Manual workflow
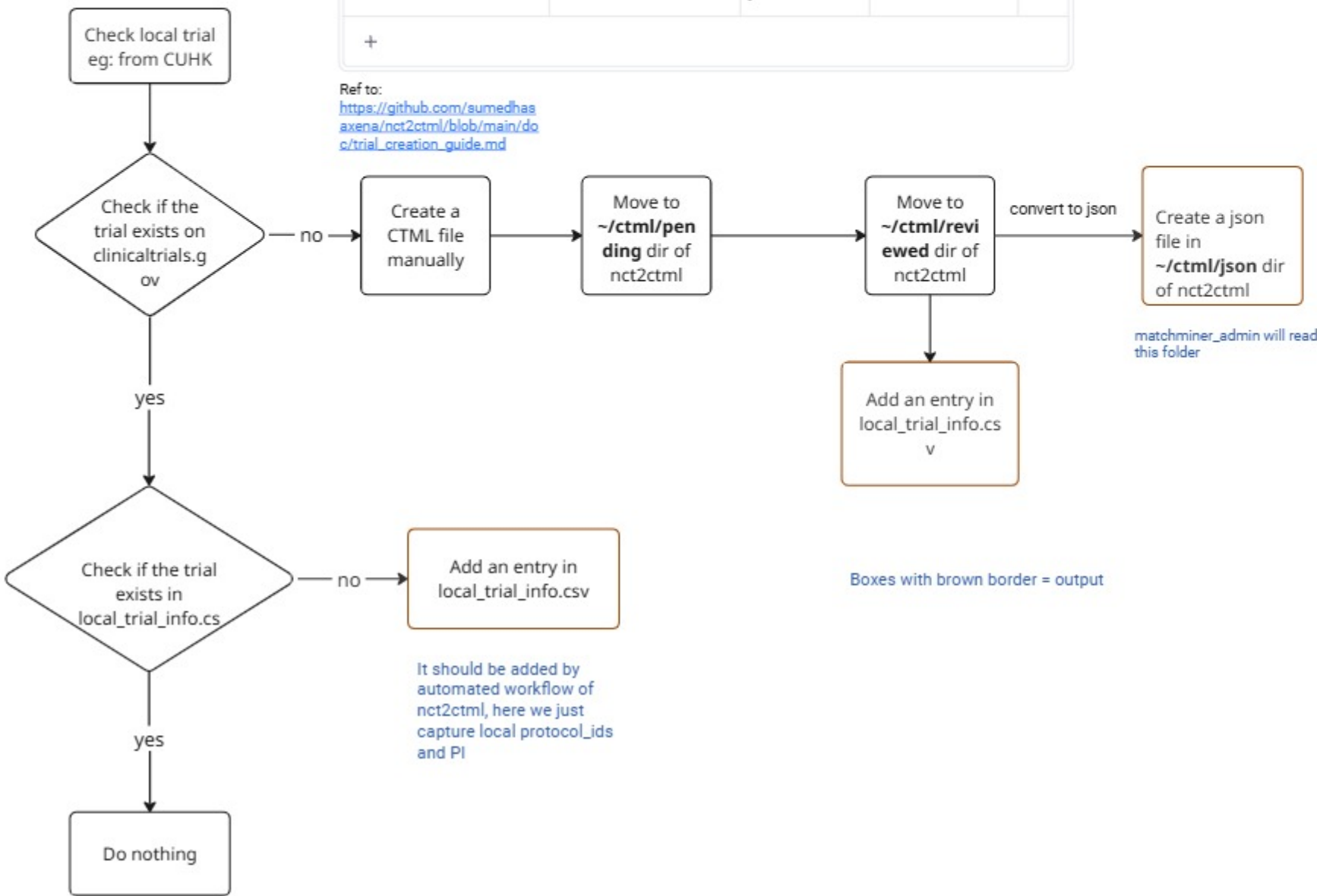
Purpose:
- Capture following info for trials that are in clinical_trials.gov
  - local protocol_ids
  - pi_name
  - pi_institution

local_trial_info.csv

	nct_id	local_protocol_ids	pi_name	pi_nstitution	+
	NCT08	ACT011, xxx	John Smith	CUHK	
	NCT10	ACT012	John Smith	CUHK	
	NA	ACT013	John Smith	CUHK	
+					

Ref to:
https://github.com/sumedhaa
axena/nct2ctml/blob/main/do
c/trial_creation_guide.md

```
Check local trial
eg: from CUHK
        │
        ▼
Check if the trial exists on clinicaltrials.gov ──no──► Create a CTML file manually ──► Move to ~/ctml/pending dir of nct2ctml ──► Move to ~/ctml/reviewed dir of nct2ctml ──convert to json──► Create a json file in ~/ctml/json dir of nct2ctml
        │ yes
        ▼                                                                                              │
Check if the trial exists in local_trial_info.csv ──no──► Add an entry in local_trial_info.csv          ▼
        │ yes                                                                                    Add an entry in local_trial_info.csv
        ▼
Do nothing
```

matchminer_admin will read this folder

It should be added by automated workflow of nct2ctml, here we just capture local protocol_ids and PI

Boxes with brown border = output

Automated workflow

1st step:
pull command: --pull all

Input: trial_status.csv and local_trial_info.csv

Purpose: Pull trial data from clinicaltrials.gov, identify NCTIds that need to be inserted, updated or closed and update trial_status.csv as required. Also merge local trials manually added in local_trial_indo.csv into trial_status.csv

