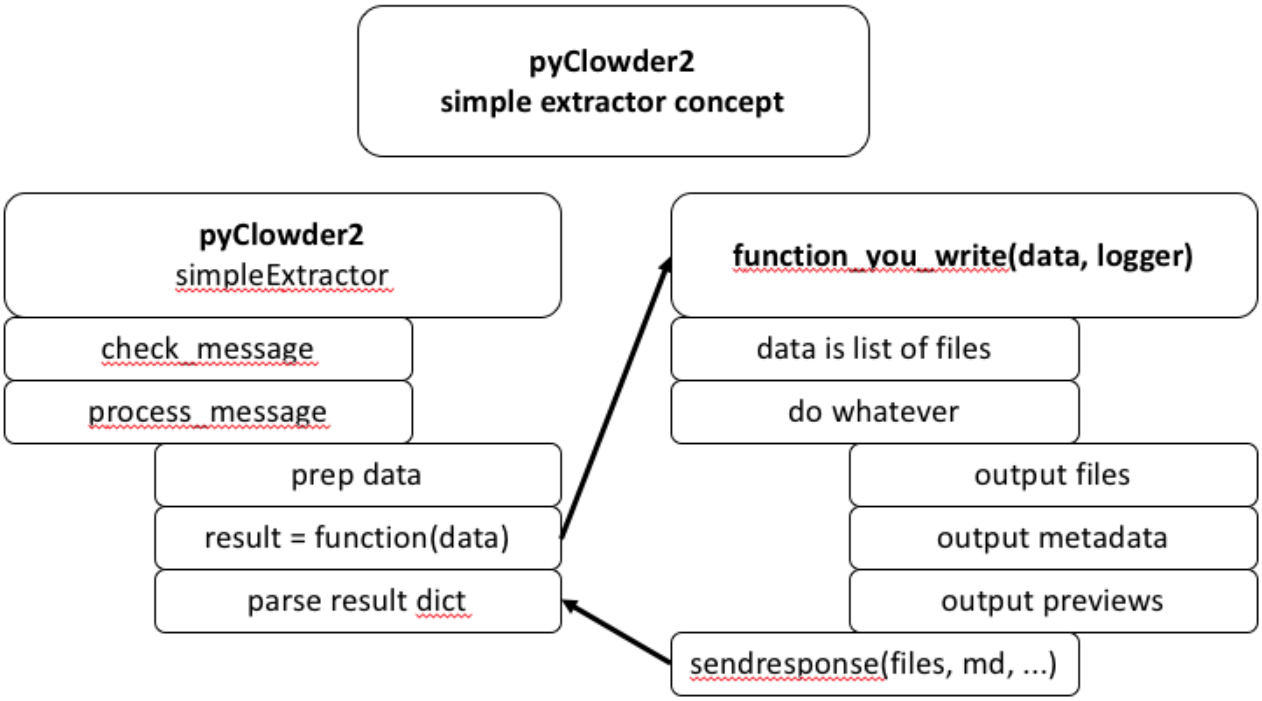
Simple Extractor wrapper for basic functions

