



Stream lookup step

- Data from more sources
- Example: find which language people speak:
 - Stream with People information along with their nationalities
 - A secondary stream with information about the languages ([countries.xml](#))

```
people.csv
people.csv x
1 ID;Country Name;Name
2 1;Russia;Mikhail Davydova
3 2;;Anastasia Davydova
4 3;Spain;Carmen Rodriguez
5 4;;Francisco Delgado
6 5;Japan;Natsuki Harada
7 6;;Emiko Suzuki
8 7;China;Lin Jiang
9 8;;Wei Chiu
10 9;United States;Chelsea Thompson
11 10;;Cassandra Sullivan
12 11;Canada;Mackenzie Martin
13 12;;Nathan Gauthier
14 13;Italy;Giovanni Lombardi
15 14;;Federica Lombardi
```

```
<?xml version="1.0" encoding="UTF-8"?>
<world>
  ...
  <country>
    <name>Russian Federation</name>
    <capital>Moscow</capital>
    <language isofficial="T">
      <name>Russian</name>
      <percentage>86.6</percentage>
    </language>
    <language isofficial="F">
      <name>Tatar</name>
      <percentage>3.2</percentage>
    </language>
    <language isofficial="F">
      <name>Ukrainian</name>
      <percentage>1.3</percentage>
    </language>
  </country>
```



Stream lookup step



1. Read file countries.xml (with countries and languages). **Get data from XML step**

Selected files:

#	File/Directory
1	\${Internal.Entry.Current.Directory}/../files/countries.xml

Step name: Get data from countries.xml

Settings

Loop XPath: /world/country/language

Encoding: UTF-8

Namespace aware?

Ignore comments?

Validate XML?

Use token

Ignore empty file

Do not raise an error if no files

Limit: 0

Prune path to handle large files:

Additional fields

Include filename in output? Filename fieldname:

Rownum in output? Rownum fieldname:

Add to result filename

Add files to result filename

Step name: Get data from countries.xml

#	Name	XPath	Element	Result type	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type	Repeat
1	country	../name	Node	Value of	String							none	N
2	language	name	Node	Value of	String							none	N
3	isofficial	isofficial	Attribute	Value of	String							none	N
4	percentage	percentage	Node	Value of	Number	0.00				.		none	N

Buttons: Help, OK, Preview rows, Cancel



Stream lookup step



2. Add **Filter rows** step, with condition **isofficial=T**

Examine preview data

Rows of step: Filter official languages (237 rows)

#	country	language	isofficial	percentage
1	Afghanistan	Pashto	T	52.40
2	Afghanistan	Dari	T	32.10
3	Albania	Albaniana	T	97.90
4	Algeria	Arabic	T	86.00
5	American Samoa	Samoan	T	90.60
6	American Samoa	English	T	3.10
7	Andorra	Catalan	T	32.30
8	Anguilla	English	T	0.00
9	Antigua and Barbuda	English	T	0.00
10	Argentina	Spanish	T	96.80
11	Armenia	Armenian	T	93.40



Stream lookup step



3. Add Text file input step

Step name: people.csv

File Content Error Handling Filters Fields Additional output fields

Filetype: CSV

Separator: ; Insert TAB

Enclosure: "

Allow breaks in enclosed fields?

Escape:

Header Number of header lines: 1

Footer Number of footer lines: 1

Wrapped lines? Number of times wrapped: 1

Paged layout (printout)? Number of lines per page: 80

Compress:

No empty rows:

Include filename in output:

Rownum in output:

Text file input

Step name: people.csv

File Content Error Handling Filters Fields Additional output fields

#	Name	Type	Format	Position	Length	Precision	Currency	Decimal	Group	Null if	Default	Trim type	Repeat
1	ID	String										none	N
2	Country_Name	String										none	Y
3	Name	String										none	N

Stream lookup step

4. Add Stream lookup step

Stream Value Lookup

Step name: lk language

Lookup step: Filter official languages

The key(s) to look up the value(s):

#	Field	LookupField
1	Country_Name	country

Specify the fields to retrieve:

#	Field	New name	Default	Type
1	language			String

Preserve memory (costs CPU)

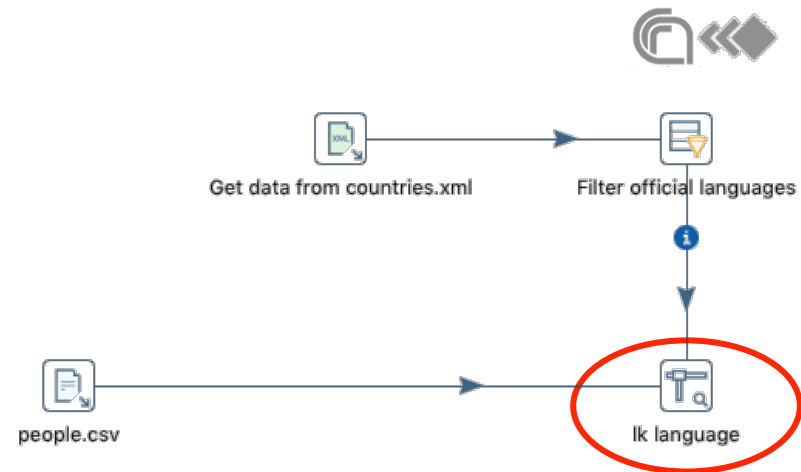
Key and value are exactly one i

Use sorted list (i.s.o. hashtable)

Help Cancel Get Fields Get lookup

This selection is for telling PDI which of the incoming streams is the stream used to look up

the names of the fields that you want back as result of the lookup



Stream lookup step

- Preview

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



- For the rows for which the lookup is successful, the values for the new fields will be taken from the lookup stream
- For the others, the fields will remain null, unless you set a default value