Appendix B: Clean-up in OpenRefine

1. Fully manual clean-up

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
DOB		
?	(blank)	8
30-31 mei 1794	30/31 mei 1794	1
Amsterdam 1899	1899 (Amsterdam moved to POB)	1
Radom 1927	1927 (Radom moved to POB)	1
ca. 1416-1417	ca. 1416/1417	1
ca. 1460-1470	ca. 1460/1470	1
POB		
19de eeuw	(blank) (The same value was already available in DOB)	23
20ste eeuw	(blank) (The same value was already available in DOB)	9
2019	(blank) (2019 moved to DOB)	1
NA december 1970	(blank) (A more specific value was already available in DOB: 6 december 1970)	1
Assebroek/Brugge	Assebroek	1
Marigot/St. Martin	Marigot	1
O.L. Vrouw-Lombeek	Onze-Lieve-Vrouwe-Lombeek (Already available value)	1

Table 14: Manual clean-up of DOB and POB in OpenRefine (presented in-text as Table 5)

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
DOD		
?	(blank)	14
17 of 18 augustus 1936	17/18 augustus 1936	1
3 of 12 februari 2008	3/12 februari 2008	1
NA;NA	(blank)	1
Hilversum 1965	1965 (Hilversum moved to POD)	1
Oxford 2009	2009 (Oxford moved to POD)	1
2 augusus 1983	2 augustus 1983	1
3 septemer 1939	3 september 1939	1
27 seprember 1915	27 september 1915	1
26 decembe 1969	26 december 1969	1
17 februari ca. 1872	ca. 17 februari 1872	1
29 juni ca. 1737	ca. 29 juni 1737	1
15 augustus ca. 1580	ca. 15 augustus 1580	1
POD		
Amsterdam (?)	Amsterdam? (Already available value)	1
Boekelo Gem. Enschede	Boekelo (Already available value)	1
NA;NA	(blank)	2

Table 15: Manual clean-up of DOD and POB in OpenRefine

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
birthDate		
000?	(blank)	19
1888-iu-20	1888-07-20	1
1927-om-Ra	1927	1
1899-te-Am	1899	1
birthPlace		
NO FULLY MANUAL CHANGES NEEDED		

Table 16: Manual clean-up of birthDate and birthPlace in OpenRefine

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
deathDate		
000?	(blank)	65
1961-no-14	1961-11-14	1
2009-or-OX	2009	1
1915-se-27	1915-09-27	1
1939-ep-3	1939-09-03	1
1965-ve-Hi	1965	1
1969-de-26	1969-12-26	1
1983-ug-2	1983-08-02	1
deathPlace		
NO FULLY MANUAL CHANGES NEEDED		

Table 17: Manual clean-up of deathDate and deathPlace in OpenRefine

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
geboortedatum		
?	(blank)	16
30-31 mei 1794	30/31 mei 1794	1
1416-1417	1416/1417	1
1430-1440	1430/1440	1
1460-1470	1460/1470	1
Amsterdam 1899	1899 (Amsterdam moved to geb_plaats)	1
Radom 1927	1927 (Radom moved to geb_plaats)	1
20 iuli 1888	20 juli 1888	1
geb_plaats		
Assebroek/Brugge	Assebroek	1
Marigot/St. Martin	Marigot	1
O.L. Vrouw-Lombeek	Onze-Lieve-Vrouwe-Lombeek (Already available value)	1

Table 18: Manual clean-up of geboortedatum and geb_plaats in OpenRefine

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
sterfdatum		
?	(blank)	62
18 \s\s oktober 2023	18 oktober 2023	1
17 of 18 augustus 1936	17/18 augustus 1936	1
3 of 12 februari 2008	3/12 februari 2008	1
Hilversum 1965	1965 (Hilversum moved to overl_plaats)	1
Oxford 2009	2009 (Oxford moved to overl_plaats)	1
14 nomvember 1961	14 november 1961	1
2 augusus 1983	2 augustus 1983	1
3 septemer 1939	3 september 1939	1
27 seprember 1915	27 september 1915	1
26 decembe 1969	26 december 1969	1
24 maart na 1684	na 24 maart 1684	1
17 februari ca. 1872	ca. 17 februari 1872	1
29 juni ca. 1737	ca. 29 juni 1737	1
15 augustus ca. 1580	ca. 15 augustus 1580	1
ca. 1493-1494	ca. 1493/1494	1
overl_plaats		
NO FULLY MANUAL CHANGES NEEDED		

Table 19: Manual clean-up of sterfdatum and overl_plaats in OpenRefine

ORIGINAL VALUE	REFINED VALUE	NUMBER OF AFFECTED CELLS
birthCountry		
angol01	Angola	1
molda01	Moldavië	1
singa01	Singapore	1
slowa01	Slowakije	1
deathCountry		
NO FULLY MANUAL CHANGES NEEDED		

Table 20: Manual clean-up of birthCountry and deathCountry in OpenRefine

2. Clean-up using regular expressions

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
DOB		
value.replace(/\(\w*\)\s/, '')	Crop out all bracketed text e.g., “12 september (gedoopt) 1599” → “12 september 1599”	6
value.replace(/(\d+)\sv\.Chr\./, '-$1')	Change postposed “v.Chr.” into preposed “-” e.g., “ca. 800 v.Chr.” → “ca. -800”	4
if(and(value.contains("?"), value.indexOf("?") == value.length() - 1), "ca. " + value.substring(0, value.length() - 1), value)	Change a postposed “?” into preposed “ca.” e.g., “1595?” → “ca. 1595”	3
POB		
value.replace(/\s\(.*\)/, '')	Crop out all bracketed text e.g., “Salatiga (Java)” → “Salatiga”	38
value.replace(/,.*/, '')	Leave out everything after a comma e.g., “Ans, bij Luik” → “Ans”	19
value.replace(/a\/d/, 'aan den')	Change “a/d” into “aan den” e.g., “Alpen a/d Rijn” → “Alphen aan den Rijn” (an already available value)	7

