



link- **LIVES**

Paradata.

Computer-Assisted Domain Expert Record Linkage

December 2025



Rigsarkivet



UNIVERSITY OF
COPENHAGEN



KØBENHAVNS
STADSARKIV

CARLSBERGFONDET

 **nnovationsfonden**

Link-Lives Paradata.

Computer-Assisted Domain Expert Record Linkage

These documents provide supplementary detail to the domain-expert, computer-assisted approach to record linkage outlined in Section 6.1 of the Guide and described in depth in Section 10. They are presented in their original form as “working documents” produced during the course of the project so they can be used to further understand the instructions and further micro-decisions of the data produced.

These materials reflect the evolving conversations, practices, and ideas of the project at different stages and with varying levels of involvement from different team members. Certain individuals were responsible for drafting, coordinating, or synthesizing specific components, and their names appear in the documents to reflect those roles; we preserve this attribution in order to keep the materials as close as possible to their original form. This should not be interpreted as an exclusive claim of authorship, but rather as an indication of who played a prominent role in the creation of each document. As working documents, they are not intended to establish authorship in the academic sense.

Table of contents

1. Best Practices 1845-1860 Only rural
2. Best Practices 1845-1860 All areas
3. Best Practices Rural parishes 1880+1901
4. Best Practices Copenhagen Burials 1861-1911
5. Best Practices PR burials focus
6. Best Practices Adjustments
7. Best Practices PR marriages and baptisms
8. Best Practices Adjustments
9. A guide to conflict solving
10. Example of Linking School Welcome Pack
11. Example of Linking Workshop

Linking best practice and recommendations (1845-1860 Only rural areas)

Link-Lives

Version number: 1

Updated: March 2020

Linking – best practice and recommendations

Linking type: Training Data

Censuses: 1845, 1850, 1860

Area type: Rural parishes

Definitions: Source1 is the census we're linking from (the newest one), Source2 is the census we're linking to (the older one).

OUR COMMON STARTING POINT

- We link a whole parish at a time
- We take decisions (ranging from "Link" to "Not found") about every person in the parish
- Generally, linking from the first to the last person is the best way, in order to make sure that no one is forgotten (but that is optional as long as the bullet above is fulfilled)
- We want to get as many people as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links. The focus is on linking the right ones.

THE COMMANDMENTS

1. Use only the information from Source 1 and Source 2
2. Always challenge the suggested links
3. Always find a competing candidate
4. When in doubt, don't link

1. SEARCHING FOR THE HOUSEHOLD

This step describes the opportunity for fast linking of entire households.

The composition of the household is the strongest linking parameter we can get. If more household members have identical information between censuses, we are rarely seeing two different households.

Look at the characteristics of the household of the person you are about to link: Is he/she likely to have been living in the same household in Source2?¹

¹ Likelihood of living in the same household: If in Source1 we want to link a 29-year-old unmarried tjenestekarl, he cannot be expected to have been tjenestekarl in the same household in Source2. If we want to link a 45-year-old

1. If Yes:

- a. Browse through the potential links of the core household members (clicking only the promising links) until one of them leads you to the right household. LINK all the relevant household members.
 - NOTE: be aware that core household members can change: the wife might be new, or different children can be living at home with the parents in different censuses
- b. If no potential links lead you to the right household, go to Step 2 and search for one of the individuals in the household using the search boxes as in step 2. Start searching for the person you think you are most likely to find quickly. If, for instance, the wife or a daughter has a particular name, you should look for her before you look for the head of household, as it will normally be the easiest way to get to the entire household.

2. SEARCHING FOR A PERSON

2.1. ALWAYS FIRST REVISE THE POTENTIAL LINKS

Even though the potential links and the families do not help you find a complete household, they may still help you find some individuals, so just browse the 10 candidates and see if there are evidences that one of them could be the corresponding person. If that fails, try searching using the search boxes:

2.2. SEARCHING WITH NAME, BIRTH YEAR AND BIRTH PLACE

Always start with typing the name, birth year and birth place in the text boxes. These are the primary variables to use for identification purposes. This is due to the fact that these variables are in principle invariable during life (or, in the case of the name, if it changes, the change can be expected to follow some relatively well-defined rules).

In general, these three variables must all contain satisfying information for a secure link to be established. This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.

The above means that unless a link is confirmed by the other household members, we should always use both of these two variable combinations in order to not miss a potential match:

- Name + birth year + birth place: for when the parish name is mentioned in the original birth place string
- Name + birth year + residence parish (using the value of birth place): for when the original birth place string says “ditto”, “samme”, “her i sognet” etc.

However, to avoid linking someone without enough evidence (when there are multiple possibilities, it is recommended to always try to find people using at least the two methods to ensure that there are not too

married husband who has a wife and a daughter of 15 years, he must be expected to live with at least the daughter in Source2, and probably also the wife (unless he was married more than once).

candidates very similar: *Jens Christensen, 1821, ditto (living in Junget)* and *Jens Christensen, 1821, Junget*. If that is the case and there is no distinguishing information, do not link.

If these combinations do not give a good match, the birth place might be recorded in an unpredictable way. In this case it makes sense to delete the birth place from the search string and block on “Residence county” instead:

- Name + birth year + residence county

This will normally give more matches – the nearest ones – and then you can browse the list in search for interesting candidates. If the name is special, the residence county should be left out of this search.

After these types of basic ways of getting searches, there may be relevant to remove one of the criteria, expand the years or remove them entirely, etc. to exhaust other alternatives. Read below for a more detailed explanations of how to better understand how we link.

3. HOW TO BEST USE THE SEARCH FIELDS

3.1. SEARCHING NAMES

Generally, you should expect the name to be precise for a good link (a part from spelling variations). If the name does not fit, we are hesitant to link unless we have household information to confirm it

Anne is not generally matched to Anne Marie, or Søren Christian should not be matched to Søren. But maybe Cathrine Marie Wilhelmina could be matched to someone with fewer names

There are four search fields for names but use only *the unstandardized search field “name”* which contains the original name string. The three *the standardized name fields*² are an experimental feature that we are testing. You should not use these fields for name searching purposes!

Type (ALWAYS use lower case letters).

It is important to use the wildcard * (which can stand for anything) when searching in order to account for spelling variants. At the same time, it is important to use it in the right places: write the constituting letters of a name in the search string, and * out the parts that can be spelled in different ways or are not necessary to distinguish from other names.

Example:

- When looking for a Bodil Cathrine Christensdatter, a good search string would be “bo* *at* *st*ns”
 - “bo” is likely to be present in most spelling variants of Bodil
 - “at” is likewise likely to be present in most spelling variants of Cathrine

² These three variables have the following content: 1) name_std: the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name) : kristen vilhelm Jakobsen or poul abraham len inger; 2) first_name_std: the first word of the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name) Kristen, poul. 3) last_name_std: the last word of the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name)

- Christensdatter is more tricky, both because Christen can be spelled in many ways and because we need to take into account that a woman can have a *-datter*-patronym in one census and a *-sen*-patronym in another (that is, we want the search string to catch Christensen as well as Christensdatter). The “st” and the “ns” will almost always be in the spelling variants
- NOTE that you regulate the minimum number of words in the name putting in the blanks (spaces) – in the example above there will be at least three words, as there are two blanks in the search string (an imaginary “borat Christensen” will not be a hit, due to the space between “bo*” and “*at”).
- NOTE that the wildcard * can mean anything – multiple letters, numbers, signs, blanks and also nothing. The * is default in the beginning and end of your search string
- When looking for Anne Maria Krøll, a good search string could be “an*ma* kr*I”
 - “an*ma*” – without a space in the middle – takes into account that sometimes “Anne Marie” can be spelled in one word (“Annemarie”)
 - “kr*I” are the constituting letters of “Krøll” and might catch most other spelling variants
 - NOTE that especially rare (sur)names are likely to be spelled in multiple ways

3.2. SEARCHING BIRTH YEAR

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

There are two parallel – search fields for this variable:

- year of birth: The calculated year of birth (census year minus age)
- age: The age given in the census

It is easier to compare birth year between two censuses, so we recommend that you use the birth year as a search parameter, although the birth year is calculated from the given age.

In the search field, you can use the wildcard “+x” to search for birth years that are within the range of x years before or after the specified one. Example: if you write “1829+3” the program will return birth years from 1826 to 1832 (inclusive).

General recommendations:

- Age 0-49: Use calculated birth year “+3”
- Age 50-64: Use calculated birth year “+5”
- Age 65-: Use calculated birth year “+10”

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range.

3.3 SEARCHING BIRTH PLACE

Generally, you should expect the birth place to be precise for a good link. Some systematic errors have been identified though:

- the more far away, the less precision in the birth place: a person not born in the parish will have a less specific place of birth (maybe the amt, even, or the country, Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong increases with the distance.
- the oldest child is relatively often born in another parish (before the parents settled down) – but is registered with the birth place as the place where he/she grew up
- parishes within the same “pastorat” are more often confused with each other than others

There are three search fields for birth places:

- birth_place: the original text string from the birth place column
- parish_std: the name of the birth parish, if it has been identified and standardized in the original birth place string
- county_std: the name of the birth county, if it has been identified and standardized in the original birth place string

At the moment – mid March 2020 – there are errors in the standardized birth places, and there is still a substantial number of birth places that have not been standardized. This means that the standardized birth place fields should be used only with their incompleteness in mind, thus, always checking also the other fields. We are currently working on improving this part of the data.

Especial warning for Viborg: our program understands now “v. amt” as belonging to “vejle amt” so our algorithm is not accurately capturing in the best matches all the potential cases.

The *original birth place field* is not without problems either, though, but here are some solutions:

- Birth places can be recorded on different levels from census to census (farm, village, parish, town, county)
 - The linker should bear it in mind as a possibility that a birth place is not always described as a parish name.
 - A sparse use of Google Maps is recommended. Not for every search but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or they could be part of the same parish.
- Birth places are often recorded as variants of “her i sognet”, “samme sted”, “ditto” and so on – that is text strings which do not include the actual name of the place
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.

4. PRIORITY LIST OF CRITERIA FOR LINKING:

Most of the links are relatively obvious and unequivocal, but above we have provided a lot of information on known variations so that we can allow for relatively conflicting information. To have an overview of how you should assess that variation, we propose you a short list to prioritize differences.

1. Household information
 - a. if the household fits and there is a minor discrepancy in the personal variables, the link is good
 - b. remember, though, that discrepancies in the personal variables CAN indicate that a new person has entered the household (new wife/another child living at home...)

- c. household information is given high priority only for the family members of the household (a servant being in the same household is not as strong identifier)
- 2. Name
 - a. if the name doesn't fit, we're generally hesitant to link
 - b. be aware of the possibility that women change surnames (from Jensdatter to Jensen, or from maiden name to married name)
- 3. Birth place – birth place should be identical, but a few systematic errors in the source material can be identified:
 - a. the more far away, the less precision
 - b. oldest child in a group of sibling are relatively often born in another parish – but doesn't know
 - c. parishes within the same "pastorat" are more often confused with each other than others (Google maps can help to disambiguate some cases).
- 4. Age
 - a. a difference of a single year must not be given any weight when we link
 - b. a difference of two years or more can have a certain influence
 - c. expect that the older the person is, the less precise is the age
- 5. Civil status
 - a. This variable can be used to exclude possible matches – bearing in mind that there can of course be errors in the data (especially there is probably a dark figure with regards to divorcees, registered as widow(er)s)
- 6. Occupation
 - a. This variable should not be used for identification purposes, unless we're talking about very specific and quite rare professions

Linking best practice and recommendations (1845-1860 All areas)

Link-Lives

Version number: 1

Updated: June 2020

Linking – best practice and recommendations (1845-1860)

Linking type: Training Data

Censuses: 1845, 1850, 1860

Area type: Rural parishes and urban areas (cities or hoods)

Definitions: Source1 is the census we're linking from (the newest one), Source2 is the census we're linking to (the older one).

CONTENT

CONTENT	2
OUR COMMON STARTING POINT	3
THE COMMANDMENTS	3
1. SEARCHING FOR THE HOUSEHOLD	3
2. SEARCHING FOR A PERSON	4
2.1. ALWAYS FIRST REVISE THE POTENTIAL LINKS	4
2.2. SEARCHING WITH NAME, BIRTH YEAR AND BIRTH PLACE	4
3. HOW TO BEST USE THE SEARCH FIELDS	5
3.1. SEARCHING NAMES	5
3.2. SEARCHING BIRTH YEAR	7
3.3 SEARCHING BIRTH PLACE	7
4. PRIORITY LIST OF CRITERIA FOR LINKING:	8
5. EXAMPLE: A HOUSEHOLD IN ODENSE 1850 AND 1845	9
6. ADDITIONAL GENERAL ADVICE FROM EXPERIENCED LINKERS	10

OUR COMMON STARTING POINT

This best practice document is directed towards linking between the censuses 1845, 1850 and 1860 – i.e. the mid 19th century censuses where birth place is registered. The best practice applies to all three censuses, but some parts of it is divided according to whether we're talking about linking in rural parishes or linking in urban areas. This is due to the fact that differences between rural and urban societies are reflected in our sources and influence how we should treat data.

Methodologically we have decided to work like this:

	Rural parishes	Urban areas
What	We link a whole parish at a time	We link samples consisting of whole pages of cities/hoods*
Who	We take decisions (ranging from “link” to “not found”) about every single person in the parish	We take decisions (ranging from “link” to “not found”) about every single person in the sample, including the rest of the households of the first and the last person on every sample page
How	It is optional whether you want to link from the first to the last person in the parish/sample or go through the “easy” links first and then the more difficult afterwards – as long as you remember to take decisions on everybody	
Quality	The focus is on linking the right ones: We want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links	

* Very small cities might be linked entirely

THE COMMANDMENTS

1. Use only the information from Source 1 and Source 2
2. Always challenge the suggested links
3. Always find a competing candidate
4. When in doubt, don't link

1. SEARCHING FOR THE HOUSEHOLD

This step describes the opportunity for fast linking of entire households.

The composition of the household is the strongest linking parameter we can get. If more household members have identical information between censuses, we are rarely seeing two different households.

Look at the characteristics of the household of the person you are about to link: Is he/she likely to have been living in the same household in Source2?¹

1. If Yes:

¹ Likelihood of living in the same household: If in Source1 we want to link a 29-year-old unmarried tjenestekarl, he cannot be expected to have been tjenestekarl in the same household in Source2. If we want to link a 45-year-old married husmand who has a wife and a daughter of 15 years, he must be expected to live with at least the daughter in Source2, and probably also the wife (unless he was married more than once).

- a. Browse through the potential links of the core household members (clicking only the promising links) until one of them leads you to the right household. LINK all the relevant household members.
 - NOTE: be aware that core household members can change: the wife might be new, or different children can be living at home with the parents in different censuses
- b. If no potential links lead you to the right household, go to Step 2 and search for one of the individuals in the household using the search boxes as in step 2. Start searching for the person you think you are most likely to find quickly. If, for instance, the wife or a daughter has a particular name, you should look for her before you look for the head of household, as it will normally be the easiest way to get to the entire household.

2. SEARCHING FOR A PERSON

2.1. ALWAYS FIRST REVISE THE POTENTIAL LINKS

Even though the potential links and the families do not help you find a complete household, they may still help you find some individuals, so just browse the 10 candidates and see if there are evidences that one of them could be the corresponding person. If that fails, try searching using the search boxes:

2.2. SEARCHING WITH NAME, BIRTH YEAR AND BIRTH PLACE

Always start with typing the name, birth year and birth place in the text boxes. These are the primary variables to use for identification purposes. This is due to the fact that these variables are in principle invariable during life (or, in the case of the name, if it changes, the change can be expected to follow some relatively well-defined rules).

In general, these three variables must all contain satisfying information for a secure link to be established. This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.

The above means that unless a link is confirmed by the other household members, we should always use both of these two variable combinations in order to not miss a potential match:

- Name + birth year + birth place: for when the parish name is mentioned in the original birth place string
- Name + birth year + residence parish (using the value of birth place): for when the original birth place string says “ditto”, “samme”, “her i sognet” etc.

However, to avoid linking someone without enough evidence (when there are multiple possibilities, it is recommended to always try to find people using at least the two methods to ensure that there are not too candidates very similar: *Jens Christensen, 1821, ditto (living in Junget)* and *Jens Christensen, 1821, Junget*. If that is the case and there is no distinguishing information, do not link.

If these combinations do not give a good match, the birth place might be recorded in an unpredictable way. In this case it makes sense to delete the birth place from the search string and block on “Residence county” instead:

- Name + birth year + residence county

This will normally give more matches – the nearest ones – and then you can browse the list in search for interesting candidates. If the name is special, the residence county should be left out of this search.

After these types of basic ways of getting searches, it may be relevant to remove one of the criteria, expand the years or remove them entirely, etc. to exhaust other alternatives. Read below for a more detailed explanation of how, to better understand how we link.

3. HOW TO BEST USE THE SEARCH FIELDS

3.1. SEARCHING NAMES

Generally, you should expect the name to be precise for a good link (a part from spelling variations). If the name does not fit, we are hesitant to link unless we have household information to confirm it.

Special characteristics for names in URBAN areas:

Names change more often between censuses in urban areas than in rural areas in this period, and that should be taken into consideration, when searching for names:

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the same person without having matching first names – and without any of the names being wrong!
 - Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
 - First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names
 - married women taking the husband’s last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen’s children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription

There are four search fields for names but use only *the unstandardized search field "name"* which contains the original name string. The three *standardized name fields*² are an experimental feature that we are testing. You should not use these fields for name searching purposes!

Type (ALWAYS use lower case letters).

It is important to use the wildcard * (which can stand for anything) when searching in order to account for spelling variants. At the same time, it is important to use it in the right places: write the constituting letters of a name in the search string, and * out the parts that can be spelled in different ways or are not necessary to distinguish from other names.

NOTE that the wildcard * can mean anything – multiple letters, numbers, signs, blanks and also nothing.

The * is default in the beginning and end of your search string

Example:

- When looking for a Bodil Cathrine Christensdatter, a good search string would be "bo* *at* *st*ns"
 - "bo" is likely to be present in most spelling variants of Bodil
 - "at" is likewise likely to be present in most spelling variants of Cathrine
 - Christensdatter is more tricky, both because Christen can be spelled in many ways and because we need to take into account that a woman can have a *-datter*-patronym in one census and a *-sen*-patronym in another (that is, we want the search string to catch Christensen as well as Christensdatter). The "st" and the "ns" will almost always be in the spelling variants
 - NOTE that you regulate the minimum number of words in the name putting in the blanks (spaces) – in the example above there will be at least three words, as there are two blanks in the search string (an imaginary "borat Christensen" will not be a hit, due to the space between "bo*" and "*at*").
 - NOTE that with this search you exclude finding a "Bodil Christensdatter" or a "Cathrine Christensdatter" – you should remember that these are possible aliases for Bodil Cathrine Christensdatter
- When looking for Anne Maria Krøll, a good search string could be "an*ma* kr*I"
 - "an*ma*" – without a space in the middle – takes into account that sometimes "Anne Marie" can be spelled in one word ("Annemarie")
 - "kr*I" are the constituting letters of "Krøll" and might catch most other spelling variants
 - NOTE that especially rare (sur)names are likely to be spelled in multiple ways

² These three variables have the following content: 1) name_std: the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name) : kristen vilhelm Jakobsen or poul abraham len inger; 2) first_name_std: the first word of the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name) Kristen, poul. 3) last_name_std: the last word of the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name)

3.2. SEARCHING BIRTH YEAR

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

There are two parallel – search fields for this variable:

- year of birth: The calculated year of birth (census year minus age)
- age: The age given in the census

It is easier to compare birth year between two censuses, so we recommend that you use the birth year as a search parameter, although the birth year is calculated from the given age.

In the search field, you can use the wildcard “+x” to search for birth years that are within the range of x years before or after the specified one. Example: if you write “1829+3” the program will return birth years from 1826 to 1832 (inclusive).

General recommendations:

- Age 0-49: Use calculated birth year “+3”
- Age 50-64: Use calculated birth year “+5”
- Age 65-: Use calculated birth year “+10”

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range.

3.3 SEARCHING BIRTH PLACE

Generally, you should expect the birth place to be precise for a good link. Some systematic errors have been identified though:

- the more far away, the less precision in the birth place: a person not born in the parish will have a less specific place of birth (maybe the amt, even, or the country, Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong increases with the distance.
- the oldest child is relatively often born in another parish (before the parents settled down) – but is registered with the birth place as the place where he/she grew up
- parishes within the same “pastorat” are more often confused with each other than others

Special characteristics for birth places in URBAN areas:

In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility which activates

1. the imprecision mentioned above (because of distance) and
2. a tendency to register birth *places* rather than birth *parishes*, probably because of the census enumerator’s lack of knowledge of (and interest in) exact parish names in the surrounding rural areas

There are three search fields for birth places:

- birth place: the original text string from the birth place column
- parish_std: the name of the birth parish, if it has been identified and standardized in the original birth place string
- county_std: the name of the birth county, if it has been identified and standardized in the original birth place string

At the moment – mid March 2020 – there are errors in the standardized birth places, and there is still a substantial number of birth places that have not been standardized. This means that the standardized birth place fields should be used only with their incompleteness in mind, thus, always checking also the other fields. We are currently working on improving this part of the data.

Especial warning for Viborg: our program understands now “v. amt” as belonging to “vejle amt” so our algorithm is not accurately capturing in the best matches all the potential cases. (perhaps no longer relevant)

The *original birth place field* is not without problems either, though, but here are some solutions:

- Birth places can be recorded on different levels from census to census (farm, village, parish, town, county)
 - The linker should bear it in mind as a possibility that a birth place is not always described as a parish name.
 - A sparse use of Google Maps is recommended. Not for every search but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or they could be part of the same parish.
- Birth places are often recorded as variants of “her i sognet”, “samme sted”, “ditto” and so on – that is text strings which do not include the actual name of the place
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.

4. PRIORITY LIST OF CRITERIA FOR LINKING:

Most of the links are relatively obvious and unequivocal, but above we have provided a lot of information on known variations so that we can allow for relatively conflicting information. To have an overview of how you should assess that variation, we propose you a short list to prioritize differences.

1. Household information
 - a. if the household fits and there is a minor discrepancy in the personal variables, the link is good
 - b. remember, though, that discrepancies in the personal variables CAN indicate that a new person has entered the household (new wife/another child living at home...)
 - c. household information is given high priority only for the family members of the household (a servant being in the same household is not a strong identifier)
2. Name
 - a. if the name doesn't fit, we're generally hesitant to link

- b. be aware of the possibility that women and children change last names (from Jensdatter to Jensen, from real patronym to inherited patronym or from maiden name to married name) – or don't have last names at all
- 3. Birth place – birth place should be identical, but a few systematic errors in the source material can be identified:
 - a. the more far away, the less precision
 - b. an urban enumerator's registration of a rural birth place often results in a birth place which is not the name of the relevant parish
 - c. oldest child in a group of sibling are relatively often born in another parish – but doesn't know
 - d. parishes within the same "pastorat" are more often confused with each other than others (Google maps can help to disambiguate some cases).
- 4. Age
 - a. a difference of a single year must not be given any weight when we link
 - b. a difference of two years or more can have a certain influence
 - c. expect that the older the person is, the less precise is the age
- 5. Civil status
 - a. This variable can be used to exclude possible matches – bearing in mind that there can of course be errors in the data (especially there is probably a dark figure with regards to divorcees, registered as widow(er)s)
- 6. Occupation
 - a. This variable should not be used for identification purposes, unless we're talking about very specific and quite rare professions
 - b. **IMPORTANT: If occupation is used to go from "not found" or "multiple" to "maybe" or "link", the link button "maybe occupation" MUST be used. This indicates that the occupation is decisive for the linking decision.**

5. EXAMPLE: A HOUSEHOLD IN ODENSE 1850 AND 1845

The screen dumps below show Johan Jensen's household in Odense (urban area) in 1850 and in 1845.

Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The wife and children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	smedemester, huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn

6. ADDITIONAL GENERAL ADVICE FROM EXPERIENCED LINKERS

...on moving between linking rural areas and linking urban areas:

6.1. You cannot link as many people in urban areas as in rural areas – that's irritating, but a fact of life

6.2. The urban households can consist of many persons and many families (sometimes even 50-100 persons). Check, thus, an extra time before linking: you might have chosen another person with the same name in the relevant household – name twins occur more often in big households

6.3. Be aware, when you search, that urban households are generally more complicated than rural households – the head of the household is more often unmarried, the household can be big, and there are many households with people on temporary stays

6.4. If a wife has the same last name as the husband, search for her also without using the last name

Linking best practice and recommendations (1845-1860)

Link Lives

Version number: 3

Updated: 28 Sep 2020

Linking type: Training Data

Censuses: 1845, 1850, 1860

Area type: Rural parishes and urban areas (cities or hoods)

Definitions: Source1 is the census we're linking from (the newest one), Source2 is the census we're linking to (the older one).

CONTENTS

OUR COMMON STARTING POINT	3
THE COMMANDMENTS	3
1. SEARCHING FOR THE HOUSEHOLD	3
2. SEARCHING FOR A PERSON	4
2.1. ALWAYS FIRST REVISE THE POTENTIAL LINKS	4
2.2. SEARCHING WITH NAME, BIRTH YEAR AND BIRTH PLACE	4
3. HOW TO BEST USE THE SEARCH FIELDS	5
3.1. SEARCHING NAMES	5
3.2. SEARCHING BIRTH YEAR	7
3.3 SEARCHING BIRTH PLACE	7
4. PRIORITY LIST OF CRITERIA FOR LINKING:	8
5. EXAMPLE: A HOUSEHOLD IN ODENSE 1850 AND 1845	9
6. ADDITIONAL GENERAL ADVICE FROM EXPERIENCED LINKERS	10
Appendix: Rural Parishes 1880+1901	11
1. Census distance	11
2. Naming of married women	11
3. Elderly people	11
4. Naming traditions	11
5. The enumerator	12

OUR COMMON STARTING POINT

This best practice document is directed towards linking between the censuses 1845, 1850 and 1860 – i.e. the mid 19th century censuses where birth place is registered. The best practice applies to all three censuses, but some parts of it is divided according to whether we're talking about linking in rural parishes or linking in urban areas. This is due to the fact that differences between rural and urban societies are reflected in our sources and influence how we should treat data.

Methodologically we have decided to work like this:

	Rural parishes	Urban areas
What	We link a whole parish at a time	We link samples consisting of whole pages of cities/hoods*
Who	We take decisions (ranging from "link" to "not found") about every single person in the parish	We take decisions (ranging from "link" to "not found") about every single person in the sample, including the rest of the households of the first and the last person on every sample page
How	It is optional whether you want to link from the first to the last person in the parish/sample or go through the "easy" links first and then the more difficult afterwards – as long as you remember to take decisions on everybody	
Quality	The focus is on linking the right ones: We want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links	

* Very small cities might be linked entirely

•

THE COMMANDMENTS

1. Use only the information from Source 1 and Source 2
2. Always challenge the suggested links
3. Always find a competing candidate
4. When in doubt, don't link

1. SEARCHING FOR THE HOUSEHOLD

This step describes the opportunity for fast linking of entire households.

The composition of the household is the strongest linking parameter we can get. If more household members have identical information between censuses, we are rarely seeing two different households.

Look at the characteristics of the household of the person you are about to link: Is he/she likely to have been living in the same household in Source2?¹

1. If Yes:

¹ Likelihood of living in the same household: If in Source1 we want to link a 29-year-old unmarried tjenestekarl, he cannot be expected to have been tjenestekarl in the same household in Source2. If we want to link a 45-year-old married husmand who has a wife and a daughter of 15 years, he must be expected to live with at least the daughter in Source2, and probably also the wife (unless he was married more than once).

- a. Browse through the potential links of the core household members (clicking only the promising links) until one of them leads you to the right household. LINK all the relevant household members.
 - NOTE: be aware that core household members can change: the wife might be new, or different children can be living at home with the parents in different censuses
- b. If no potential links lead you to the right household, go to Step 2 and search for one of the individuals in the household using the search boxes as in step 2. Start searching for the person you think you are most likely to find quickly. If, for instance, the wife or a daughter has a particular name, you should look for her before you look for the head of household, as it will normally be the easiest way to get to the entire household.

2. SEARCHING FOR A PERSON

2.1. ALWAYS FIRST REVISE THE POTENTIAL LINKS

Even though the potential links and the families do not help you find a complete household, they may still help you find some individuals, so just browse the 10 candidates and see if there are evidences that one of them could be the corresponding person. If that fails, try searching using the search boxes:

2.2. SEARCHING WITH NAME, BIRTH YEAR AND BIRTH PLACE

Always start with typing the name, birth year and birth place in the text boxes. These are the primary variables to use for identification purposes. This is due to the fact that these variables are in principle invariable during life (or, in the case of the name, if it changes, the change can be expected to follow some relatively well-defined rules).

In general, these three variables must all contain satisfying information for a secure link to be established. This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.

The above means that unless a link is confirmed by the other household members, we should always use both of these two variable combinations in order to not miss a potential match:

- Name + birth year + birth place: for when the parish name is mentioned in the original birth place string
- Name + birth year + residence parish (using the value of birth place): for when the original birth place string says “ditto”, “samme”, “her i sognet” etc.

However, to avoid linking someone without enough evidence (when there are multiple possibilities, it is recommended to always try to find people using at least the two methods to ensure that there are not too candidates very similar: *Jens Christensen, 1821, ditto (living in Junget)* and *Jens Christensen, 1821, Junget*. If that is the case and there is no distinguishing information, do not link.

If these combinations do not give a good match, the birth place might be recorded in an unpredictable way. In this case it makes sense to delete the birth place from the search string and block on “Residence county” instead:

- Name + birth year + residence county

This will normally give more matches – the nearest ones – and then you can browse the list in search for interesting candidates. If the name is special, the residence county should be left out of this search.

After these types of basic ways of getting searches, it may be relevant to remove one of the criteria, expand the years or remove them entirely, etc. to exhaust other alternatives. Read below for a more detailed explanation of how, to better understand how we link.

3. HOW TO BEST USE THE SEARCH FIELDS

3.1. SEARCHING NAMES

Generally, you should expect the name to be precise for a good link (a part from spelling variations). If the name does not fit, we are hesitant to link unless we have household information to confirm it.

Special characteristics for names in URBAN areas:

Names change more often between censuses in urban areas than in rural areas in this period, and that should be taken into consideration, when searching for names:

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the same person without having matching first names – and without any of the names being wrong!
 - Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
 - First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names
 - married women taking the husband’s last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen’s children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription

There are four search fields for names but use only *the unstandardized search field "name"* which contains the original name string. The three *standardized name fields*² are an experimental feature that we are testing. You should not use these fields for name searching purposes!

Type (ALWAYS use lower case letters).

It is important to use the wildcard * (which can stand for anything) when searching in order to account for spelling variants. At the same time, it is important to use it in the right places: write the constituting letters of a name in the search string, and * out the parts that can be spelled in different ways or are not necessary to distinguish from other names.

NOTE that the wildcard * can mean anything – multiple letters, numbers, signs, blanks and also nothing. The * is default in the beginning and end of your search string

Example:

- When looking for a Bodil Cathrine Christensdatter, a good search string would be "bo* *at* *st*ns"
 - "bo" is likely to be present in most spelling variants of Bodil
 - "at" is likewise likely to be present in most spelling variants of Cathrine
 - Christensdatter is more tricky, both because Christen can be spelled in many ways and because we need to take into account that a woman can have a *-datter*-patronym in one census and a *-sen*-patronym in another (that is, we want the search string to catch Christensen as well as Christensdatter). The "st" and the "ns" will almost always be in the spelling variants
 - NOTE that you regulate the minimum number of words in the name putting in the blanks (spaces) – in the example above there will be at least three words, as there are two blanks in the search string (an imaginary "borat Christensen" will not be a hit, due to the space between "bo*" and "*at*").
 - NOTE that with this search you exclude finding a "Bodil Christensdatter" or a "Cathrine Christensdatter" – you should remember that these are possible aliases for Bodil Cathrine Christensdatter
- When looking for Anne Maria Krøll, a good search string could be "an*ma* kr*I"
 - "an*ma*" – without a space in the middle – takes into account that sometimes "Anne Marie" can be spelled in one word ("Annemarie")
 - "kr*I" are the constituting letters of "Krøll" and might catch most other spelling variants
 - NOTE that especially rare (sur)names are likely to be spelled in multiple ways

² These three variables have the following content: 1) name_std: the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name) : kristen vilhelm Jakobsen or poul abraham len inger; 2) first_name_std: the first word of the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name) Kristen, poul. 3) last_name_std: the last word of the name string in its standardized form (if the exact word has not been standardized, it is transformed into the closest standard name)

3.2. SEARCHING BIRTH YEAR

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

There are two parallel – search fields for this variable:

- year of birth: The calculated year of birth (census year minus age)
- age: The age given in the census

It is easier to compare birth year between two censuses, so we recommend that you use the birth year as a search parameter, although the birth year is calculated from the given age.

In the search field, you can use the wildcard “+x” to search for birth years that are within the range of x years before or after the specified one. Example: if you write “1829+3” the program will return birth years from 1826 to 1832 (inclusive).

General recommendations:

- Age 0-49: Use calculated birth year “+3”
- Age 50-64: Use calculated birth year “+5”
- Age 65-: Use calculated birth year “+10”

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range.

3.3 SEARCHING BIRTH PLACE

Generally, you should expect the birth place to be precise for a good link. Some systematic errors have been identified though:

- the more far away, the less precision in the birth place: a person not born in the parish will have a less specific place of birth (maybe the amt, even, or the country, Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong increases with the distance.
- the oldest child is relatively often born in another parish (before the parents settled down) – but is registered with the birth place as the place where he/she grew up
- parishes within the same “pastorat” are more often confused with each other than others

Special characteristics for birth places in URBAN areas:

In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility which activates

1. the imprecision mentioned above (because of distance) and
2. a tendency to register birth *places* rather than birth *parishes*, probably because of the census enumerator’s lack of knowledge of (and interest in) exact parish names in the surrounding rural areas

There are three search fields for birth places:

- birth place: the original text string from the birth place column
- parish_std: the name of the birth parish, if it has been identified and standardized in the original birth place string
- county_std: the name of the birth county, if it has been identified and standardized in the original birth place string

As of mid March 2020, there are errors in the standardized birth places, and there is still a substantial number of birth places that have not been standardized. This means that the standardized birth place fields should be used only with their incompleteness in mind, thus, always checking also the other fields. We are currently working on improving this part of the data.

The *original birth place field* is not without problems either, though, but here are some solutions:

- Birth places can be recorded on different levels from census to census (farm, village, parish, town, county)
 - The linker should bear it in mind as a possibility that a birth place is not always described as a parish name.
 - A sparse use of Google Maps is recommended. Not for every search but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or they could be part of the same parish.
- Birth places are often recorded as variants of “her i sognet”, “samme sted”, “ditto” and so on – that is text strings which do not include the actual name of the place
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.

4. PRIORITY LIST OF CRITERIA FOR LINKING:

Most of the links are relatively obvious and unequivocal, but above we have provided a lot of information on known variations so that we can allow for relatively conflicting information. To have an overview of how you should assess that variation, we propose you a short list to prioritize differences.

1. Household information
 - a. if the household fits and there is a minor discrepancy in the personal variables, the link is good
 - b. remember, though, that discrepancies in the personal variables CAN indicate that a new person has entered the household (new wife/another child living at home...)
 - c. household information is given high priority only for the family members of the household (a servant being in the same household is not a strong identifier)
2. Name
 - a. if the name doesn't fit, we're generally hesitant to link
 - b. be aware of the possibility that women and children change last names (from Jensdatter to Jensen, from real patronym to inherited patronym or from maiden name to married name) – or don't have last names at all
3. Birth place – birth place should be identical, but a few systematic errors in the source material can be identified:

- a. the more far away, the less precision
 - b. an urban enumerator's registration of a rural birth place often results in a birth place which is not the name of the relevant parish
 - c. oldest child in a group of sibling are relatively often born in another parish – but doesn't know
 - d. parishes within the same "pastorat" are more often confused with each other than others (Google maps can help to disambiguate some cases).
4. Age
- a. a difference of a single year must not be given any weight when we link
 - b. a difference of two years or more can have a certain influence
 - c. expect that the older the person is, the less precise is the age
5. Civil status
- a. This variable can be used to exclude possible matches – bearing in mind that there can of course be errors in the data (especially there is probably a dark figure with regards to divorcees, registered as widow(er)s)
6. Occupation
- a. This variable should not be used for identification purposes, unless we're talking about very specific and quite rare professions
 - b. **IMPORTANT: If occupation is used to go from "not found" or "multiple" to "maybe" or "link", the link button "maybe occupation" MUST be used. This indicates that the occupation is decisive for the linking decision.**

5. EXAMPLE: A HOUSEHOLD IN ODENSE 1850 AND 1845

The screen dumps below show Johan Jensen's household in Odense (urban area) in 1850 and in 1845.

Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The wife and children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	amedemester. huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn

6. ADDITIONAL GENERAL ADVICE FROM EXPERIENCED LINKERS

...on moving between linking rural areas and linking urban areas:

6.1. You cannot link as many people in urban areas as in rural areas – that's irritating, but a fact of life

6.2. The urban households can consist of many persons and many families (sometimes even 50-100 persons). Check, thus, an extra time before linking: you might have chosen another person with the same name in the relevant household – name twins occur more often in big households

6.3. Be aware, when you search, that urban households are generally more complicated than rural households – the head of the household is more often unmarried, the household can be big, and there are many households with people on temporary stays

6.4. If a wife has the same last name as the husband, search for her also without using the last name

Appendix: Rural Parishes 1880+1901

This document is a supplement for the Best Practices for linking the censuses 1845-1860. Basically, we use the same techniques for linking, but with a few important points of interest that should be born in mind while linking rural parishes 1880 + 1901.

1. Census distance

There are 20 years between 1880 and 1860, and there are 21 years between 1901 and 1880. Households change a lot in 20 years: children often do not live at home any more and many couples weren't married in both censuses.

Consequence: In many cases we get no household support when we have to decide between possible candidates.

2. Naming of married women

Very often married women have only a first name and their husband's last name (married name) – just as we see earlier in the urban censuses.

Consequence: If the couple weren't married in the earlier census, we have only first name, age and birth place to link on, and that's rarely enough.

3. Elderly people

It is our impression that elderly people are more often registered in a separate household immediately next to the younger generation's household than in the earlier censuses, even if they are actually "aftægtsfolk". This might be due to a change in the registrational practice of separating households, or to a physical change in the historic reality – i.e. that it became more frequent that aftægtsfolk had their own independent houses and households. Please note that we do not know whether such a change actually happened!

Consequence: We should be aware that a small household of elderly people right next to a young household might be a sign of a family relation (and thus give household support to linking).

4. Naming traditions

A part from the married names mentioned above, naming traditions have also changed generally from the traditional rural names to a more urban model with rare names and/or multiple names – creating some of the difficulties which we also found when we started linking the urban population.

Consequence:

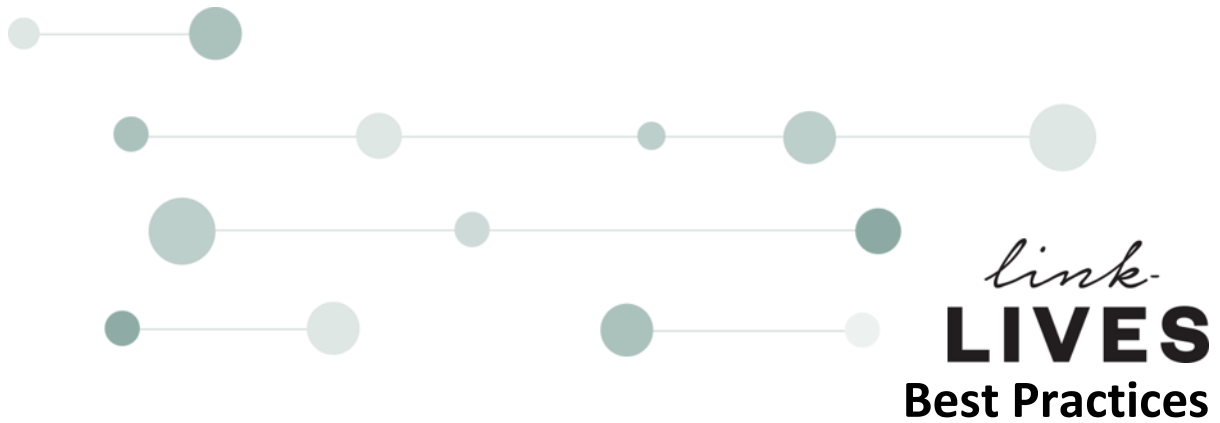
- Rare names are good for identification, but might be transcribed erroneously more often (count that in, when you search).

- Longer names tend more often to create diminutives in the sources (Vilhelmine/Mine, Nielsine/Sine) which is also a fact to take into account when searching.
- Multiple names can be replaced with only the name most frequently used.
- It is more frequent to see abbreviations (i.e. only the initial letters).
- Or it can be a combination of (some of) those: Vilhelmine Marie Sofie can be named only “Mine” or “V. M. S.” in the target census (or the other way around).

5. The enumerator

The original censuses of 1880 and 1901 are written by different persons, creating a changing environment of registrational practices even within the same parish in the same year.

Consequence: The linker should be aware that established patterns may change during a linking task (typically a parish). In the first three pages women may be registered exclusively with their married names, while in the following pages with their birth names.



Description	This document outlines the common approach to linking using ALA in the Link-Lives project. It contains universal guidance for all sources as well as specific considerations for individual sources (in tables).
Linking type	Training Data
Sources	KBH Burials 1861-1911 (specific guidance highlighted blue) Censuses 1845, 1850, 1860, 1880, 1885 (KBH only), 1901 (specific guidance highlighted yellow)
Definitions	Source1 is the census we link from, Source2 is the census we link to.
Authors	OR, AT
ALA release	ALA v. 1.3.9. 15 Mar 2021.

What's new?

1. Linking KBH burials, we have fewer variables to link on as they rarely include households or places of birth. It might therefore feel harder to link than census-census.
2. There are button changes in ALA relating to Primary vs. Secondary variables.
3. There is no Occupation button. This has been superseded by the +Secondary button and link approach.
4. A new OBS button applies to cases where we cannot make a decision.
5. Guidelines for writing conflict checking comments have been added.
6. Detailed descriptions of sources are included.
7. A linking decision flowchart is included for quick reference

Contents

From source to ALA.....	3
1. Source types.....	3
A) Censuses 1845-1901.....	3
B) Copenhagen Burials 1861-1911.....	4
2. Cleaning & preparing	4
3. Generating potential links.....	5
Principles of linking	5
1. Linking process.....	5
2. Linking decisions	5
3. Linking decisions chart	7
Search strategies.....	10
1. Linking units	10
2. Primary variables	10
A) Household/Family.....	11
B) Name.....	11
C) Birth places	12
D) Birth year	12
E) Civil status & gender	13
3. Secondary variables	13
A) Occupations.....	14
B) Residential information / Institution	14
C) Comments.....	15
Conflict checking.....	15
FAQs	15
Appendix	16
1. Quick guide linking decision tree	16
2. Purposes of linking decisions	18
A) Examples.....	19
B) Census illustration	20

From source to ALA

This section offers some background on the sources, their genesis and how they are presented in ALA.

1. Source types

A) Censuses 1845-1901

Censuses (*folketællinger*) are snapshots of the population – as individuals placed in households – on a particular day in a particular year. The Danish censuses generally cover the entire Danish population.¹ In the period 1845-1901 censuses were made every 5th or 10th year, usually on 1st February. The census enumerator gathered information from each household in the order they appeared in the street, building etc., so that we can be relatively sure that households that appear next to one another in the census were likely to be neighbours in real life. The personal information registered in this period is name, age, birth place (from 1845), civil status, position in the household and occupation (see figures below).

Gadens Navn, samt Gaardens eller Husejers Nummer og Staaens Antal.	Samtlige Personers Navne i hver Gaard, hvert Hus, eller Familie.	Alder, Høft, Civilstand, Ende eller Kærlig.	Personens Titel, Embede, Forretning, Gaardbrug, Høftingsbet, Stilling i Familien eller hvad de leve af.	At hvilke værdi
284	Gottlieb Pihl, Husejer	28 år	Handelsmand	
	Johanne Emili Pihl	24 år	Kone	
	Johanne Vilhelmine Pihl	5 år	Datter	
	Emilie Carol Pihl	2 år	Datter	
	Christiane Pihl	10 år	Datter	

Figure 1: Census 1850, Prinsessegade, Copenhagen

The censuses have been transcribed by volunteers for the National Archives' Dansk Demografisk Database-project. They are fully transcribed for the years 1845, 1850, 1860, 1880 and (almost) 1901, while the years 1855, 1870 and 1890 are not yet complete. Some proofreading of the transcriptions has occurred, but not for every one. i.e. errors occur. Because the census is recorded in headed columns, each field in the census corresponds to a field in the transcribed data. To make the data searchable and to generate potential links in ALA, Link-Lives also carry out a cleaning and standardizing process and have also added further fields to make searching simpler. E.g. a field has been added to calculate a person's year of birth, based on the age given in the census.

¹ Apart from Sønderjylland south of Kongeåen from where no census has been preserved from the German period (1864-1920)

Hovedliste.

Opgtægnis paa Folketallet den 1^{de} Februar 1901 i Ejendommen, Matr.-Nr. *11* i *St. Sofies* By i *St. Sofies* Sogn i *St. Sofies* Herred i *St. Sofies* Amt.

(Her optæges ikke de Personer, som ere midlertidigt forsvarede, se Tilføjelsen paa Bagkappen.)

Familien- nr. Læse- nr.	Samlige Personers Navn (optagne Personer). Vel Bør, ældre eller Næst, efter "Børne" eller "Fætre".	Ages- skeds- nr. Ind- føds- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fødsels- nr.	Fø
----------------------------------	---	--	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	-----------------	----

Figure 2: Census 1901, Krønge parish (opslag 26)

B) Copenhagen Burials 1861-1911

The KBH burials (*Begravelsesprotokoller*) exist for the years 1861-1940. The years 1861-1911 (inclusive) have been transcribed by Stadsarkivet and that data has been delivered to Link-Lives, cleaned and prepared and added to ALA. Image scans of the original source can be found below. Because the forms were not tabulated (i.e. were not in fixed columns), some mixing of information and fields has occurred. E.g. *Kontorist Petersens Søn Sofus* contains information that a transcriber must allocated to 4 separate data fields: Father's occupation, father's surname, deceased surname and deceased given name. *Bopæl* recorded the residence of the deceased at the time of their death, but if this wasn't a residential address (e.g. hospital) the transcriber might have recorded it in the Institution field instead. Please be aware, therefore, that the transcriptions can vary and are not always accurate.

Because of data gaps and inconsistencies in the 1880 and 1901 censuses, we are only linking the KBH burials to the 1885 and 1860 censuses at present.

7070

N ^o 1	X	2	
dagen den		Kl.	begraves
Andersson			
Gövarð Valdemar, Poul Lund			
18. Aar gl. Bopæl		Andersstræde	død d. 28/2
Ugift. Gift. Enkelskab.			
af Lempestvej		fra	F. H.
Garni		til	Gindlyfte
Jorden		Kr.	Öre

Figure 3: Skrædersvend Edvard Valdemar Povl Lund, KBH Begravelsesprotokoller, Jan 1890

No 24 X 2
 Tre dage den 3 Jan. Kl. 10 1/2 begravet
 Kantorist Petersens Son
 Julius
 1 Dag Aar gl. Bopæl i København d. 29/2
 Ugift. Gift Enkestand
 af Mørshøj fra Nørre
 Math. til Nørre
 Ass. Kirkg.
 Jorden 8 Øre

Figure 4: Kontorist Petersens søn Sofus, KBH Begravelsesprotokoller, Jan 1890

2. Cleaning & preparing

The data has been prepared in a way that makes sense for both Link-Lives and ALA linkers. Only fields relating to demographic information have been included (e.g. the cost of the burial is not relevant for our purposes). A set of instructions was given to the volunteer transcribers directing them on which field to record each piece of information in. As with any historical source, however,

be aware that errors, omissions and subjective decisions are inevitable at each stage. For example, we know that some people's residential addresses were mis-recorded in the institution field, or the burial chapel appeared in their residential address.

3. Generating potential links

Potential links are generated for each record entry (where possible) by running a rule-based algorithm based on name and age. The most likely 10 are then displayed on the right side of the ALA screen in order of probability score (0 is the closest match). These act as start-point suggestions for linkers. We might find the correct link within this list, or we might not, but we always attempt to challenge them.

Principles of linking

We link according to an agreed set of guidelines that help our team link consistently. We make a linking decision for every person in a given linking unit. The focus is on linking the right ones: we want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links. The process is as follows:

1. Linking process

1. Linking
 - a. Each linking unit is always linked by at least two linkers.
2. Conflict checking
 - a. After each finishes their linking unit, our linking coordinator identifies all the contested links (i.e. those which the linkers disagreed on) and creates a conflict file in Excel for each linker to check.
 - b. Each linker works through the conflict file to check and comment on each contested link.
3. Conflict solving
 - a. The coordinator solves any remaining contested links
 - b. A consolidated training data file is generated and stored.

Golden Linking Rules

1. Only link one person to one person
2. Always challenge a potential link
3. Always search for a competing candidate
4. Always link using primary variables first, before assessing secondary variables
5. Never use unauthorized sources outside of ALA
6. When in doubt, do not link

2. Linking decisions

We link using two sets of variables: primary variables in principle do not vary during life (or, in the case of the name, if it changes, the change can be expected to follow some relatively well-defined rules). Secondary variables may contain information that gives context to the individual's life, but

which generally changes over time. Link Lives uses the following fields to distinguish primary from secondary:

Primary variables	Secondary variables
Name	Occupation
Birth year	Address
Birth place	Comments
Family/household	Other
Gender	
Civil Status	

- Identify a candidate by searching on primary variables first.
- It may then be relevant to remove one of the criteria, expand the years or remove them entirely, in order to exhaust alternatives.
- Only then should you consider the secondary variables.

The following chart shows the different linking decisions, when to make them and how they are used.² See the Appendix for a flowchart diagram to assist in your linking, a fuller list of examples and an overview of the consequences of our linking decisions.

² When the linking decisions are processed, Maybes, Links and those with +Secondary are marked 'positive' (shaded green below) and the rest are marked 'negative' (shaded red below). This distinction simply helps to train the machine algorithm.

3. Linking decisions chart

Decision	Definition	How it is commonly used
Link	<p>When you are confident that you have found the correct link.</p> <p><i>Quantitative hint: You should be at least 99% sure.</i></p>	<p>We are usually most sure of links when primary variables all contain satisfying information for a secure link to be established. E.g. there are specific characteristics of name, age, and (for censuses) household and/or birth place.</p> <p>This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.</p>
Maybe	<p>When you are almost sure you have found the correct link, but you want to flag that there is some information missing.</p> <p><i>Quantitative hint: You should be at least 95% sure.</i></p>	<p>We tend to link “maybe” when one of the characteristics is missing but the others are solid. E.g. distinctive name, wrong age.</p> <p>Candidates often include wives who take their husbands’ names, but whose other characteristics are otherwise a match.</p>
<p>+Secondary³ can only be selected once you have clicked Maybe, Multiple, Not Found</p>	<p>When you use information secondary variables to move from a “Maybe”, “Multiple” or “Not Found” to a single plausible candidate.</p> <p>Maybe +Secondary One option is made more certain by using secondary information.</p> <p>Multiple +Secondary 2-5 options narrowed to one using secondary information.</p> <p>Not Found +Secondary > 5 options narrowed to one using secondary information.</p> <p><i>Quantitative hint: After using the secondary information, you should be at least 95% sure.</i></p>	<p>Candidates in census linking often include people with distinctive professions or trades involving some training, e.g. lærer, or occupations that are out of place in the local context. However, they are unlikely to include tjenestefolk, arbejdsmand, landmand etc.</p> <p>In other sources (e.g. burials) we might use the address or the occupation to narrow a selection to one candidate.</p>

³ For linkers who have used ALA versions before, please note there is no longer an ‘Occupation’ decision button.

Decision	Definition	How it is commonly used
Multiple	When you have a well-defined group of no more than 5 plausible candidates, but you cannot choose between them.	<p>e.g. 3 x Maren Madsens from the same parish of the same age and no household support to distinguish between them.</p> <p>e.g. Peter Jens Knudsen died in 1880, and there are 3 Peter Knudsen candidates and 2 Jens Knudsen candidates all in the same age range.</p>
Not found	<p>Not found can be applied to several situations. You do not have to mark which is which. The following is just for explanatory purposes:</p> <p>1) Unfound When you find no plausible candidate, even with the presence of distinctive characteristics.</p> <p>2) Too many When there are too many candidates (>5) to choose from and you lack primary information to narrow the pool (though see +Secondary above).</p> <p>3) Unfindable Where there is missing information that makes it impossible to identify any candidate (though see +Secondary above).</p>	<p>Examples include:</p> <p>1) Distinctive characteristics unfound e.g. Robert Anders Ludviggo Lund but despite multiple search strategies, you did not find him.</p> <p>2) Lots of local possibilities e.g. one Maren Madsen from the right parish, but her age is 4 years out, then half a dozen Maren Madsens born in neighbouring parishes with correct ages.</p> <p>3a) Common given names e.g. Marie Sørensen, wife of Hans Sørensen, no maiden name.</p> <p>3b) Vague places of birth e.g. Hans Hansen, b. 1875, Slesvig.</p> <p>3c) Missing transcription information e.g. Ubeboet, Mangler, Jens B[?], no place of birth, no household support.</p>

Decision	Definition	How it is commonly used
Unborn	If it is probable that the person was not yet born at the time of the census.	Used for people whose birth years are calculated to within a year of a census year and who are missing at the end of a family grouping. If no household support, simply mark Not Found.
OBS	<p>When you find records which you simply cannot link and would like to flag them for later analysis. E.g.</p> <p>1) two people listed in one record</p> <p>2) empty addresses or fields</p>	<p>There are several hundred records which contain more than one deceased person, especially mothers and infants or stillborns but also two children, twins etc. Look for records containing 'og' or '&'.</p> <p>Some records identify empty addresses, or the hand-writing in the source was illegible to the transcriber.</p>

Search strategies

Refer to the User Guide for full details of search functions, shortcut keys and search terms.

Census->census	KBH Burials->census
In census->census linking, we are able to focus on several primary variables to identify and place a person in an earlier census. Household position is one of the strongest indicators we have, though be aware that the young people who leave home, servants who move household and others in institutions are harder to link because of this. Search on name, year of birth and place of birth, then use household position in the search results to confirm a candidate.	In our census->census linking, the composition of the household was the strongest linking parameter we could get and we were often able to link entire households. With the KBH burials, however, we are only linking one person (deceased) back to their appearance in a census. This means it is very rare that we can use household information. In fact, we do not have even a birth place. Instead, we are restricted to a much narrower range of variables on which to make a linking decision: our primary identifiers are name, age, gender and civil status and we use secondary information to help search for the right candidate (e.g. occupation of father, residential address at time of death) and in some cases identify a link using these secondary variables.

1. Linking units

Census->census	KBH Burials->census
<p>You will be asked to link a number of records usually based on a parish or a street. The records will appear in the same order as in the original source.</p> <p>Complete all records in your linking unit: for a parish, complete all records; for a sample of a parish or street, complete the pages allocated to you in whole households (link entire households, even if they run over the page).</p> <p>Work through the records from start to finish but you are not required to link in order. If you change your mind, simply click a new link button and your data file will record the last decision made.</p>	<p>You will be asked to link a number of records based on a date range. The records will appear in the same order as in the original source, though you may notice some missing numbers in the ID field, as these had a death date in a different year. Do not be concerned about this.</p> <p>Work through the records on the pages allocated to you, as you would have in census->census linking. Since there are no households, there is no need to start or finish before or after the sample.</p> <p>Work through the records from start to finish but you are not required to link in order. If you change your mind, simply click a new link button and your data file will record the last decision made.</p>

2. Primary variables

Primary variables in principle do not vary during life (or, in the case of the name, if it changes, the change can be expected to follow some relatively well-defined rules).

