

Sometimes you get a file...

Checking data
and finding the gaps in Wikidata
with OpenRefine

User:A_ka_es



- **explore the data**
- **compare and analyze**
- **match the rows with wikidata**
- **recheck the data**
- **combine projects**
- **find the gaps**
- **enrich the wikidata-items**

explore

Firefox Developer Edition File Edit View History Bookmarks Tools Window Help

OpenRefine x +

127.0.0.1:3333 Search

 **OpenRefine** *A power tool for working with messy data.*

Create Project

Open Project

Import Project

Language Settings

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.

Get data from

Locate one or more files on your computer to upload:

This Computer

Browse...

No files selected.

Next »

[Web Addresses \(URLs\)](#)

[Clipboard](#)

[Database](#)

[Google Data](#)



Version 3.2 [55c921b]

[Preferences](#)

[Help](#)

[About](#)

explore

Firefox Developer Edition File Edit View History Bookmarks Tools Window Help
EP_info_files EP 2019 07 19 csv x OpenRefine x +
127.0.0.1:3333 Search

OpenRefine A power tool for working with messy data.

Create Project

Open Project

Import Project

Language Settings

« Start Over Select Files to Import

Configure Parsing Options »

There are several files available. Please select the ones to import.

0 of 65 files selected

Select All

Unselect All

Select by Extension

.csv 63 files

Select

Unselect

.html 1 file

Select

Unselect

.txt 1 file

Select

Unselect

Select by Regex on File Names

Select

Unselect

Import?	Name	Media type	Format	Size
<input type="checkbox"/>	EP_info_files/European-Parliament-2019-07-19.html	unknown	unknown	2 MB
<input type="checkbox"/>	EP_info_files/notes.txt	unknown	text/line-based	2.3 KB
<input type="checkbox"/>	EP_info_files/EP-2019-07-19.csv	unknown	text/line-based/*sv	189.5 KB
<input type="checkbox"/>	EP_info_files/committees/DEVE.csv	unknown	text/line-based/*sv	14.2 KB
<input type="checkbox"/>	EP_info_files/committees/CPCO.csv	unknown	text/line-based/*sv	6.4 KB
<input type="checkbox"/>	EP_info_files/committees/INTA.csv	unknown	text/line-based/*sv	21.2 KB
<input type="checkbox"/>	EP_info_files/committees/ENVI.csv	unknown	text/line-based/*sv	38.4 KB
<input type="checkbox"/>	EP_info_files/committees/ITRE.csv	unknown	text/line-based/*sv	36.8 KB
<input type="checkbox"/>	EP_info_files/committees/TRAN.csv	unknown	text/line-based/*sv	25 KB
<input type="checkbox"/>	EP_info_files/committees/CONT.csv	unknown	text/line-based/*sv	14.8 KB
<input type="checkbox"/>	EP_info_files/committees/IMCO.csv	unknown	text/line-based/*sv	22.5 KB
<input type="checkbox"/>	EP_info_files/committees/JURI.csv	unknown	text/line-based/*sv	12.7 KB
<input type="checkbox"/>	EP_info_files/committees/SEDE.csv	unknown	text/line-based/*sv	14.8 KB
<input type="checkbox"/>	EP_info_files/committees/CULT.csv	unknown	text/line-based/*sv	15.7 KB
<input type="checkbox"/>	EP_info_files/committees/PETI.csv	unknown	text/line-based/*sv	15.4 KB
<input type="checkbox"/>	EP_info_files/committees/ECON.csv	unknown	text/line-based/*sv	29.5 KB
<input type="checkbox"/>	EP_info_files/committees/AFCO.csv	unknown	text/line-based/*sv	15.1 KB
<input type="checkbox"/>	EP_info_files/committees/BCPR.csv	unknown	text/line-based/*sv	2.9 KB
<input type="checkbox"/>	EP_info_files/committees/REGI.csv	unknown	text/line-based/*sv	20.8 KB
<input type="checkbox"/>	EP_info_files/committees/PECH.csv	unknown	text/line-based/*sv	14.3 KB
<input type="checkbox"/>	EP_info_files/committees/LIBE.csv	unknown	text/line-based/*sv	35.3 KB
<input type="checkbox"/>	EP_info_files/committees/BURO.csv	unknown	text/line-based/*sv	5.4 KB
<input type="checkbox"/>	EP_info_files/committees/AFET.csv	unknown	text/line-based/*sv	36.5 KB
<input type="checkbox"/>	EP_info_files/committees/FEMM.csv	unknown	text/line-based/*sv	17.7 KB
<input type="checkbox"/>	EP_info_files/committees/QUE.csv	unknown	text/line-based/*sv	1.2 KB
<input type="checkbox"/>	EP_info_files/committees/BUDG.csv	unknown	text/line-based/*sv	20.3 KB
<input type="checkbox"/>	EP_info_files/committees/EMPL.csv	unknown	text/line-based/*sv	27.2 KB



