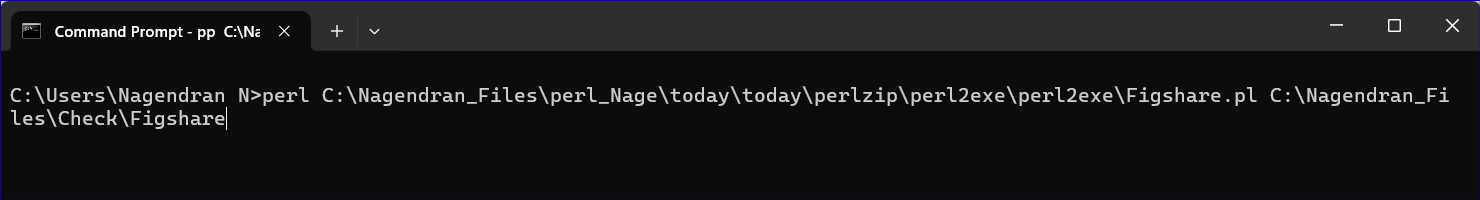
1. Collect supplementary files from the articles' ‘suppl’ folder (all file types are accepted).
   1. Use only the supplementary files listed in the ‘\*\*\*-metadata.xml’. Remove any files not listed and follow the sequence order as specified in the metadata.xml.
   2. Rename the supplementary file names using the following format.  
      ’jounal name’‘journal ID’\_si\_00‘seqence order’.extension  
      eg., rsos240132\_si\_001.pdf
2. Additionally, we will create two types of XML files for this process.;
   1. ’jounal name’‘journal ID’\_si\_00‘seqence order’\_metadata.xml  
      Eg., rsos240132\_si\_001\_metadata.xml
   2. ’jounal name’‘journal ID’\_si\_collection.xml  
      Eg., rsos240132\_si\_collection.xml
3. Process of metadata.xml as below; use the following file for sample. 
   1. Add the article ID in <uid> tag – rsos240132001
   2. <title> [like a label of the supplement] – find from the metadata.xml “ESM file 1 title” and use data of cd\_value.  
        
      use ‘Supplementary Materials’, if title is unavailable.
   3. <description> [like a caption of the supplement] – find from the metadata.xml “ESM file 1 caption” and use data of cd\_value.  
        
      caption if not in metadata
      1. use caption from supplement file. Or;
      2. use supplement file name as per input received.
   4. embargodate - current+7days
   5. submission\_date – <submitted\_date> (use from meta-data.xml)
   6. acceptance\_date – same as submission date
4. In the “doi\_si\_001\_collection.xml” changes below;
   * **title** - article\_title
   * **description** - abstract (if any doubt check with input pdf)
   * **authors** – author in metadata.xml (if any doubt check with input pdf)
   * **tags** – use subjects from the pdf
   * **categories** – any one category allowed, subject match with json file.
5. Zip with main folder.
6. Upload to the FTP site. (inside the ‘data’)

### **Fighshare tool Key Steps:**

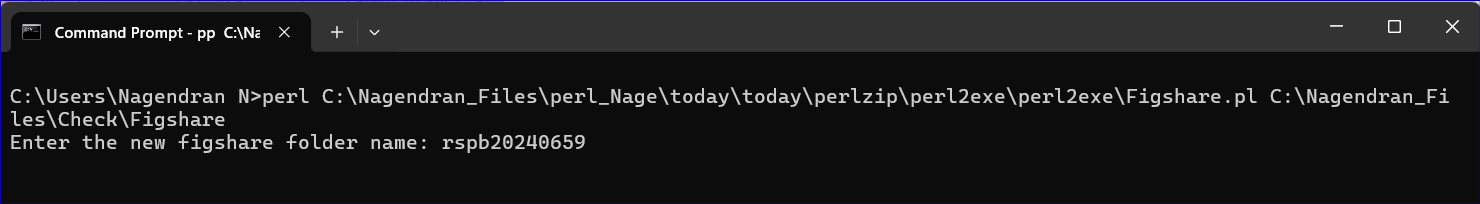
1. **Extraction of ZIP Files:**
   * The Perl script reads and extract files from ZIP archives found in the input directory.
   * Creates necessary directory as per Fighshare structures for extracted files.
2. **Metadata Extraction:**
   * Reads META XML files to extract information such as submission dates, journal abbreviations, article titles, DOIs, and custom fields.
   * Collects author details and other relevant metadata.
3. **Supplementary Data Handling:**
   * Identifies and copies files from ‘suppl\_data’ subdirectories that match metadata entries to the output directory.
4. **File Renaming and XML Generation:**
   * Renames files based on metadata information and creates ‘\*\_metadata.xml’ & ‘\*\_collection.xml’ files with updated metadata.