Steps:
1. Query trial status via API request along with 'LastUpdatePostDate'
2. Create 3 empty lists: 'to_be_inserted', 'to_be_updated', 'to_be_closed'
3. For each trial in API response,
   a. If the **trial is not in trial_status.csv**
      i. If **status = open**, add nct_id to 'to_be_inserted' list
      ii. If **status = closed**:
         1. If trial is in local_trial_info.csv -> add to 'to_be_inserted' list.
         2. else -> do nothing
   b. If the **trial is in trial_status.csv**
      i. If **status has changed to 'closed'** -> add nct_id to 'to_be_closed' list
      ii. else if 'LastUpdatePostDate' > 'trial_last_updated_date' -> add nct_id to 'to_be_updated' list
      iii. else -> do nothing
4. For all nct_ids in 'to_be_inserted':
   a. Pull trial JSON from clinical_trials.gov via API and save in cache/nct dir.
   b. Update trial_status.csv:
      i. Insert new row with appropriate status
      ii. Set 'trial_last_updated_date' as 'LastUpdatePostDate'
      iii. Set 'entry_last_updated_date' as current date
5. For all nct_ids in 'to_be_updated':
   a. Pull trial JSON from clinical_trials.gov via API and save in cache/nct dir.
   b. Update trial_status.csv:
      i. Update existing row with 'trial_last_updated_date' as 'LastUpdatePostDate'
      ii. Update existing row with 'entry_last_updated_date' as current date
6. For all nct_ids in 'to_be_closed':
   a. Update trial_status.csv:
      i. Update existing row with 'last_updated_date' as 'LastUpdatePostDate' and 'status' as 'closed'
      ii. Update existing row with 'entry_last_updated_date' as current date
7. Merge local trials in trial_status.csv:
   a. Read local_trial_info.csv
   b. For rows with nct_id as NA,
      i. look up trial_status.csv with local_protocol_id(s), and insert the trial in trial_status.csv if not already present
      ii. Set nct_id = NA, local_protocol_ids = local ids, status = open, entry_last_updated_date = current date
   c. For rows with nct_id,
      i. lookup trial_status.csv, to see if any row exists for corresponding nct_id.
      ii. If yes, check if local_protocol_ids are present. If not, add local_protocol_id(s) and update entry_last_updated_date = current date

Output:
1. Downloaded trial files in cache/nct
2. Updated trial_status.csv

2nd step:
map command: --map all
Input:
1. Downloaded trial files in cache/nct
2. trial_status.csv
3. local_trial_info.csv

Purpose: For all studies with a nct_id, map/convert trial files to CTML

Steps:
1. Read all files in cache/nct dir. For each file,
2. If trial in trial_status.csv has 'entry_last_updated_date' as current_date
   a. Map to a CTML file.
   b. if trial in trial_status.csv has local_protocol_ids(s), append local trial info from local_trial_info.csv in the CTML file
   c. Save ctml file in cache/ctml

Output: Mapped trial files in cache/ctml, along with local trial info, if available

---

3rd step:
Manual review
Input:
1. Mapped trial files in cache/ctml
Purpose: Review CTML files and convert CTML files to JSON files.

Steps:
1. Read all files in cache/ctml dir. For each file,
2. If trial in trial_status.csv has 'entry_last_updated_date' as current_date
   a. Review trial info
   b. Convert to JSON file
   c. Save the file in ctml/json dir

Output: Converted JSON files in ctml/json dir

---

trial_status.csv

	nct_id	local_protocol_ids	status	trial_last_updated_date	entry_last_...	+
	NCT01		open	Jul 22, 2025		
	NCT02		open	Jul 22, 2025		
	NCT03		open	Jul 22, 2025		
	NCT04		closed	Jul 22, 2025		
	NCT05		open	Jul 22, 2025		
	NCT06		open	Jul 22, 2025		
	NCT07		open	Jul 22, 2025		
	NCT08		open	Jul 22, 2025		
	NA	ACT013	open	Jul 22, 2025		
	NCT09		open	Jul 22, 2025		
	NCT10		open	Jul 22, 2025		
+						

---

matchminer_admin

.env:
LAST_RUN: {date_of_last_run}

1. Read trial_status.csv
2. For each row, if entry_last_updated_date > LAST_RUN:
3. **If the row has nct_id:**
   a. Try to pull trial record from matchminer_db by nct_id
   b. If already exists in matchminer_db: **(update workflow)**
      i. In trial_status.csv, if 'status' == 'open'
         1. Read the corresponding NCTxx.json from ~ctml/json
         2. **Update** the record in matchminer_db
      ii. In trial_status.csv, if 'status' == 'closed'
         1. Update 'status' to 'closed' in system
   c. If does not exist in matchminer_db: **(new record workflow)**
      i. Read the corresponding NCTxx.json from ~ctml/json
      ii. **Insert** the record in matchminer_db
   d. Move/delete file from ~ctml/json
4. **If the row does not have nct_id or nct_id == NA:**
   a. **Try to pull trial record from matchminer_db by local_protocol_ids in CTML**
   b. **Read the file from ctml/json and Insert/update as needed**

```
records inserted/updated in matchminer DB
```