Version 3.2 [55c921b]

Preferences

Help

About

compare and analyze

Firefox Developer Edition | File Edit View History Bookmarks Tools Window Help | 127.0.0.1:3333/project?project=1722156626755

OpenRefine EP_committees_csv | Permalink | Open... Export Help

Facet / Filter | Undo / Redo 0 / 0 | 1945 rows | Extensions: Wikidata

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

All	File	MEP	President	Vice-President	Chair	Vice-Chair	Member	Substitute	Country	EP_group	Group_position	Em
★	1. EP_info_files/committees /DEVE.csv	Tomas TOBÉ			DEVE		CPCO d-il dmed LIBE	d-tr ITRE	Sweden	PPE		tomas.t
★	2. EP_info_files/committees /DEVE.csv	Erik MARQUARDT				DEVE	d-af dmas	d-iq dmed LIBE	Germany	Greens		erik.ma
★	3. EP_info_files/committees /DEVE.csv	Chrysoula ZACHAROPOULOU				DEVE	dacp	d-cl dlat FEMM	France	Renew		chryso
★	4. EP_info_files/committees /DEVE.csv	Pierrette HERZBERGER-FOFANA				DEVE	dcar dpap	dacp FEMM	Germany	Greens		pierrett fofana
★	5. EP_info_files/committees /DEVE.csv	Norbert NEUSER				DEVE	dacp	AFET dase DROI	Germany	S&D		norbert
★	6. EP_info_files/committees /DEVE.csv	Bernhard ZIMNIOK					dacp DEVE dmas dmed	AFET dcas	Germany	Identity		bernhai
★	7. EP_info_files/committees /DEVE.csv	György HÖLVÉNYI					dacp DEVE DROI	darp	Hungary	PPE		gyorgy.
★	8. EP_info_files/committees /DEVE.csv	Lukas MANDL				SEDE	AFET DEVE dsee	dacp EMPL	Austria	PPE		lukas.m
★	9. EP_info_files/committees /DEVE.csv	Miguel URBÁN CRESPO					DEVE dlat dmer DROI	d-br LIBE	Spain	GUE		miguel.
★	10. EP_info_files/committees /DEVE.csv	Martin HORWOOD					d-ir DEVE	AFET d-ru dnat SEDE	United Kingdom	Renew	Vice-Chair	martin.f

match with wikidata

Firefox Developer Edition File Edit View History Bookmarks Tools Window Help

EP_info_files EP 2019 07 19 csv X

127.0.0.1:3333/project?project=1605134250829

OpenRefine EP_info_files EP 2019 07 19 csv Permalink

Open... Export Help

Facet / Filter Undo / Redo 2 / 2

747 rows

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 25 next > last »

