Country Concordances

☐

Contents

Summary

Country concordance is an important aspect of a data technician's job to maintain the International Futures database. Country concordance refers to the differences between the IFs Country list and another organization's country list and merging them to ensure the IFs country list in IFs system. For example, in some organization "Türkiye" is the name displayed, but for IFs we use the name "Turkey". Therefore IFs data technicians need to change the name to "Turkey" for the system to process it. To do this there are multiple steps that can be taken and they are shown below:

GitHub Repositories

There are two repositories essential to this process; a mini app that can handle concordance for files less than 15mb and a centralized GitHub repository for all country concordance tables.

The app allows users to upload datasets from other organizations and utilizes fuzzy matching against standardized International Futures territory names. This is to make it easier for data technicians to ensure all country data points are being pulled correctly. You can clone the repository but a link to the application can be found here.

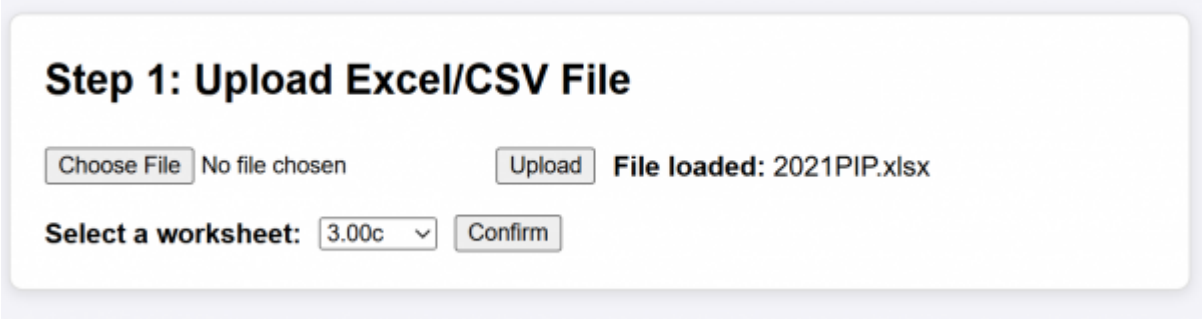
The centralized GitHub is a growing project thank contains excel sheets for mapping tables between an organization's country list and International Futures country names.

The format for these excels are formatted as:

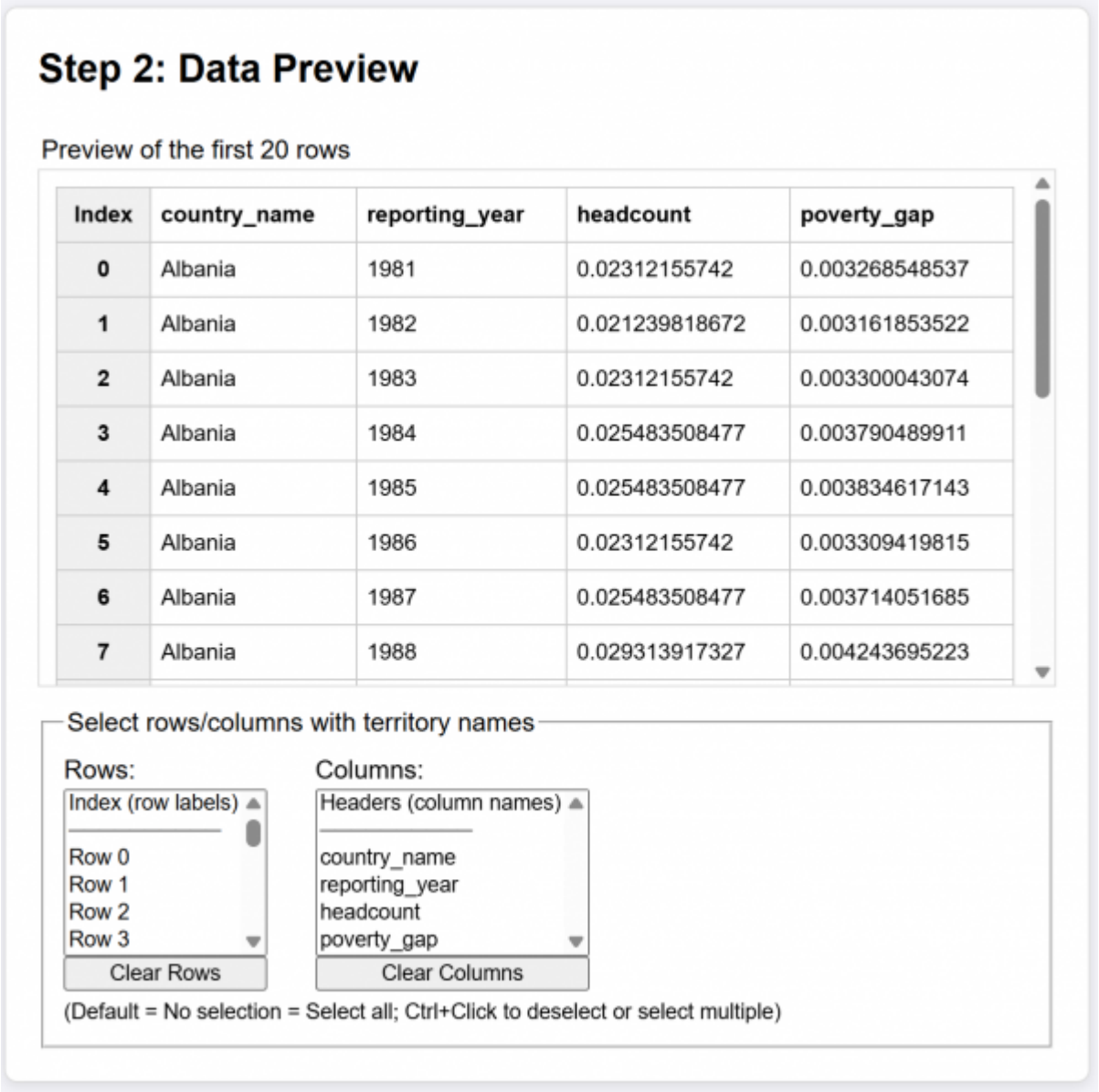
IFs Country	**ex: ITU Country**
Afghanistan	Afghanistan
Albania	Albania
and so on...	

Process to Create Country Concordance

Use the app to import your dataset. Once you choose a file, and click upload and select the worksheet if needed:

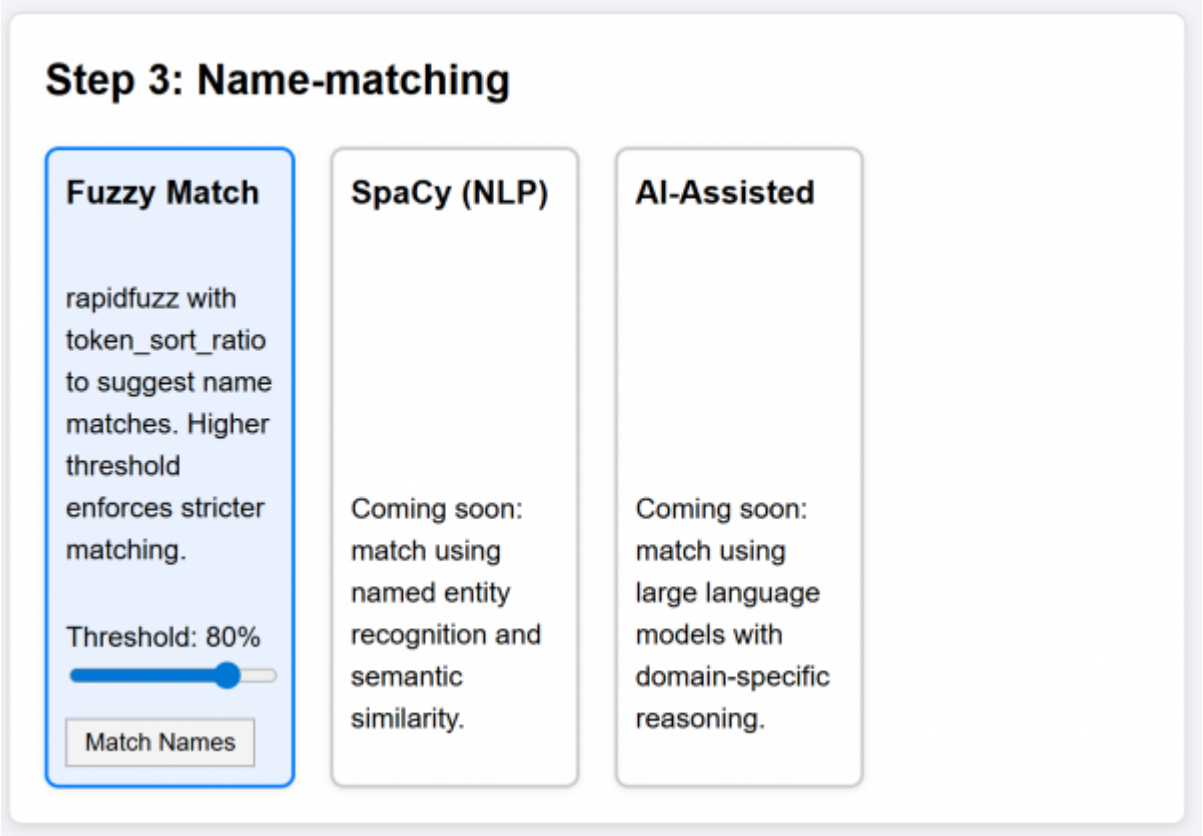


The page will go to Data Preview. Select the row or column that contains the country names. For this example it is the column "country_name"

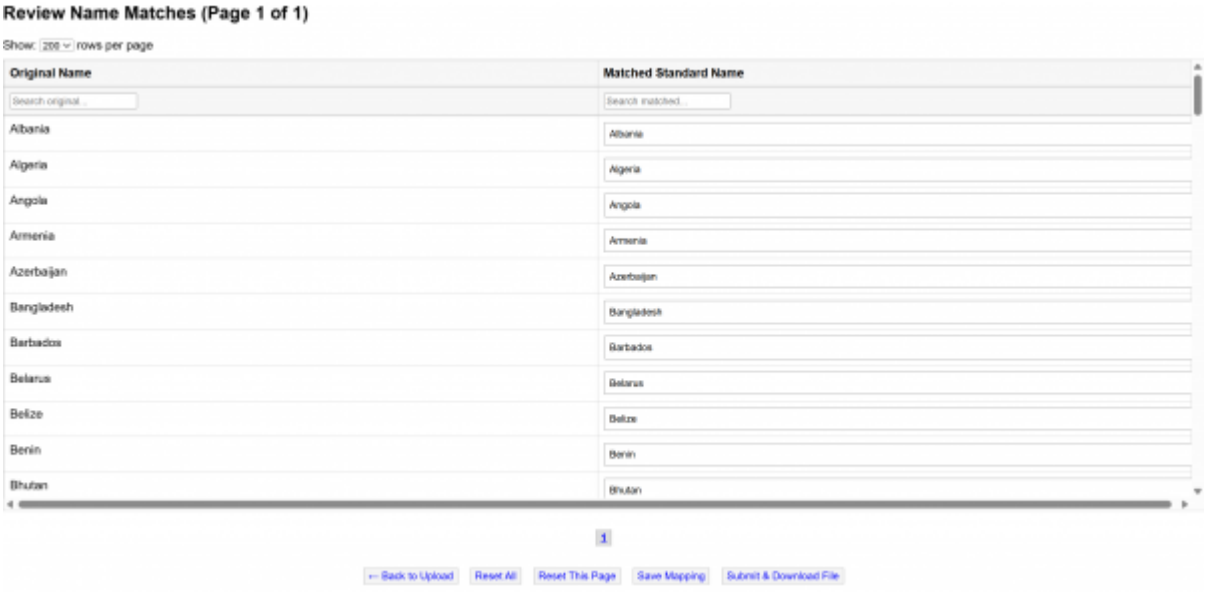


Select Fuzzy Match and at least 80%. You can go higher but high percentages mean that

names can be missed if they fall short.



Once you click match names you will be led to this page:



In the bottom of the page there are two important options; "Save Mapping" and "Submit & Download File."

"Submit & Download File" changes the name in the original file to the IFs country names so the data technician can use the IFs interface to import into a db file. **Before doing this make sure that all the country names have been changed correctly**. Sometimes Turkey can be missed due to a less percentage in FuzzyMatch.

"Save Mapping" will download a file that shows the organization's country list compared to IFs country list.