A) Household/Family

Census->census	KBH Burials->census
<p>If the household fits and there is a minor discrepancy in the personal variables, the link is good</p> <ul style="list-style-type: none"> • However, discrepancies in the personal variables can indicate that a new person has entered the household, e.g. a new wife/another child living at home. • Household information is given high priority only for the family members of the household, e.g. a servant being in the same household is not a strong identifier. 	<p>Although there is no household context available, we might find familial connections through a named parent or spouse. The link is therefore good if this household support is present.</p>

B) Name

Generally, you should expect the name to be precise for a good link (apart from spelling variations). If the name does not fit, we are hesitant to link unless we have supporting information to confirm it. Names change more often between census periods in the urban areas than in rural areas so that should be taken into consideration when searching for names.

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the same person without having matching first names – and without any of the names being wrong!
 - Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
 - Watch out for common diminutives: Vilhelmine/Mine, Nielsine/Sine etc.
 - First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names than men
 - married women taking the husband's last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen's children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription

C) Birth place

Census->census	KBH Burials-> census
<p>Generally, you should expect the birth place to be exact for a good link. Some systematic errors have been identified though:</p> <ul style="list-style-type: none"> the further away, the less precise the birth place: a person not born in the parish will have a less specific place of birth (maybe just the amt is listed or even just the country, e.g. Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong probably increases with the distance. an oldest child was often born in another parish (before the parents settled down) but in the census is actually registered with the birth place where he/she actually grew up instead. parishes within the same "pastorat" are more often confused with each other than others. the administrative levels of birth places can vary from census to census (e.g. farm, village, parish, town, county, country) <ul style="list-style-type: none"> Bear it in mind as a possibility that a birth place is not always recorded as a parish name. A sparse use of contemporary geographic search engines is recommended. Not for every search, but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or if they were part of the same parish. <ul style="list-style-type: none"> https://www.danishfamilysearch.dk/ http://www.krabsen.dk/stednavnebase/ Google maps Birth places are often recorded as variants of "her i sognet", "samme sted" or "ditto", i.e. text strings which do not include the actual name of the place. <ul style="list-style-type: none"> Look for the person in the birth parish (as Residence parish) in order to account for potential matches with "her i sognet" as birth place. Look for the person in the standardized birth parish (Parish Std) in order to account for potential matches which appear simply with 'ditto' marks. <p>i) Special characteristics for birth places in URBAN areas</p> <p>In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility, causing:</p> <ol style="list-style-type: none"> the imprecision mentioned above (because of distance) and a tendency to register birth <i>places</i> rather than birth <i>parishes</i>, probably because of the census enumerator's lack of knowledge of (and interest in) exact parish names in the surrounding rural areas. 	<p>Birth places are rarely found in this source, though if a birth place is listed (even in the comments field) you should treat it as a primary variable. Be aware that the comments field sometimes contains information added by transcribers. Only link using this field if it is clear that it is from the original source.</p>

D) Birth year

The age of a person in the sources is converted into ALA as a year of birth, i.e. [census year]-[age]=[birth year]. E.g. a 23 year old man in 1880 will be given the birth **year** 1857, even though his birth **date** might be any time before or after the census, which was usually held on 1 February.

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

General age searching recommendations

1. Age 0-49: Use calculated birth year "+3"
2. Age 50-64: Use calculated birth year "+5"
3. Age 65-: Use calculated birth year "+10"
4. To search for any age, run a 'blank age search' using #

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range.

i) Unborns

Census->census	KBH Burials->census
Used for people whose birth years are calculated to within a year of a census year and who are missing at the end of a family grouping. If there is no household support, simply mark Not Found.	Children who died as newborns or infants (dødfødt or similar) are pre-marked Unborn if they are under 1 year. Check their age at death in months, weeks and days to avoid searching for them in the census unnecessarily. If a deceased person has no age at all, look at their death cause and Comments in case they suggest death in infancy (e.g. dødfødt). If so, mark Unborn. If not, continue to search as you would normally.

E) Civil status & gender

These are considered to be primary variables. They are often used to exclude possible matches.

Bear in mind, though, that

- civil status can change naturally between censuses, e.g. unmarried to married, married to widowed or even single to divorced or widowed.
- there can of course be errors in the data, e.g. divorced people registered as widow(er)s.

3. Secondary variables

All variables in ALA that are not listed above as primary variables are considered secondary information. In all sources we have access to extra fields containing information that might be relevant for finding someone and in certain cases to make linking decisions. These include:

Census->census	KBH Burials->census
Occ. information: an amalgamation of the data derived from the household position and occupation columns across all censuses. Searching for unusual or distinctive occupations can help identify a candidate.	Occ. information: this field records the occupation of the deceased person.
Place, Res. parish, Res. county, Res. information, Plot no.: these record the parish or sometimes street name and	Add. Occupations: this records the occupation of the deceased's spouse or parent, which can help you to find them in the free text searches.
	Res. information: this records the street name and number of the deceased at the time of death, and

number of a household, and are weak indicators of a link. As the century wore on, it became increasingly unusual for individuals to stay living at the same address for their entire lives.

may or may not be the same place as where they were living in the census you are linking to.

Institution: this records the institution where the deceased was living (possibly temporarily) at the time of death (similar to Res. information).

Comments: this contains a mixture of information derived from the source and from the transcribers.

Important. Non-primary-variable fields should be used with caution. **Make your decision based on the primary variables first.** Only then should you consider secondary variables. The decision tree in the Appendix can help with this.

A) Occupations

Census->census	KBH Burials->census
<p>The occupation of a person listed in the census appears under the heading 'Occ. information', which also includes a person's position in the family. For this reason, be vigilant not to link on occupation as a primary variable, even though it appears in same place as other primary variables (e.g. household position).</p> <p>In any case, the most common occupational titles such as "tjenestekarl", "arbejdsmand", "gårdmand" and "tyende" actually have no value as identifiers at all. Occupation types should be distinctive to be useful identifiers. If you link on a distinctive occupation, be sure to use the +Secondary button.</p>	<p>This is a set of fields that records the occupation of either the deceased, their spouse or their parent. The occupation of the deceased appears under the heading Occ. Information in Additional information on the left side of the ALA screen. In a column next to it (called Add. Occupations) you will find the occupations of their relatives (e.g. ægtefælles or fars erhverv).</p> <p>The most common occupational titles such as "tjenestekarl", "arbejdsmand", "gårdmand" and "tyende" actually have no value as identifiers at all. Occupation types should be distinctive to be useful identifiers.</p>

B) Residential information / Institution

Census->census	KBH Burials->census
<p>Potential links and search results display both the stated place of birth (including 'her i sognet' etc.) and the parish in which the census was taken. At the bottom right hand side of the screen, under Additional Personal Information, you will see the specific address of the person highlighted (shown under Res. information), which might contain data such as the matrikel no, street address or farm/village name within the residential parish. The Place field then shows information on the residential parish and county in which they are living.</p> <p>Address information is not considered very secure since people moved so often, but it can be a useful indicator if the address is distinctive enough.</p>	<p>The Res. information field contains either a street name, a street name and property number, a number or is just blank. Institutions appear in a separate column along with their addresses. Assess both together as it is possible to have a residential address and an institution. E.g. a child, whose parents are living at Gothersgade, died in the Fødselsstiftelsen. Both Gothersgade and Fødselsstiftelsen are recorded in his burial entry. In these cases it is likely the residential information is most useful for connecting census to burial records, but use your judgement. Transcription errors can often be found in these fields. The closer the source is in time to the census, the greater weight you should give the contents of this field.</p>

C) Comments

Census->census	KBH Burials->census
There is no comment field in the census but sometimes comments were inserted into the Occ. information field (formerly called HH position) along with occupations and household position. Transcribers sometimes add information in this field too: be careful not to make decisions based on these.	This field appears in Full Person Info (scroll right to the bottom) and if you click on it, a pop up window displays its full content. The field can contain a mixture of extra information in the record and the comments of the archival transcriber. Here you might find information that helps you identify the candidate in the census, but be aware of the mixed nature of the information. Where it is possible to identify, you should not use anything added by the transcriber to make a linking decision.

Important reminders

1. Match the primary variables (if available) first, then use the secondary variables to potentially help secure a link.
2. Do not use the secondary variables before the primary ones.

Conflict checking

Our conflict solver takes the final decision on contested links – those where both linkers disagree. All information provided in the comments section helps, but the conflict solver has helpfully outlined the following to be most useful:

1. If the place of birth is a place name in the parish we are looking for, please write that. It saves me a lot of time.
2. Have you used the reverse search to get to your result? If so it should be clearly stated.
3. How did you search for your result? If you did something special to find it, which you might think the other linker and the solver wouldn't think of, share it!
4. If you have a link and the other person says it's a multiple, and you decided that you still think your link is the best, then say in the comments why you still think it's a link.
5. If you cannot decide, please write that in the comments and why.
6. I am in love with the commenters that put the id of the other multiples in the comments and shall grant them my firstborn 😊 It saves me a ton of work, though it's, in a strict sense, not necessary because I will do a search in ALA myself anyway.

FAQs

- **I'm worried I am linking incorrectly or am making the wrong decision**

Do not be concerned if you are unsure of your linking decisions as we have built in checks at many stages of our process. The Link Lives approach has adopted the principle of always having two linkers link each unit. Any 'contested' links (those the two linkers do not agree on) are then solved by a third linker (the conflict solver). Interestingly, when each pair of link files are compared, there are on average only 10-15% of links that are not identical, which means that 85-90% of the time we are all linking in the same way. Those 10-15% of links are

also usually halved after the conflict-checking stage, so we can be confident that up to 95% of our links are 'uncontested'.

- **Can I use other sources outside ALA to help me link?**

No, in order to train the algorithm you must never use sources outside ALA apart from those listed by name in 2.Primary Variables/C) Birthplaces.

