For our purposes, the process of extracting the relevant information from articles entails two general processes:

1. Cataloging the relevant information about the article itself.
2. Cataloging the white matter structures they discuss, and those constituent discussions.

For now, we are approaching this by creating a directory for each article, simply with the name doc[x], where [x] simply corresponds to some sequential number. Within this directory there are two general Excel files:

1. A file with data about the article itself, entitled doc.xlsx
2. Some number of excel files, entitled tract1.xlsx, tract2.xlsx, tract3.xlsx, etc. for each white matter structure discussed.

Thankfully, the first of these doesn’t actually entail that much work. Because we are able to query much of the information using DOI. As such we really only need to record:

1. The curator’s name
2. The doi
3. The pertinent species
4. The pertinent methods

All other article information can be automatically populated.

The creation of excel spreadsheets for tracts (i.e. white matter structures) is somewhat more laborious. It entails (for each tract) filling out the following 3 fields.

1. tractTermsUsed: In each cell, one of the terms used to refer to the white matter structure in question, either the full name, abbreviation, or alternate names provided.
2. descriptions: using the new Google Chrome “link to highlighted text” capability (described here <https://support.google.com/chrome/answer/10256233>) create a link to each individual section that discusses the tract in question. In the Chrome browser, highlight the relevant text, and then right click the highlight and select the “**copy link to highlight**” option. I don’t actually know how this works exactly (I’ve asked [here](https://support.google.com/chrome/thread/120988476/text-extraction-using-new-link-to-highlight-chrome-feature?hl=en)), but it allows us to link to specific regions of text on a website. To some degree, there is an element of subjectivity as to what is pertinent. Remember our main interest here are descriptions of anatomy, and that to some extent, more granular selections are preferred to selection with a lot of “chaff”
3. Figures: links to *the images themselves* for any depictions of the tract. Typically these should be a web link that has a .jpg, .png or some other image file type at the end.

For an example of this, done on Bullock et al. 2019, check the directory containing this document.

NOTE: the .json file that you may or may not find here is the product of a code pipeline that extracts information from the excel files. Curators are not expected to produce these.