Extensions: Wikidata

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

	All	MEP	President	Vice-President	Chair	Vice-Chair	Member	Substitute	Country	EP_group	Group_position	Email 1
1.	Abir AL-SAHLANI						d-iq EMPL	dsee LIBE	Sweden	Renew		abir.alsahlani@europarl.europa.eu
2.	Adam BIELAN						dase IMCO	AFET darp	Poland	ECR		adam.bielan@europarl.europa.eu
3.	Adam JARUBAS						d-jp ENVI	BUDG depa dsca ITRE	Poland	PPE		adam.jarubas@europarl.europa.eu
4.	Adina-Ioana VĂLEAN				ITRE		CPCO d-us	depa dsca ENVI	Romania	PPE		adinaioana.valean@europarl.europa.eu
5.	Adrian-Dragoș BENEĂ						AGRI d-md REGI	d-ua PETI	Romania	S&D		adrian-dragos.benea@europarl.europa.eu
6.	Adriana MALDONADO LÓPEZ						dmag dmed IMCO	d-us ITRE	Spain	S&D		adriana.maldonado@europarl.europa.eu
7.	Agnes JONGERIUS						dase EMPL	dacp dpap INTA	Netherlands	S&D		agnes.jongerus@europarl.europa.eu
8.	Agnès EVREN						d-af dmed dpal ENVI PETI	ECON	France	PPE		agnes.evren@europarl.europa.eu
9.	Aileen McLEOD						AFCO deea ENVI		United Kingdom	Greens		aileen.mcleod@europarl.europa.eu
10.	Aldo PATRICIELLO						d-in ITRE	d-ru ENVI	Italy	PPE		aldo.patriciello@europarl.europa.eu
11.	Alessandra BASSO						dacp IMCO	d-cn DEVE	Italy	Identity		alessandra.basso@europarl.europa.eu
12.	Alessandra MORETTI						d-jp ENVI	dcas FEMM TRAN	Italy	S&D		alessandra.moretti@europarl.europa.eu
13.	Alessandro PANZA						deea REGI	CULT depa	Italy	Identity		alessandro.panza@europarl.europa.eu
14.	Alex AGIUS SALIBA						dacp dmed IMCO PETI	dpal EMPL	Malta	S&D		alex.agiussaliba@europarl.europa.eu
15.	Alexander ALEXANDROV YORDANOV						AFET d-mk	AFCO d-md depa EMPL	Bulgaria	PPE		alexander.yordanov@europarl.europa.eu
16.	Alexander BERNHUBER						depa dsca ENVI PETI	CULT d-m FEMM	Austria	PPE		alexander.bernhuber@europarl.europa.eu
17.	Alexandr VONDRA						d-us ENVI	d-il LIBE	Czechia	ECR		alexandr.vondra@europarl.europa.eu

match with wikidata

OpenRefine EP_info_files EP 2019 07 19 csv [Permalink](#) Open... Export Help

Facet / Filter Undo / Redo 238 / 238 Extensions: Wikidata

Refresh Reset All Remove All Show as: **rows** records Show: 5 10 25 50 rows « first < previous 1 - 4 next > last »

MEP: judgment change
1 choices Sort by: name count
none 4
[Facet by choice counts](#)

MEP: best candidate's score change reset

 Numeric Non-numeric Blank Error
2 0 2 0

Flagged Rows change invert reset
2 choices Sort by: name count
false 743
true 4 exclude
[Facet by choice counts](#)

Country change
2 choices Sort by: name count Cluster
Greece 2
Netherlands 2
[Facet by choice counts](#)

4 matching rows (747 total)

All	MEP	President	Vice-President	Chair	Vice-Chair	Member	Substitute	Country	EP_group	Group_position	Email 1
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

73. Athanasios KONSTANTINOU
 Create new item
[Search for match](#)

410. Lefteris NIKOLAOU-ALAVANOS
 Create new item
[Search for match](#)

707. Tom BERENDSEN
 Tom Berendsen (50)
 Create new item
[Search for match](#)

727. Vera TAX
 Vera Tax (50)
 Create new item
[Search for match](#)

recheck with wikidata

Firefox Developer Edition | File | Edit | View | History | Bookmarks | Tools | Window | Help | Thu 15. Aug 10:59

EP_info_files_EP-2019-07-19-cs X | 127.0.0.1:3333/project?project=1695799956492 | 90% | Search

OpenRefine EP_info_files_EP-2019-07-19-csv_v2 | Permalink | Open... | Export | Help

Facet / Filter | Undo / Redo 241 / 242 | 747 records | Extensions: Wikidata

Show as: rows records | Show: 5 10 25 50 records | « first < previous 1 - 10 next > last »

Using facets and filters
Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.
Not sure how to get started? Watch these screencasts