A	B
original_name	matched_name
Angola	Angola
Albania	Albania
Armenia	Armenia
Azerbaijan	Azerbaijan
Burundi	Burundi
Benin	Benin
Burkina Faso	Burkina Faso
Bangladesh	Bangladesh
Bulgaria	Bulgaria
Bosnia and Herzegovina	Bosnia and Herzegovina
Belarus	Belarus
Belize	Belize
Barbados	Barbados
Bhutan	Bhutan
Botswana	Botswana
Central African Republic	Central African Republic
China	China
Cote d'Ivoire	Cote D'Ivoire
Cameroon	Cameroon
Congo, Dem. Rep.	Congo, Dem. Republic of the
Congo, Rep.	Congo
Comoros	Comoros
Cabo Verde	Cabo Verde
Djibouti	Djibouti
Algeria	Algeria
Egypt, Arab Rep.	Egypt
Estonia	Estonia
Ethiopia	Ethiopia
Fiji	Fiji

PIP2021Update_mapping

Before inputting the excel file into this GitHub change "original_name" to Country or IFs Country and "matched_name" to the organization's name. For example this is from World Bank's Poverty and Inequality Platform so the name would PIP Country.

If you are using a code to import directly into a db file, here is an example using Python and panadas:

country = pd.read_excel('ITU Country.xlsx') #get this excel from the repo listed above

dt = country.merge(df2, on='entityName', how='left') #entityName is what the list is called in ITU

dt = dt.drop(columns='entityName') #after merging you must drop this column before importing

dt = dt.rename(columns={'IFs Country':'Country'}) #this must be changed before importing as the column name in IFs is Country

Special Cases

Each organization might have special cases due to the nature of their country list. Therefore in the "Notes"/"Data Notes" section of each data source page (example: FAOSTAT Land Use#Notes) contains information necessary to complete the data pulling tasks (example: FAOSTAT Land Use needs disaggregation, notes for Ethiopia, and notes for Kosovo and Serbia).

Proxies

Proxies are used when a certain country does not have data points and estimates are needed. Proxies should be used on a case by case basis and for certain series. Proxies essentially use a similar country to the country being approximated (in terms of population, GDP, etc.) and calculates the scale of what the approximated countries value should be.

An example from IHME's series is Kosovo and Albania. Albania is similar and geographically close to Kosovo making it an ideal for a proxy. For a series that deals with death such as SocietalViolenceDeathsTotal use the following equation:

- Kosovo’s number of deaths = Albania’s number of deaths * (Kosovo’s Population / Albania’s Population)

Disaggregation

A lot of organizations have values for dissolved states such as USSR, Yugoslavia, Czechoslovakia, and more. This then leads to gaps in values for the newly formed such as Serbia, Slovakia, Czechia, etc. Therefore, disaggregation is a great tool to fill in these gaps. Below is the disaggregation steps for common groups:

1. Czechoslovakia = Slovakia and Czech Republic
2. Yugoslav SFR = Slovenia, Croatia, Bosnia and Herzegovina, North Macedonia, Serbia, and

Montenegro

3. Serbia and Montenegro = Serbia and Montenegro
4. Sudan (former) = Sudan and South Sudan
5. USSR = Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine, and Uzbekistan

For disaggregation you are adding up the values for the year after the state disbanded, using the data when the state existed, have the values for the new states after the entity disbanded, and then extrapolate (find the percentage of each state in the data from the OG state and then multiply it to get the right amount) data for the new states in the previous years.

1. USSR add up the values for 1992 (year it disbanded); have totals for 1961-1991; have the values for 15 states for 1992; extrapolate data for 15 states 1961-1991
2. Czechoslovakia add up the values for for 1993 (year it disbanded); have totals for 1961-1992; have the values for 2 states for 1993; extrapolate data for 2 states 1961-1992
3. Sudan (former) add up the values for 2012 (year it disbanded); have totals for 1961-2011; have the values for 2 states for 2012; extrapolate data for 2 states 1961-2011
4. Yugoslav SFR add up the values for 1992 (year it disbanded); have totals for 1961-1991; have the values for 6 states for 1992; extrapolate for 6 states 1961-1991
5. Serbia and Montenegro add up the values for 2006 (year it disbanded); have totals for 1992-2005; have the values for 2 states for 2006; extrapolate data for 2 states 1992-2005

An example of this disaggregation can be found in FAOSTAT Land Use.