### **Steps to Run the Script**

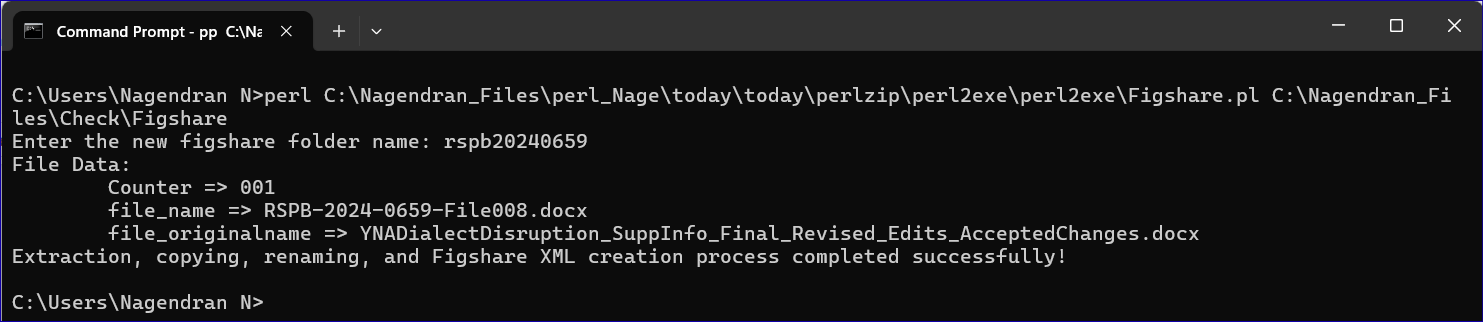
1. **Execute the Script:**
   * Run the script from the command line with the directory path containing the ZIP files.



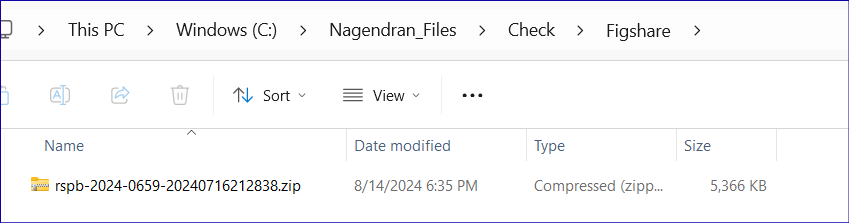
1. **Provide Output Directory Name:**
   * When prompted, enter the name for the new output folder where processed files will be stored.



* Enter processed\_files when prompted for the new output folder name.
* The script will create a processed\_files directory, extract ZIP files into it, and process the metadata as described.

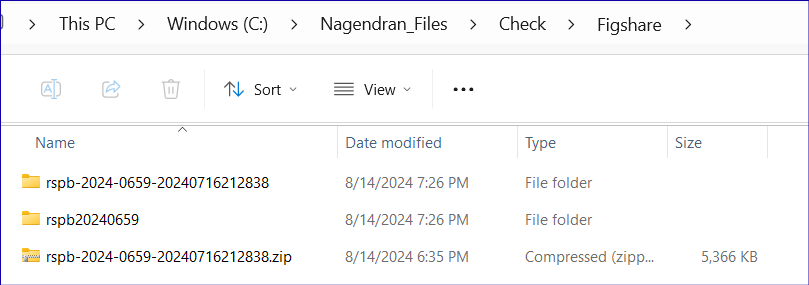


### **Input**



### **Output**

**Screenshot 1**



**Screenshot 2**