	All	MEP	new_Item	President	Vice-President	Chair	Vice-Chair	Member	Substitute	Country	EP_group	Group_position	Email 1	Email 2
1.	Abir Al-Sahlani	Choose new match						d-iq EMPL	dsee LIBE	Sweden	Renew		abir.alsahlani@europarl.europa.eu	
2.	Adam Bielan	Choose new match						dase IMCO	AFET darp	Poland	ECR		adam.bielan@europarl.europa.eu	
3.	Adam Jarubas	Choose new match						d-jp ENVI	BUDG depa dsca ITRE	Poland	PPE		adam.jarubas@europarl.europa.eu	
4.	Adina-Ioana Vălean	Choose new match				ITRE		CPCO d-us	depa dsca ENVI	Romania	PPE		adinaioana.valean@europarl.europa.eu	
5.	Dragoș Benea	Choose new match						AGRI d-md REGI	d-ua PETI	Romania	S&D		adrian-dragos.benea@europarl.europa.eu	
6.	Adriana Maldonado López	Choose new match						dmag dmed IMCO	d-us ITRE	Spain	S&D		adriana.maldonado@europarl.europa.eu	
7.	Agnes Jongerius	Choose new match						dase EMPL	dacp dpap INTA	Netherlands	S&D		agnes.jongerius@europarl.europa.eu	
8.	Agnès Evren	Choose new match						d-af dmed dpal ENVI PETI	ECON	France	PPE		agnes.evren@europarl.europa.eu	
9.	Aileen McLeod	Choose new match						AFCO deea ENVI		United Kingdom	Greens		aileen.mcleod@europarl.europa.eu	
10.	Aldo Patriciello	Choose new match						d-in ITR edit	d-ru ENVI	Italy	PPE		aldo.patriciello@europarl.europa.eu	

recheck with wikidata

Firefox Developer Edition | File | Edit | View | History | Bookmarks | Tools | Window | Help | Thu 15. Aug 13:26

EP_info_files EP 2019 07 19 csv X | erforderlich - Englisch-Überset X | +

127.0.0.1:3333/project?project=1605134250829 | Search

OpenRefine EP_info_files EP 2019 07 19 csv | Permalink | Open... | Export | Help

Facet / Filter | Undo / Redo 243 / 243 | 747 rows | Extensions: Wikidata

Show as: rows records | Show: 5 10 25 50 rows | « first < previous 1 - 10 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

All	MEP	MEP directory ID	Twitter usernam	new_item	President	Vice-President	Chair	Vice-Chair	Member	Substitute	Country	EP_group
1.	Abir Al-Sahlani <small>Choose new match</small>								d-iq EMPL	dsee LIBE	Sweden	Renew
2.	Adam Bielan <small>Choose new match</small>	23788	AdamBielan						dase IMCO	AFET darp	Poland	ECR
3.	Adam Jarubas <small>Choose new match</small>	197517							d-jp ENVI	BUDG depa dsca ITRE	Poland	PPE
4.	Adina-Ioana Vălean <small>Choose new match</small>	37324	AdinaValean				ITRE		CPCO d-us	depa dsca ENVI	Romania	PPE
5.	Dragoș Benea <small>Choose new match</small>								AGRI d-md REGI	d-ua PETI	Romania	S&D
6.	Adriana Maldonado López <small>Choose new match</small>	197735	_AMaldonado_						dmag dmed IMCO	d-us ITRE	Spain	S&D
7.	Agnes Jongerius <small>Choose new match</small>	125021	a_jongerius		edit				dase EMPL	dacp dpap INTA	Netherlands	S&D
8.	Agnès Evren <small>Choose new match</small>	197535	AgnesEvren						d-af dmed dpal ENVI PETI	ECON	France	PPE
9.	Aileen McLeod <small>Choose new match</small>	30490	AileenMcLeodSNP						AFCO deea ENVI		United Kingdom	Greens
10.	Aldo Patriciello <small>Choose new match</small>	36392	PatricielloAldo						d-in ITRE	d-ru ENVI	Italy	PPE

compare

- via Twitter (transform the data)
- via political party
- manually, using domain knowledge
- via different Wikipedias

result:

- 744 okay
- 3 new entries

\o/

find the gaps via SPARQL and OpenRefine

Firefox Developer Edition | File | Edit | View | History | Bookmarks | Tools | Window | Help | Fri 16. Aug 12:06

2019_08_16_wikidata_query - O | EP_info_files EP 2019 07 19 csv | tmp - OpenRefine | Wikidata Query Service

https://query.wikidata.org/#SELECT %3Fmep %3FmepLabel %3FtermLabel %

Wikidata Query Service | Examples | Help | More tools | English

```

1 SELECT ?mep ?mepLabel ?termLabel ?startTime ?endTime ?parGroup ?parGroupLabel WHERE {
2   ?mep wdt:P31 wd:Q5;
3     p:P39 ?statement.
4   ?statement ps:P39 wd:Q27169;
5     pq:P2937 ?term;
6     pq:P580 ?startTime;
7     pq:P4100 ?parGroup.
8   OPTIONAL {
9     ?statement pq:P582 ?endTime.
10  }
11  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
12 }
13 ORDER BY ?startTime ?mepLabel
  