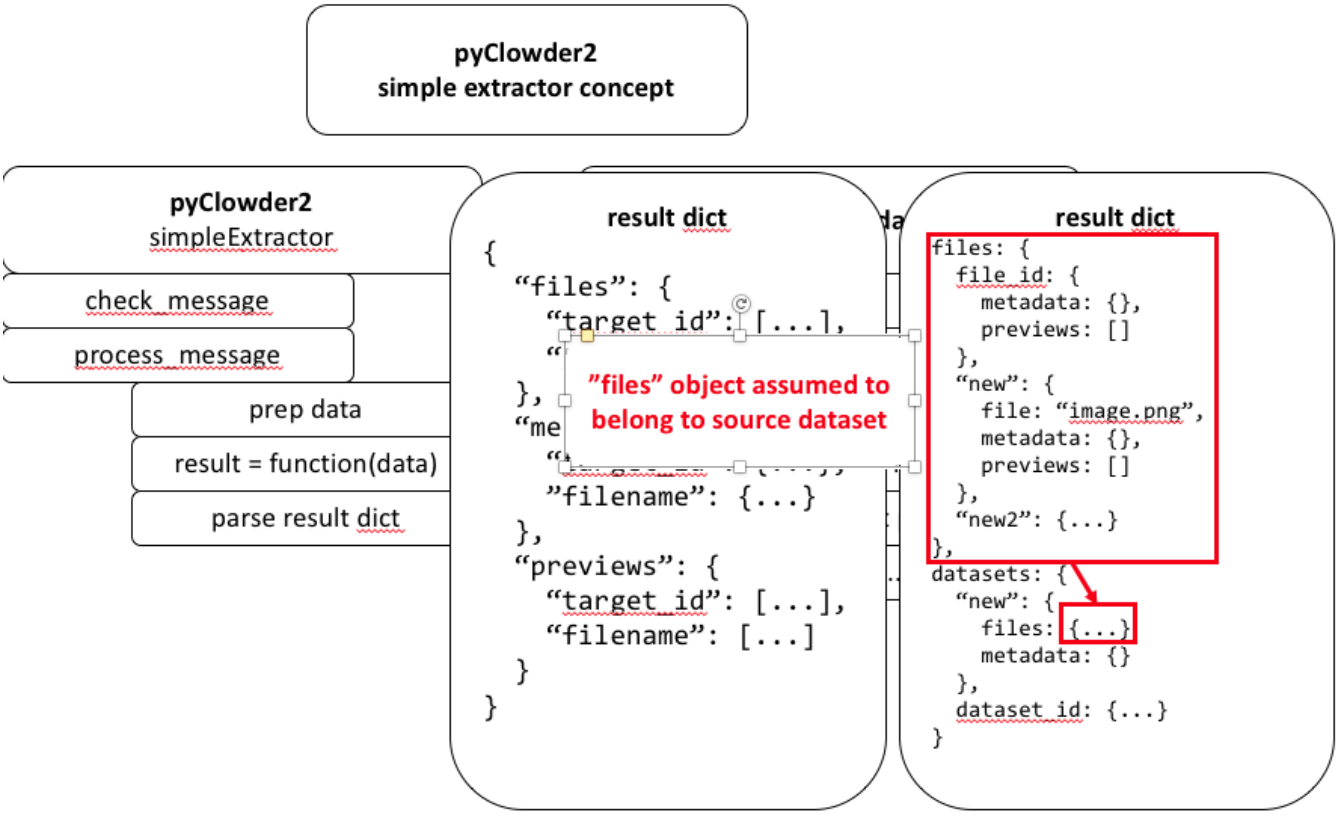
This extractor would utilize pyClowder framework to handle simple check_message() and process_message() components, but with the chance to call an external function to do the actual extractor work.

Developers would write a function that takes an input or data (and probably the logger used by simple extractor) and returns a JSON dict with any new files, metadata, previews etc. that result from the function:



So here, the simple extractor would handle everything and call the configured function (configured on initialization) with the data to get a result, then parse the result in a standard way.

The result dict would have a structure that allows users to define outputs for their function:



So in this example response object:

- "files" key has a dict with any files to be uploaded.
  - using file_id as subkey allows uploading metadata and previews to existing file.
  - using "new_123" or any "new" prefix key will result in new file created that can also include metadata and previews
  - dict can include as many files as desired
  - files in the "files" portion are assumed to belong to dataset that triggering input file belongs to.
- "datasets" key has a dict if the extractor needs to create NEW datasets, or upload files to datasets different the one that triggered the extractor
  - "new" or "new_123..." will create a new dataset that can include files and dataset metadata.
  - existing dataset ID will load files or metadata to that dataset.
  - the "files" dict has same structure described above.

Big idea is to let developers simply write code to process an input file based on 2 parameters: the data and the logger.

At the end, it can call a sendrespose(files, metadata) function of some kind to auto build the dict for simple extractor to parse. We should think about this - maybe different sendresponse() functions if file vs. dataset extractor? Don't want users to have to build the JSON object themselves necessarily, although maybe they have to if the JSON object is complex and sendresponse() is just for basic responses?

Single File Extractor: