

# Documentation for Eve's packages

November 19, 2023

## Contents

<b>1</b>	<b>Instructions for uploading to PoolParty from MDR staging</b>	<b>1</b>
1.1	Download the tabular data from MDR staging . . . . .	1
1.2	Processing of the basket csv to be ready for PoolParty . . . . .	4
1.2.1	Windows . . . . .	4
1.2.2	Linux/Mac . . . . .	4
1.3	Uploading the csv to PoolParty . . . . .	4
<b>2</b>	<b>Instructions for spotting similar questions using bag of words</b>	<b>5</b>
2.0.1	Windows . . . . .	5
2.0.2	Linux/Mac . . . . .	6

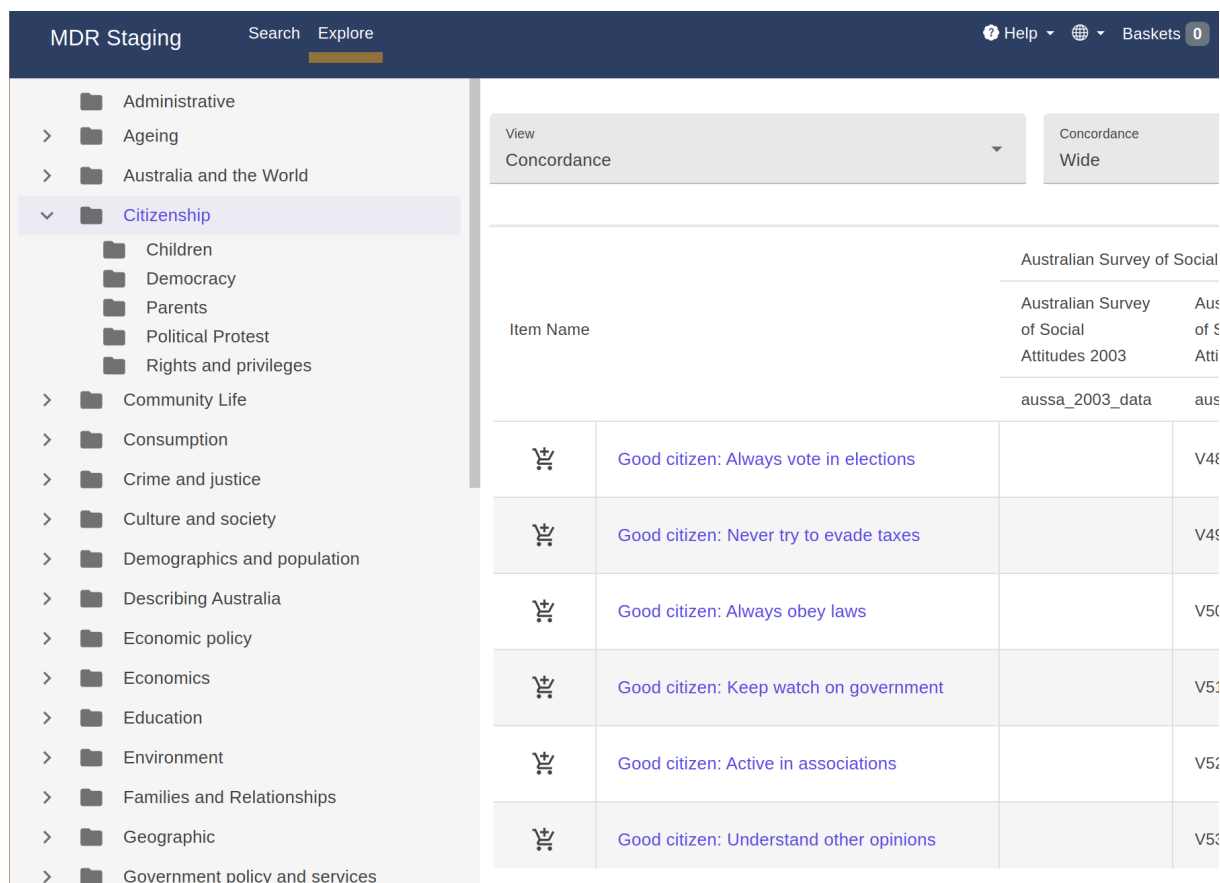


Figure 1: Step 2

## 1 Instructions for uploading to PoolParty from MDR staging

This section explains how to download the data from MDR staging and then upload them to PoolParty.

### 1.1 Download the tabular data from MDR staging

Step 1: create a new basket in your account and go to the survey of interest.

Step 2: select the vocabularies using the shopping carts (see Figure 1)

Step 3: Go to your basket and click download metadata, and select csv as the option (see Figure 2).

The csv file will look something like Fig 3.

## Basket

Variables (69)		
<div>Download Metadata ▾</div> <div> <div>PDF Codebook</div> <div>CSV</div> <div>Excel</div> <div>DDI 3.2 XML</div> </div>		
	MODE	Mode of survey completion <a href="#">AUSSI-ESS</a> > <a href="#">AuSSI-ESS 2020 Published</a> > <a href="#">AuSSI-ESS_2020_Published</a>
	SEX	SEX. What sex are you? <a href="#">AUSSI-ESS</a> > <a href="#">AuSSI-ESS 2020 Published</a> > <a href="#">AuSSI-ESS_2020_Published</a>
PPLTRST		A4. Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? <a href="#">AUSSI-ESS</a> > <a href="#">AuSSI-ESS 2020 Published</a> > <a href="#">AuSSI-ESS_2020_Published</a>
NETUSTM		A3. On a typical day, about how much time do you spend using the internet on a computer, tablet, smartphone or other device, whether for work or personal use? (in minutes) <a href="#">AUSSI-ESS</a> > <a href="#">AuSSI-ESS 2020 Published</a> > <a href="#">AuSSI-ESS_2020_Published</a>
NETUSOFT		A2. People can use the internet on different devices such as computers, tablets and smartphones. How often do you use the internet on these or any other devices, whether for work or personal use? <a href="#">AUSSI-ESS</a> > <a href="#">AuSSI-ESS 2020 Published</a> > <a href="#">AuSSI-ESS_2020_Published</a>
NWSPOL		A1. On a typical day, about how much time do you spend watching, reading, or listening to news about politics and current affairs? (in minutes) <a href="#">AUSSI-ESS</a> > <a href="#">AuSSI-ESS 2020 Published</a> > <a href="#">AuSSI-ESS_2020_Published</a>

Figure 2: Step 3

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Name Label Type Codes Description Study Dataset												
2	MODE "Mode of survey completion" Code "1"	Online   2	Phone	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published									
3	SEX "SEX. What sex are you?" Code "1"	Male   2	Female   3	Refused   4	Don't know	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published							
4	PPLTRST "A4. Generally speaking	would you say that most people can be trusted or that you can't	0 - You can't be	1   3	2   4	3   5	4   6	5   7	6   8	7   9	8   10	9   11	
5	NETUSTM "A3. On a typical day	about how much time do you spend using the	tablet	smartphone or o whether for work	5   2	9   3	10   4	15   5	20   6	30   7	38   8	40   9	
6	NETUSOFT "A2. People can use the internet on different device	tablets and smartphones. How often do you use whether for work	Never   2	Only occasional	A few times a wk	Most days   5	Every day   6	Refused   7	Don't know	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_P			
7	NWSPOL "A1. On a typical day	about how much time do you spend watching	reading	or listening to ne	0   2	1   3	2   4	3   5	5   6	7   7	10   8	15   9	20   10
8	POLNTR "B1. How interested would you say you are in politics	Very interested   2	Quite interested	Hardly interested	Not at all interest	Refused   6	Don't know	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published					
9	PSPPSOVA "B2. How much would you say the political system	Not at all   2	Very little   3	Some   4	A lot   5	A great deal   6	Refused   7	Don't know	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published				
10	HMSACLD "B36. Agree or disagree - Gay male and lesbian co	Agree strongly   2	Agree   3	Neither agree n	Disagree   5	Disagree strongl	Refused   7	Don't know	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published				
11	CTZSHIPC "C21. What citizenship do you hold?" Code "14	Australia   48	China   56	Don't know   77	United Kingdom	Indonesia   105	India   158	Malaysia   171	New Zealand   1	Pakistan   189	Refused   888	Others	"AuSSI-ESS 2020 P
12	BRNCNTR "C22. Were you born in Australia?" Code "1"	Yes   2	No   3	Refused   4	Don't know	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published							
13	EDULVLBP "F44. What is the highest level of education your h	wife	or partner has s	Less than Grade	Grade / Year 6 to Year 10 to Year	Certificate I or II	Year 12   6	Certificate III   7	Certificate IV   8	Associate Degree	Diploma or Adv	Bachelor's	
14	P_GENDER "What is your gender?" Code "1"	Refused   2	Don't know   3	Not stated / Unk	Male   5	Female   6	Other	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published					
15	P_AGE "Age as of 29th February 2020" Code "1"	Unknown   2	18   3	19   4	20   5	21   6	22   7	23   8	24   9	25   10	26   11	27   12	28   13
16	P_AGE_GROUP "Age group as of 29th February 2020" Code "	Unknown   2	18-24 years   3	25-34 years   4	35-44 years   5	45-54 years   6	55-64 years   7	65-74 years   8	75 or more years	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published			
17	D_COB_GROUP "Respondent country of birth grouping - deriv	Refused   2	Don't know   3	Not stated / Unk	Australian born   Mainly NESB	ba Mainly ESB	background	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published					
18	P_CITIZEN "Are you an Australian citizen?" Code "1"	Refused   2	Don't know   3	Yes   4	No	"AuSSI-ESS 2020 Published" AuSSI-ESS_2020_Published							
19	D_EDUCATION "Highest educational qualification - derived" C	Not stated / Unknown   2	Postgraduate De	Graduate Diplo	Bachelor Degree	Advanced Diplo	Certificate III & I	Secondary Educ	Secondary Educ	Certificate I & II	Secondary Education - Years 9 and below	"A	
20													

Figure 3: The basket example file

## ADEMP-PreReg Template for Simulation Studies

	A	B	C	D	E	F	G	H	I	J	K	
1	scheme	concept										
2	Mode of survey completion											
3		1, Online										
4		2, Phone										
5	SEX. What sex are you?											
6		1, Male										
7		2, Female										
8		3, Refused										
9		4, Don't know										
10	A4. Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?											
11		1, 0 – You can't be too careful										
12		2, 1										
13		3, 2										
14		4, 3										
15		5, 4										
16		6, 5										
17		7, 6										
18		8, 7										
19		9, 8										
20		10, 9										
21		11, 10 – Most people can be trusted										
22		12, Refused										
23		13, Don't know										
24	A3. On a typical day, about how much time do you spend using the internet on a computer, tablet, smartphone or other device, whether for work or personal use? (in minutes)											
25		1, 5										
26		2, 9										
27		3, 10										
28		4, 15										
29		5, 20										
30		6, 30										
31		7, 38										
32		8, 40										
33		9, 45										
34		10, 50										
35		11, 59										
36		12, 60										
37		13, 61										
38		14, 65										
39		15, 70										
40		16, 75										
41		17, 80										
42		18, 90										
43		19, 100										
44		20, 105										
45		21, 120										
46		22, 130										
47		23, 135										
48		24, 140										
49		25, 148										
50		26, 150										
51		27, 152										
52		28, 160										
53		29, 165										
54		30, 180										
55		31, 190										

Figure 4: The output file

## 1.2 Processing of the basket csv to be ready for PoolParty

### 1.2.1 Windows

This step assumes the user installed Python and Make on their system (it is easier to install Make if you install Chocolatey, and then type 'choco install Make' in the terminal). Download the folder that corresponds to Windows.

step 1: go into the directory of your choice that contains the basket csv file and the python script Scraper.py and Makefile.

step 2: open a terminal from the directory and run:

```
make
```

step 3: run

```
.\myenv\Scripts\python.exe Scraper.py Name/of/your/basket/csv/file
```

### 1.2.2 Linux/Mac

This step assumes the user installed Python and Make on their system. Download the folder that corresponds to Linux/Mac.

step 1: go into the directory of your choice that contains the basket csv file and the python script Scraper.py and Makefile.

step 2: open a terminal from the directory and run:

```
make
```

step 3: run

```
python Scraper.py Name/of/your/basket/csv/file
```

The output file (output.csv) will look something like Figure 4 and it is now ready for PoolParty upload.

## 1.3 Uploading the csv to PoolParty

Following the instructions on how to upload to PoolParty on the PoolParty website ([PoolParty link](#)), we can now import the output.csv into PoolParty (see figure 5).

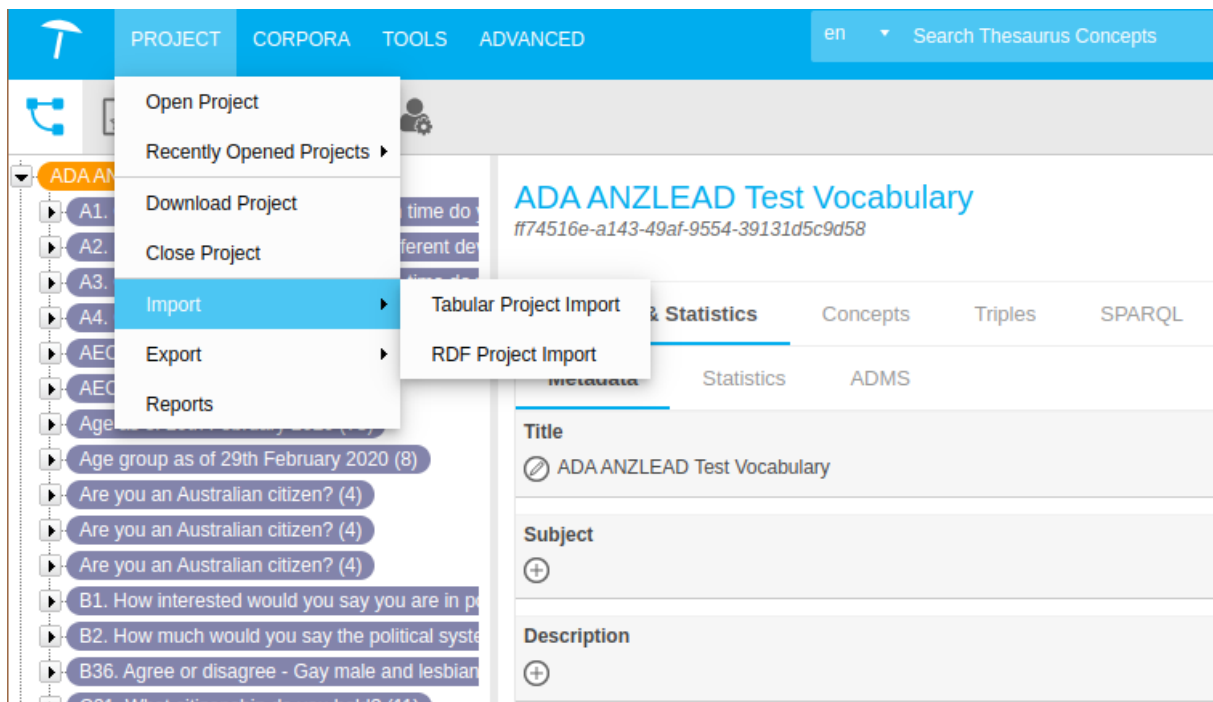


Figure 5: How to import to PoolParty

## 2 Instructions for spotting similar questions using bag of words

Please see the presentation for details on the input variables. There are four input variables required:

1. name of your input csv file
2. length of the search box, an integer is expected.
3. threshold. a float number between 0 and 1.
4. run option. "fuzzy" or "exact"

### 2.0.1 Windows

This step assumes the user installed Python and Make on their system (it is easier to install Make if you install Chocolatey, and then type 'choco install Make' in the terminal). Download the folder that corresponds to Windows.

step 1: go into the directory of your choice that contains the basket csv file and the python script Scraper.py and Makefile.

step 2: open a terminal from the directory and run:

```
make
```

step 3: run

```
.\myenv\Scripts\python.exe bow.py Name/of/your/input_csv/file  
length/of/the/search/box threshold run_option
```

## 2.0.2 Linux/Mac

This step assumes the user installed Python and Make on their system. Download the folder that corresponds to Linux/Mac.

step 1: go into the directory of your choice that contains the basket csv file and the python script Scraper.py and Makefile.

step 2: open a terminal from the directory and run:

```
make
```

step 3: run

```
python bow.py Name/of/your/input_csv/file length/of/the/search/box thresh-  
old run_option
```