```

7121 results in 5404 ms | Code | Download | Link

Search

mep	mepLabel	termLabel	startTime	endTime	parGroup	parGroupLabel
Q87485	Martin Sonneborn	Ninth European Parliament	t1369968799		Q1440626	Non-Inscrits
Q435867	Eileen Desmond	First European Parliament	16 July 1979	7 July 1981	Q51137669	Socialist Group
Q1687809	Jerry Cronin	First European Parliament	16 July 1979	23 July 1984	Q5413043	European Progressive Democrats
Q6211198	Joe McCartin	First European Parliament	16 July 1979	23 July 1984	Q635616	European People's Party group
Q193630	John Hume	First European Parliament	16 July 1979	23 July 1984	Q51137669	Socialist Group

find the gaps via SPARQL and OpenRefine

Firefox Developer Edition | File Edit View History Bookmarks Tools Window Help | 127.0.0.1:3333/project?project=1998442345221

OpenRefine wikidata_query | Permalink | Open... Export Help

Facet / Filter | Undo / Redo 37 / 37 | 110 matching rows (3749 total) | Extensions: Wikidata

Refresh | Reset All | Remove All | Show as: rows records | Show: 5 10 25 50 rows | Sort

Facet: ninth_parliament | change invert reset | 2 choices Sort by: name count | Cluster | false 3639 | true 110 | exclude | Facet by choice counts

All	mep	wikidata_q_num	mepLabel	parGroup	termLabel	ninth_parliamen	startTime	parGroupLat
<input type="checkbox"/>	<input type="checkbox"/>	121. http://www.wikidata.org/entity/Q114667	Sarah Wiener	http://www.wikidata.org/entity/Q751935	Ninth European Parliament	true	2019-07-02T00:00:00Z	The Greens–European Free Alliance
<input type="checkbox"/>	<input type="checkbox"/>	276. http://www.wikidata.org/entity/Q12028942	Kateřina Konečná	http://www.wikidata.org/entity/Q753711	Ninth European Parliament Eighth European Parliament Fifth European Parliament	true	2019-07-02T00:00:00Z	European United Left–Nordic Greer Left
<input type="checkbox"/>	<input type="checkbox"/>	278. http://www.wikidata.org/entity/Q12036669	Michaela Šojdrová	http://www.wikidata.org/entity/Q635616	Ninth European Parliament Eighth European Parliament	true	2019-07-02T00:00:00Z	European People' Party group
<input type="checkbox"/>	<input type="checkbox"/>	286. http://www.wikidata.org/entity/Q12055995	Stanislav Polčák	http://www.wikidata.org/entity/Q635616	Ninth European Parliament Eighth European Parliament	true	2019-07-02T00:00:00Z	European People' Party group
<input type="checkbox"/>	<input type="checkbox"/>	358. http://www.wikidata.org/entity/Q12726340	Dan Motreanu	http://www.wikidata.org/entity/Q635616	Ninth European Parliament	true	2019-07-02T00:00:00Z	European People' Party group
<input type="checkbox"/>	<input type="checkbox"/>	359. http://www.wikidata.org/entity/Q12726380	Daniel Buda	http://www.wikidata.org/entity/Q635616	Ninth European Parliament Eighth European Parliament	true	2019-07-02T00:00:00Z	European People' Party group
<input type="checkbox"/>	<input type="checkbox"/>	363. http://www.wikidata.org/entity/Q12735517	Mihai Tudose	http://www.wikidata.org/entity/Q507343	Ninth European Parliament	true	2019-07-02T00:00:00Z	Progressive Allian of Socialists and Democrats
<input type="checkbox"/>	<input type="checkbox"/>	369. http://www.wikidata.org/entity/Q12742616	Tudor Ciuhodaru	http://www.wikidata.org/entity/Q507343	Ninth European Parliament	true	2019-07-02T00:00:00Z	Progressive Allian of Socialists and Democrats
<input type="checkbox"/>	<input type="checkbox"/>	486. http://www.wikidata.org/entity/Q1381218	Seán Kelly	http://www.wikidata.org/entity/Q635616	Ninth European Parliament Eighth European Parliament Seventh European Parliament	true	2019-07-01T00:00:00Z	European People' Party group
<input type="checkbox"/>	<input type="checkbox"/>	549. http://www.wikidata.org/entity/Q14541202	Günther Sidl	http://www.wikidata.org/entity/Q507343	Ninth European Parliament	true	2019-07-02T00:00:00Z	Progressive Allian of Socialists and Democrats

