

Annotation principles – Finnish migration records (church books)

1 Table structure

- 1) Start by annotating the borders of headers and tables. Page headers, table headers, and data tables are all annotated as separate table objects. When the table borders are annotated, add rows and columns for the data tables. Rows and columns are not added to the headers.

Pag. 107.		S. 107.		Längd öfver inflyttade för år 18		i	församling.
Månad. Kuuksius.	Dag. Päivä.	Allérens Attestens nummern.	De inflyttades ständ och namn. Tulluksen sääty ja nimi.	Personernas antal. Ihmis-luku.		För Förrum i kyrkoboken. Utan nummer korrespondens.	Anmärkningar. Huomallusia.
				Mänkin Miehet	Kvinnor. Naissuhteet		
Jouluks.	17	148	Rentti Verner Valdemar Nieminen	1	162	Hämeenlinna 252 7/2	
"	20	149	Piika Amanda Aluoroinen	1.	Jaakkimatti 350 7/2		
"	22	150	Amanda Bib. Kallio	1.	Uljala 389 7/2		
"	"	151	Rentti Jaha Wilhelm Kar-			443	7/2

- 2) Annotate also empty tables and pages.

- 3) If a page does not consist of moving data, it is not annotated. For example the right side page below.

1.1 Column annotation principles

- 1) Annotate all columns for each table, even if the table is empty/half empty. Rows are not annotated for empty tables, or empty parts of partially filled tables.
 - 2) If column dividers are visible, try to follow them. For columns with subheadings, annotate subheadings as separate columns.

Månad. Kuuksiusi.	Dag. Päivä	Alkutens nummer: Mordin numero	De Inflytades ständ och namn. Tulleen sääty ja nimi.	Personernas antal. Ihmis-luku.	Från hvilken för- samling inflyt- ningen skett Lähet numero kirkohetjysa	Volum i kyrkohöken.	Anmärkningar. Huomattavia.
Jouluk.	17	148	Rentti Verner Valdemar Nieminen	133	162	Hameenlinna	252 $\frac{1}{2}$
"	20	149	Piika Amanda W. Vuorinen	1	1.	Savonlinna	350 $\frac{1}{2}$
"	22	150	" Amanda Olof. Rallio	1	1.	Nyala	389 $\frac{1}{2}$
"	"	151	Rentti Juhha Wilhelm Kar- hunen, o. ja sr	2	2.	Kylinahenk	68 $\frac{1}{2}$
	23	152	Rouva Hilda Elin Lindberg	1	1.	Elimäki	57 $\frac{1}{2}$

3) If the table is skewed, try to adjust the rows and columns to match the table (press H or V and use the keyboard “arrows” to adjust).

4) If some columns go over the text, try to edit the cells one by one (selection mode in transkribus, click the cell and adjust the corners of the cell). However, this is not necessary if every cell would have to be modified separately.

5) If there are one table openings, the page line in the middle of the book doesn't need to be annotated as a separate column.

<u>kaa</u> seura-	kunnasta muutta
försam-	ling utflyttade år 18
<u>Systuna</u> Föderie	festl. Blåmbl. Blå- mänen.
Falka. Ort.	Kalling. Kaka. Gift. Oppl.
<u>Ljagrönvita</u> "fälten"	mainut "

6) If the text systematically goes over the printed columns (not respecting the book layout), the columns don't need to be added to the table.

Födelse		Länsdels postnummer		Namn		Postort		Namn		Postort	
År och Datum	Ort	Kommun namn	K postnr	A postnr	B postnr	C postnr	D postnr	E postnr	F postnr	G postnr	H postnr
X-1910-01-01	Stockholm	Stockholms Stad	100000	Anders	Andersson	Stockholms Stad	100000	Anders	Andersson	Stockholms Stad	100000

1.2 Row annotating principles

1) One row generally consist of one text line, but may include also several lines, depending on the table structure.

2) If the table includes visible printed/drawn row lines, use those to annotate rows. *Each person and its information has its own row.*

Elc	41 Laanisen Olli Pelti, Pietarsaari	1	400 Kungsjöön	mainit	talot, talot, veivis	200 Kansi	1/145
"	Mustonen Ann Fabri, Tyggrundat	1	401 "	"	"	"	"
"	Laaninen Olli	1	402 Kansi	antti	salottitila typpi	"	"
"	" Lani-ais	1	403 Kansi	"	"	"	"
"	42 Mäkinen Arno Antti t.	1	404 Kungsjöön	mainit	talot, veivis	200 Kungsjöön	1/112
sys	7 Anttilainen Karina Lehtiniemi	1	405 "	"	"	"	3.161
"	12 Toponens Ropponen metsäntekijä Utkilma	1	406 Heinävesi	"	"	300 Heinävesi	1/268

- 3) Especially when annotating handdrawn images without visible row lines, the rows should in general follow the running number if such exist. *Each person does not necessarily have its own row in all book formats.*

1.	Emilia H. 38a	Elternteile: Opa Edwin Jagt Großmutter: Maria	gämsmauer	1903-28-28-161	1	14	Emilia	6	492	Hattinen joku testo. Igt. ja seni voimassa	Schaukate	1903-31-25-88	1
2.	-	- 17a	Hattinen Elmer 308	testo		15	-	8	593	Hattinen joku testo. Igt. ja seni voimassa	Schaukate	1904-7-21-7-21	
3.	-	21a	Elisabetta Maria Editha Gustaf Kihlberg voimassa	Gävleholma	1903-11-19-18-1	16	-	11	520	Istor. pohd. Hämeenlinna			
4.	-	-	Gävleholma ja ka. Poh. Le Helsingfors	-	-	17	-	4	595	Hattinen Ester Paul Kuusimaa-parkkaisten	Kuusimaa	1904-5-5-9-1	

- 4) However, if there is a handdrawn table where running number is for whole family, but each person has its information in one row, in that case, the table is annotated by following the rule “one person in one row”.

7	1	Hattie Frick Altagoita Montero Linda smallwood Hattie		1	8	2.	2.	Wendy Annie John, Foster wife, Foster wife, Foster Mrs. Maria Louise, Foster wife	608	Residence	Japan
1	7	2	Cathleen Lewis Brett Gustavson, Stewart	Linda	81	7	4	J. William James Waini Webster	-	-	Japan
7	7	2	G. Carter Witherspoon, Foster wife		-	-	-	Albuquerque, New Mexico	-	-	Japan
			S. Lydia Maria		-	-	-	Albuquerque, New Mexico	-	-	Japan
			Frans Brugel		-	-	-	Albuquerque, New Mexico	-	-	Japan
9	7	1	Hattie Foster wife, Anna, Maria, Ruthie	Foster wife	62	7	1	Ruthie Foster wife, Anna, Maria, Ruthie	90	Residence	Japan
3	7	1	Ruthie Foster wife, Anna, Maria, Ruthie	Foster wife	287	7	1	Marguerite Frans Alfred Horowitz, Brown	40	Residence	Japan
			Lyat 8/1923, from 8/23					de Lyat 8/1923, Marguerite Frans Alfred Horowitz, Brown	40	Residence	Japan

- 5) Subheadings in the middle of the table (typically year information) are annotated as separate rows.

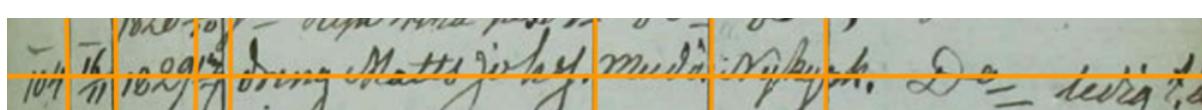
- 6) Annotate also empty rows that are *in the middle* of the table as separate rows, try to be systematic.

Lukitakun	1. 16. Jukka Antti, Tuomonen, Lempäälä	1.	1902	Tiesi	animato	Kuorijyys	kuorijyys	100%	
"	5. 17. Tuula Terhi, Tuomonen, Lempäälä	1.	1903	Tiesi	"	Pielislaat	Tiesi	100%	
"	10. 18. Alvaro Berndt Lantue, apulainen	1.	1905	Tiesi-pitkä	animato	Pappi	Lapinvirta	100%	
"	22. 19. Anna Ranta, Kotka-Pietarsaari, sit. palka	1.	1906	Tiesi	animato	Pielislaat	Korttilahti	100%	
			1						
Tuokkulan	2. 20. Liisa Ylkkönen, Tärkkeli, sit. palka	1.	1908	Tiesi	animato	Pielislaat	Tiesi	100%	
"	15. 21. Liisa Karri Ylkkönen, Tärkkeli, sit. palka	1.	1909	Tiesi	"	"	Ylkkönen	100%	
"	18. 22. Kalle Frans Eskola, valokuvaaja	1.	1910	Pontanen	animato	Kästölä	Pontanen	100%	
"	" 23. Frans Eskola, Kokkola	1.	1910	"	"	"	"	100%	
"	" 24. Alvaro Antti	1.	1910		animato	"	"	100%	
"	" 25. Hilda Alvaro	1.	1910		"	"	"	100%	
"	" 26. Kalle Liina	1.	1910		"	"	"	100%	
"	" 27. Tuula Frans	1.	1910		"	"	"	100%	
"	" 28. Hilda Alvaro	1.	1910		"	"	"	100%	
"	27. 29. Klemuni Maria Mantsi, last	1.	1912		animato	Piikkilä	Laki, metsätaloust	100%	
			1						
Korikun	30. 24. Alvaro Kaisa Ylkkönen, Leivonkoski	1.	1910	Tiesi	Ierki	Kuorijyys	Pielislaat	100%	
"	" 25. Herman	1.	1910	"	animato	"	"	100%	
"	" 26. Tuula Frans	1.	1910	"	"	"	"	100%	
"	" 27. Alvaro	1.	1910	"	"	"	"	100%	
"	38. 26. Alvaro Ylkkönen, Raumala, Loviisa	1.	1910	"	animato	"	Ylkkönen	100%	
"	" 28. Alvaro Ylkkönen	1.	1910	"	"	"	"	100%	
"	" 29. Leppi-Leija Marja	1.	1910	Mallila	Leppi	"	"	100%	
			1						
Kurikan	7. 26. Alvaro Antti Ylkkönen, taloja	1.	1912	Tiesi	animato	"	Nurmes	100%	
"	8. 27. Alvaro Antti Ylkkönen, taloja	1.	1912	"	"	"	Ylkkönen	100%	

7) Empty rows *in the end/in the beginning* of the table, don't need to be annotated. Remember to always annotate columns, even if the table is empty.

57.	"	<i>Gata i Maria</i>	"	" 94	1
58.	Frank L	344 lot 2 Dead. A. T. & C. Birth	"	42 481	
59.	"	41 76 G. L. Smith Birth	"	42 481	
60.	"	1 Tot. 1000. G. L. Smith Birth	"	42 481	
61.	"	1 Tot. 1000. G. L. Smith Birth	"	42 481	
62.	"	1 Tot. 1000. G. L. Smith Birth	"	42 481	
63.	"	1 Tot. 1000. G. L. Smith Birth	"	42 481	
64.	"	1 Tot. 1000. G. L. Smith Birth	"	42 481	
65.	"	1 Tot. 1000. G. L. Smith Birth	"	42 481	

8) If the letters go over the row line, try to annotate the row a little bit higher/lower. Try not to annotate the row line over text so that text would be divided into two rows. If it is not possible to separate the rows perfectly, try to adjust it so that the text is mainly in one row. E.g. in the below example, try to annotate the row divider above the text.



2 Cell annotation

2.1 Annotate cell as **Empty** if

- 1) The cell is empty so that the text recognition model is not expected to produce any output for it.
- 2) The cell is empty also if it only includes letters overflowing from other cells.
- 3) The cell includes something which is not text and not meant to be read from that cell alone (includes e.g. part of line crossing the whole page) → text recognition is not expected to produce anything.
- 4) Note: Hyphens (–) are annotated as text cells.
- 5) Note: Crossed out text is annotated as text cell if the text is still clearly visible and readable. If not visible (only black) then cell is annotated as empty.

2.2 Annotate cell as **Same-as** if

- 1) Has only double quote (“) mark
- 2) Has only dot (.) and from the following context it can be understood as same-as marking
- 3) Has only D or Do marking (Note: if the cell has Dit or Dito as text, it is annotated as text cell)
- 4) Has only S or Sn marking which can be understood as same-as marking from the context
- 5) Several repeated same-as marking (e.g. “ or D or Do) and nothing else
- 6) Note: If a cell includes text in addition to same-as markings, it is annotated as text cell

2.3 Annotate cell as **Multi-line** if

- 1) The cell includes two or more text line on top of each other, and these are meant to be read separately
- 2) If extra text line is part of two cells (not able to separate it to only one), try to add it to the one where it semantically belongs and mark that to be multi-line cell. If text is between rows, do not mark both cells to multi-line only because one line is on top of two rows, try to add it to the row where it belongs to. However, if text goes over columns, so that a substantial amount of text is part of both columns, both cells can be annotated as multi-line.
- 3) Dates e.g. 12/1 1910, where date and month are written on top of each other, are not multiline, as they are not meant to be read separately. Dates are annotated as text cells. However, date can also be multi-line if it is clearly written as two separate lines.

3 Text recognition

- 1) Annotate the text as it appears in the image, do not normalize abbreviations or intended orthographic variation (e.g. Hfors is not normalized into Helsingfors). However, annotate also letters which are not easily readable from the image, if those can be derived from the context, and they follow the normal orthographic intention (e.g. annotate Helsingfors even if the final s is not clearly visible).
- 2) If the text is partly unreadable, add ? to mark letters which are unclear, and cannot be derived from the context.
- 3) If text is completely unreadable, add ? to mark to unclear text (as opposed to empty cell).