



Standardizing information

- Input: **countries.xml**
 - E.g. **Russia** vs **Russian Federation** in the XML file → We need to standardize the information



Examine preview data

Rows of step: lk language (14 rows)

#	ID	Country_Name	Name	language
1	1	Russia	Mikhail Davydova	<null>
2	2	Russia	Anastasia Davydova	<null>
3	3	Spain	Carmen Rodriguez	Spanish
4	4	Spain	Francisco Delgado	Spanish
5	5	Japan	Natsuki Harada	Japanese
6	6	Japan	Emiko Suzuki	Japanese
7	7	China	Lin Jiang	Chinese
8	8	China	Wei Chiu	Chinese
9	9	United States	Chelsea Thompson	English
10	10	United States	Cassandra Sullivan	English
11	11	Canada	Mackenzie Martin	French
12	12	Canada	Nathan Gauthier	French
13	13	Italy	Giovanni Lombardi	Italian
14	14	Italy	Federica Lombardi	Italian

Close



Standardizing information

1. Add a **Value Mapper** step

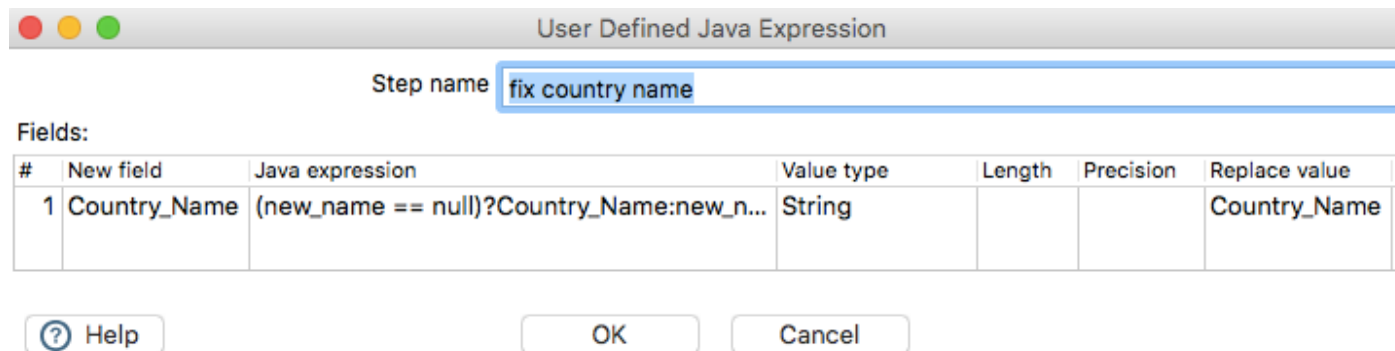
#	Source value	Target value
1	USA	United States
2	U.S.A.	United States
3	Russia	Russian Federation





Standardizing information

2. Add **UDJE** step to overwrite the country names that didn't have the standard value (example Russia → Russian Federation)





Standardizing information

3. Add **Lookup Stream** step

Examine preview data

Rows of step: lk language (14 rows)

#	ID	Country_Name	Name	new_name	language
1	1	Russian Federation	Mikhail Davydova	Russian Federation	Russian
2	2	Russian Federation	Anastasia Davydova	Russian Federation	Russian
3	3	Spain	Carmen Rodriguez	<null>	Spanish
4	4	Spain	Francisco Delgado	<null>	Spanish
5	5	Japan	Natsuki Harada	<null>	Japanese
6	6	Japan	Emiko Suzuki	<null>	Japanese
7	7	China	Lin Jiang	<null>	Chinese
8	8	China	Wei Chiu	<null>	Chinese
9	9	United States	Chelsea Thompson	<null>	English
10	10	United States	Cassandra Sullivan	<null>	English
11	11	Canada	Mackenzie Martin	<null>	French
12	12	Canada	Nathan Gauthier	<null>	French
13	13	Italy	Giovanni Lombardi	<null>	Italian
14	14	Italy	Federica Lombardi	<null>	Italian

Close