connect the projects an enrich data

Firefox Developer Edition | File Edit View History Bookmarks Tools Window Help | Sat 17. Aug 16:29

EP_info_files EP 2019 07 19 csv | wikidata_query_20_30 - OpenRefine

127.0.0.1:3333/project?project=1605134250829

OpenRefine EP_info_files EP 2019 07 19 csv | Permalink | Open... | Export | Help

Facet / Filter | Undo / Redo 859 / 859 | 747 rows | Schema | Issues 1 | Preview | Extensions: Wikidata

Show as: rows records | Show: 5 10 25 50 rows | « first < previous 1 - 10 next > last »

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
[Watch these screencasts](#)

	All	MEP	verified	original_name	wikidata_Q_num	status_ID	wikidata_MEP_c	ID	status_parliame	wikidata_parliam	status_parliame	wikidata_parliam
1.	Abir Al-Sahlani	okay	Abir AL-SAHLANI	Q4118653	true		197400	false				
2.	Adam Bielan	okay	Adam BIELAN	Q2507571	true	23788	23788	false	Seventh European Parliament Sixth European Parliament Fifth European Parliament	false	European Conservatives and Reformists	Choose new match
3.	Adam Jarubas	okay	Adam JARUBAS	Q9139694	true	197517	197517	false				
4.	Adina-ioana Vălean	okay	Adina-Ioana VĂLEAN	Q274739	true	37324	37324	true	Ninth European Parliament Eighth European Parliament Seventh European Parliament Sixth European Parliament	true	European People's Party group	Choose new match
5.	Adrian-dragoș Benea	okay	Adrian-Dragoș BENEA	Q28035943	true		197658	true	Ninth European Parliament	true	Progressive Alliance of Socialists and Democrats	Choose new match
6.	Adriana Maldonado López	okay	Adriana MALDONADO LÓPEZ	Q64091813	true	197735	197735	false				
7.	Agnes Jongerius	okay	Agnes JONGERIUS	Q2745560	true	125021	125021	false	Eighth European Parliament	false	Progressive Alliance of Socialists and Democrats	Choose new match
8.	Agnès Evren	okay	Agnès EVREN	Q33109862	true	197535	197535	false				

data cleaning is a circle

Firefox Developer Edition | File Edit View History Bookmarks Tools Window Help | Sat 17. Aug 01:51 | 127.0.0.1:3333/project?project=2215802419267

OpenRefine query 8 json | Permalink | Open... Export Help

Facet / Filter | Undo / Redo 0 / 0 | 751 matching rows (7171 total) | Extensions: Wikidata

Refresh | Reset All | Remove All | Show as: rows records | Show: 5 10 25 50 rows | « first < previous 1 - 10 next > last »

_ - termLabel | change invert reset | 17 choices | Sort by: name count | Cluster

- Eighth European Parliament 911
- Fifth European Parliament 874
- First European Parliament 550
- Fourth European Parliament 750
- Ninth European Parliament 751** | exclude
- P16836722 48
- P17315694 20
- P17315702 35
- P17315703 30
- P17315704 28
- P17315706 30

_ - endTime | change | 3 choices | Sort by: name count | Cluster

- 2019-07-02T00:00:00Z 2
- 2019-07-06T00:00:00Z 1
- 2019-07-18T00:00:00Z 1
- (blank) 747

