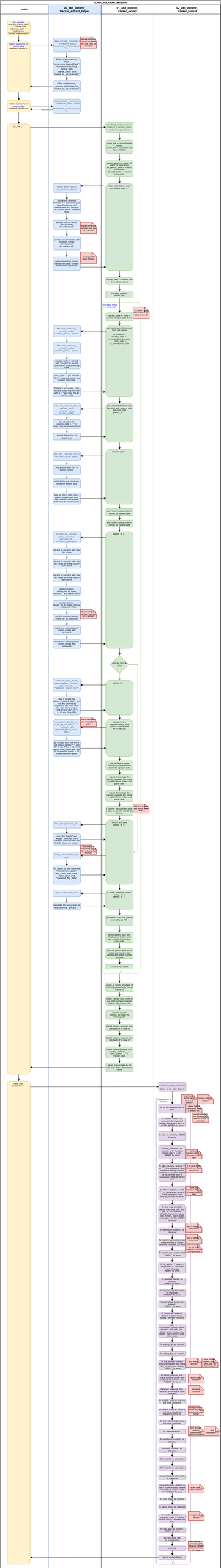
10_Run_add_patient_tracker_extraction
main
00_add_patient_tracker_extract_helper
01_add_patient_tracker_extract
02_add_patient_tracker_format
Set variables incoming_tracker_path = "Tracker.xlsx" sheet_name = "MASTER" header_row = 0
read in synonyms for column_name and coldbook_patient =
read_column_synonyms(coldbook_path, sheet_name, header_row)
Read in an .xlsx file sheet "sheet_name" and "coldbook" ; transform into long format with "column_name" and "name_to_be_matched"
Filter empty rows; remove duplicates for "name_to_be_matched"
read in synonyms for column_name and coldbook_product =
read_column_synonyms(coldbook_path, sheet_name, header_product)
df_raw =
clean_add_patient_tracker(df_raw, coldbook_patients)
clean_sheet_date(df_patient_data)
"months" has different formats: >= 5 columns (in the case of 2020 tracker, 5 columns named) and >= 3 columns (in the case of 2021 tracker) ; month
update column name and synonym names: "month" to "month_dt"
replace matching column names with actual column names from synonyms
return cleaning!
clean_df_columns(df_patient_data)
df_patient_data =
"country_code" = "country" ; "country" = "country_name" (column and actual country)
if "country_code" = get self with "China" or second column also between gen. code
if "China" in col. country, "country_code" = "country_name" ; "country" = "country_code"
df_patient_data["country_code", "country", "country_name"]
find all cells with "US" in second column
extract this row as column names for patient data
only for 2019, 2020, 2021: patient header span over two rows, so column names have been now in column name
extract_header_row(df_patient_data, 0)
df_patient_data
find all cells with "US" in second column
extract this row as column names for patient data
header_row = 0
df_patient_data
drop_rows_before_header(df_patient_data, header_row)
df_patient_data
drop_rows_before_header(df_patient_data, header_row)
df_patient_data
drop_rows_before_header(df_patient_data, header_row)