Table 21: Clean-up using regular expressions of DOB and POB in OpenRefine (presented in-text as Table 6)

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
DOD		
value.replace(/\(\w*\)\s/, '')	Crop out all bracketed text e.g., “13 oktober (begraven) 1679” → “13 oktober 1679”	5
value.replace(/(\d+)\sv\.Chr\./, '-$1')	Change postposed “v.Chr.” into preposed “-” e.g., “347 v.Chr.” → “-347”	2
if(and(value.contains("?"), value.indexOf("?") == value.length() - 1), "ca. " + value.substring(0, value.length() - 1), value)	Change a postposed “?” into preposed “ca.” e.g., “1484?” → “ca. 1484”	4
POD		
value.replace(/\s\(.*\)/, '')	Crop out all bracketed text e.g., “Minneapolis (Minnesota)” → “Minneapolis”	13
value.replace(/,.*/, '')	Leave out everything after a comma e.g., “Willemstad, Curaçao” → “Willemstad”	19
value.replace(/a\/d/, 'aan den')	Change “a/d” into “aan den” e.g., “Alpen a/d Rijn” → “Alphen aan den Rijn” (an already available value)	4

Table 22: Clean-up using regular expressions of DOD and POD in OpenRefine

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
birthDate		
NO CHANGES INCLUDING REGULAR EXPRESSIONS NEEDED		
birthPlace		
value.replace(/\s\(.*\)/, '')	Crop out all bracketed text e.g., “Driewegel (bij Terneuzen)” → “Driewegel”	5
value.replace(/a\/d/, 'aan den')	Change “a/d” into “aan den” e.g., “Alpen a/d Rijn” → “Alphen aan den Rijn” (an already available value)	5

Table 23: Clean-up using regular expressions of birthDate and birthPlace in OpenRefine

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
deathDate		
NO CHANGES INCLUDING REGULAR EXPRESSIONS NEEDED		
deathPlace		
value.replace(/a\/d/, 'aan den')	Change “a/d” into “aan den” e.g., “Alpen a/d Rijn” → “Alphen aan den Rijn” (an already available value)	2

Table 24: Clean-up using regular expressions of deathDate and deathPlace in OpenRefine

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
geboortedatum		
value.replace(/\?\((\w*\s?eeuw)\)/, "$1")	Leave out brackets and “?” around centuries e.g., “?(13de eeuw)” → “13de eeuw”	1,936
value.replace(/\(\w*\)\s/, '')	Crop out all bracketed text e.g., “15 april (gedoopt) 1783” → “15 april 1783”	6
if(and(value.contains("?"), value.indexOf("?") == value.length() - 1), "ca. " + value.substring(0, value.length() - 1), value)	Change a postposed “?” into preposed “ca.” e.g., “1595?” → “ca. 1595”	3
geb_plaats		
value.replace(/\s\(.*\)/, '')	Crop out all bracketed text e.g., “Cirebon (Java)” → “Cirebon”	40
value.replace(/\.\s.*/, '')	Leave out everything after a dot e.g., “Detroit. Michigan” → “Detroit”	27
value.replace(/a\/d/, 'aan den')	Change “a/d” into “aan den” e.g., “Alpen a/d Rijn” → “Alphen aan den Rijn” (an already available value)	9

Table 25: Clean-up using regular expressions of geboortedatum and geb_plaats in OpenRefine

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
sterfdatum		
value.replace(/\?\((\w*\s?eeuw)\)/, "$1")	Leave out brackets and “?” around centuries e.g., “?(6de eeuw)” → “6de eeuw”	68
value.replace(/\(\w*\)\s/, '')	Crop out all bracketed text e.g., “10 april (begraven) 1575” → “10 april 1575”	6
if(and(value.contains("?"), value.indexOf("?") == value.length() - 1), "ca. " + value.substring(0, value.length() - 1), value)	Change a postposed “?” into preposed “ca.” e.g., “1867?” → “ca. 1867”	4
overl_plaats		
value.replace(/\s\(.*\)/, '')	Crop out all bracketed text e.g., “Reston (Virginia)” → “Reston”	14
value.replace(/\.\s.*/, '')	Leave out everything after a dot e.g., “Dandy. Vermont” → “Dandy”	25
value.replace(/a\/d/, 'aan den')	Change “a/d” into “aan den” e.g., “Alpen a/d Rijn” → “Alphen aan den Rijn” (an already available value)	3

Table 26: Clean-up using regular expressions of sterfdatum and overl_plaats in OpenRefine

REGULAR EXPRESSION	EFFECT AND EXAMPLE	NUMBER OF AFFECTED CELLS
geboortedatum		
NO CHANGES INCLUDING REGULAR EXPRESSIONS NEEDED		
geb_plaats		
NO CHANGES INCLUDING REGULAR EXPRESSIONS NEEDED		

Table 27: Clean-up using regular expressions of birthCountry and deathCountry in OpenRefine