_ - startTime | change | 4 choices | Sort by: name count | Cluster

- 2019-07-01T00:00:00Z 12
- 2019-07-02T00:00:00Z 737
- 2019-07-04T00:00:00Z 1
- 2019-07-10T00:00:00Z 1

All	_ - mep	_ - mepLabel	_ - termLabel	_ - startTime	_ - id	_ - endTime
6421.	http://www.wikidata.org/entity/Q559402	Billy Kelleher	Ninth European Parliament	2019-07-01T00:00:00Z	197818	
6422.	http://www.wikidata.org/entity/Q380310	Ciarán Cuffe	Ninth European Parliament	2019-07-01T00:00:00Z	197654	
6423.	http://www.wikidata.org/entity/Q2975532	Clare Daly	Ninth European Parliament	2019-07-01T00:00:00Z	197731	
6424.	http://www.wikidata.org/entity/Q540812	Frances Fitzgerald	Ninth European Parliament	2019-07-01T00:00:00Z	197720	
6425.	http://www.wikidata.org/entity/Q16866329	Grace O'Sullivan	Ninth European Parliament	2019-07-01T00:00:00Z	197889	
6426.	http://www.wikidata.org/entity/Q64034615	Isabel Wiseler-Santos Lima	Ninth European Parliament	2019-07-01T00:00:00Z	197421	
6427.	http://www.wikidata.org/entity/Q3778169	Luke 'Ming' Flanagan	Ninth European Parliament	2019-07-01T00:00:00Z	124985	
6428.	http://www.wikidata.org/entity/Q467633	Mairead McGuinness	Ninth European Parliament	2019-07-01T00:00:00Z	28115	
6429.	http://www.wikidata.org/entity/Q63978256	Maria Walsh	Ninth European Parliament	2019-07-01T00:00:00Z	197863	
6430.	http://www.wikidata.org/entity/Q16979858	Matt Carthy	Ninth European Parliament	2019-07-01T00:00:00Z	124986	

result

Firefox Developer Edition | File Edit View History Bookmarks Tools Window Help | Sat 17. Aug 10:42 | 127.0.0.1:3333/project?project=1867593029113

OpenRefine query 11 json | Permalink | Open... | Export | Help

Facet / Filter | Undo / Redo 0 / 0 | 751 matching rows (7171 total) | Extensions: Wikidata

Refresh | Reset All | Remove All | Show as: rows records | Show: 5 10 25 50 rows | « first < previous 1 - 10 next > last »

_ - termLabel | change | invert | reset | Cluster

17 choices | Sort by: name count

- Achtes Europäisches Parlament 911
- Drittes Europäisches Parlament 612
- Erstes Europäisches Parlament 550
- Fünftes Europäisches Parlament 874
- Neuntes Europäisches Parlament** 751 | exclude
- P16836722 48
- P17315694 20
- P17315702 35
- P17315703 30
- P17315704 28
- P17315706 30

_ - startTime | change | Cluster

3 choices | Sort by: name count

- 2019-07-02T00:00:00Z 749
- 2019-07-04T00:00:00Z 1
- 2019-07-10T00:00:00Z 1

Facet by choice counts

_ - endTime | change | Cluster

3 choices | Sort by: name count

- 2019-07-02T00:00:00Z 2
- 2019-07-06T00:00:00Z 1
- 2019-07-18T00:00:00Z 1
- (blank) 747

Facet by choice counts

All	_ - mep	_ - mepLabel	_ - termLabel	_ - startTime	_ - id	_ - endTime
6421.	http://www.wikidata.org/entity/Q4118653	Abir Al-Sahlani	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	197400	
6422.	http://www.wikidata.org/entity/Q2507571	Adam Bielan	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	23788	
6423.	http://www.wikidata.org/entity/Q9139694	Adam Jarubas	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	197517	
6424.	http://www.wikidata.org/entity/Q274739	Adina Vălean	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	37324	
6425.	http://www.wikidata.org/entity/Q2745560	Agnes Jongerius	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	125021	
6426.	http://www.wikidata.org/entity/Q33109862	Agnès Evren	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	197535	
6427.	http://www.wikidata.org/entity/Q405025	Aileen McLeod	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	30490	
6428.	http://www.wikidata.org/entity/Q1077655	Aldo Patriciello	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	36392	
6429.	http://www.wikidata.org/entity/Q2640766	Aleksandar Jordanow	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	197836	
6430.	http://www.wikidata.org/entity/Q15055614	Alessandra Moretti	Neuntes Europäisches Parlament	2019-07-02T00:00:00Z	124799	

javascript: {}

conclusion

OpenRefine is a powerful tool for working with data, not only an upload tool for Wikidata.

We should

- spread the word
- develop services and modules to make some steps easier and safer for others
- we should really use it to improve the quality for our data