- **Can't I just accept the potential link with the lowest score rather than searching each time?**

We always challenge the potential links to make sure we've been thorough. Ideally the algorithm is good enough to find each link, but then we wouldn't need human linkers! Eventually we test how good the algorithm is by comparing its links to those made by humans. Humans are good at spotting a subtle spelling error, or age difference, that a more rigid rule-based algorithm might miss.

- **I've found duplicate records. What should I do?**

A duplicate record is where you might find the same person (though often it's a household) listed twice. This can often be a transcription error (two transcribers transcribed the same household, or transcribed it twice) or a duplication in the source itself.

Simply add the details to our [duplicates file](#) for us to follow up on later.

- **I've found a fun or interesting record**

Add it to our Freak Links file

- **What do I do with a blank record?**

Sometimes there is no information in a record. Simply mark it with the >1 button for us to deal with later.

- **My ALA has shut down with no warning. Why?**

Check if you have left one of your link (.csv) files open. If so, and if you tried to save some linking decisions right before ALA shut down, this is the explanation. Close your links file and restart ALA.

If it was not because of the links-file, please describe in as much detail as possible what you did just before ALA shut down – and try to reproduce the error. Please send a mail with your description to Asbjørn or Atlanta (at@sa.dk / olivia@hum.ku.dk).

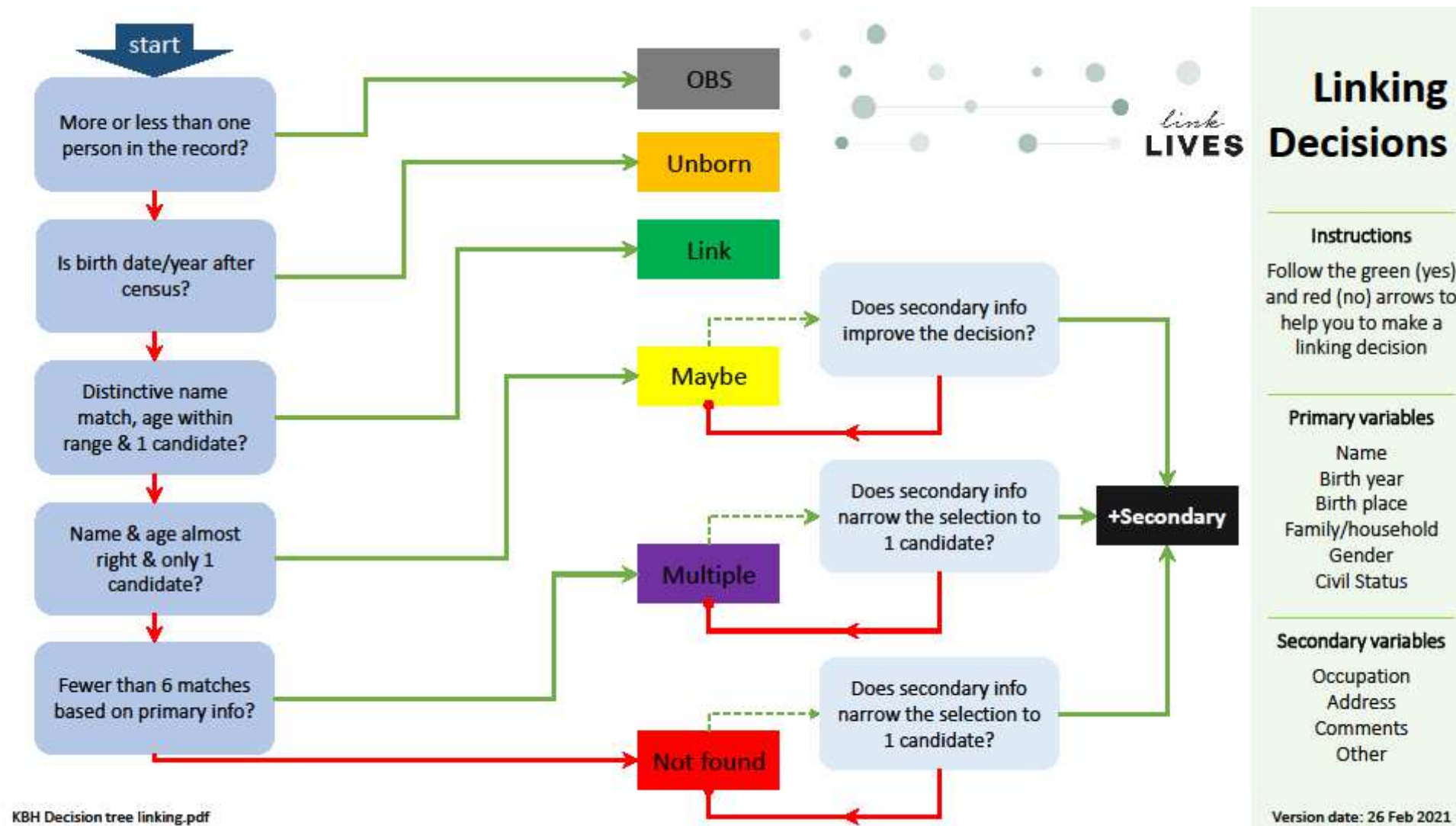
- **What kind of comments should I write in my conflict checking file?**

See the section above called Conflict Checking for more guidance.

Appendix

1. Quick guide linking decision tree

This flow diagram is a summary of the descriptions above, and should act as a guide when linking. Follow the green (yes) and red (no) lines to help you to make a decision.



2. Purposes of linking decisions

This table shows the consequences of our linking decisions.

Decision	Purpose/function
Link	<ul style="list-style-type: none"> • Trains the machine that this is a secure link. • Data is used to train the algorithm. • Helps to build life courses. • The positive decisions (ie. maybe, link etc) are usually merged for Machine Learning purposes.
Maybe	<ul style="list-style-type: none"> • Offers the ML team evidence of a link that is missing something. They are then able to identify what that missing information is and use that as a dataset in itself. • The positive decisions are usually merged for ML purposes.
+Secondary can only be selected once you have clicked Maybe, Multiple, Not Found	<ul style="list-style-type: none"> • Because the 'accepted' approach to linking is using only unchanging variables, BRE would like to capture all instances where we use OTHER information in the source to make a link.
Multiple	<ul style="list-style-type: none"> • Trains the machine that there is more than one plausible candidate based on the information available. They then assess what data is missing that is preventing a "link". • Conflict solvers use this to help them investigate a finite potential pool of candidates.
Not found	<ul style="list-style-type: none"> • Trains the machine that there is no link for this candidate based on the information available. <p>Even though it feels like a negative decision, rest assured that there is a good chance that we'll find this person later on and link him/her based on adding information from other sources. You do not need to mark things as "multiple" or "maybe" to show you have made an effort 😊.</p>
Unborn	<ul style="list-style-type: none"> • These are excluded from our training data.
>1	<ul style="list-style-type: none"> • We analyse this dataset later using more advanced techniques, so all we need to do is flag records for this purpose. • We link on a one-to-one basis, so if there is more or less than one person in a record it cannot be included in our standard training data.

A) Examples

Here are some examples of how to use the different linking decisions in different circumstances.

Source 1 (linking from)	Source 2 (linking to)	Reasoning	Decision
Maren Henriette Poulsen (wife of Rasmus Poulsen) b. 1850 in Græsted.	Maren Henriette Ferdinandsdatter, tjenestepige b. 1850 in Græsted.	No other Maren Henriettes in the parish or amt but with no household support there's a small chance there is more than one Maren Henriette of that age born in the parish.	Maybe
Hans Krøll's wife Christine Krøll, b 1820 Krønge.	Hans Krøll's wife Kirstine Jensdatter, b 1824 Krønge.	We cannot be sure whether he remarried, because we don't have her birth name and the age is slightly differing. We take "Maybe" to flag the doubt, but we're not REALLY worried.	Maybe
Laura Nielsen født Petersen in 1863 in Gilleleje.	There are two Laur* Pe*ers* in Gilleleje: Laure Jensine Kristine Pedersen and Laura Johanne Marie Pedersen, both 2 years out. Another Laura Petersen is the same age but born in the neighbouring parish.	There is a small set of identifiable candidates and we are relatively sure the right link is one of them, but we are missing household information and further middle name identifiers.	Multiple
Petrine Petersen, wife of Laurits Petersen, b. 1861 Gilleleje. The source is also Gilleleje, so it's likely she would be there 20 years before but with her maiden name.	There are 2 Petrine names registered in the parish within her age range and a further 3 Petronella variations.	A lack of information therefore stops us from deciding on one candidate, even though the name is quite unusual.	Not found
Peter Nielsen b 1853 Græsted.	11 people are returned on a search for name, age and Græsted as birth place, some with multiple given names, but no occupation or household support to distinguish between them.	A number of people are identifiable but a lack of further distinguishing information stops us from narrowing the pool to less than 6	Not found
Robert Anders Ludviggo Lund (distinctive name).	Cannot be found despite multiple versions of his name/age/birth place searches.	This is not lack of information but some anomaly in the transcription or source (e.g. he was out of the country, not enumerated, not transcribed, source illegible etc.)	Not found
Marie Hansen (married to Johan	We cannot know what her maiden name is, even though we have a definite place of	We lack information on her surname to even begin to be able to search for her.	Not found

Hansen) b. 1875 in Stege living in Sorø.	birth, because her name is too common.		
Peter Hansen b. 1875 in Årre, tjenestekarl.	There are 18 Peter Hansens in Ribe but none born in Årre. Several are born in neighbouring parishes but none in 1874-6	With no distinguishing characteristics (e.g. unusual name(s)), and no household support, it is not possible to reduce the pool of candidates.	Not found
Burials examples to follow.			

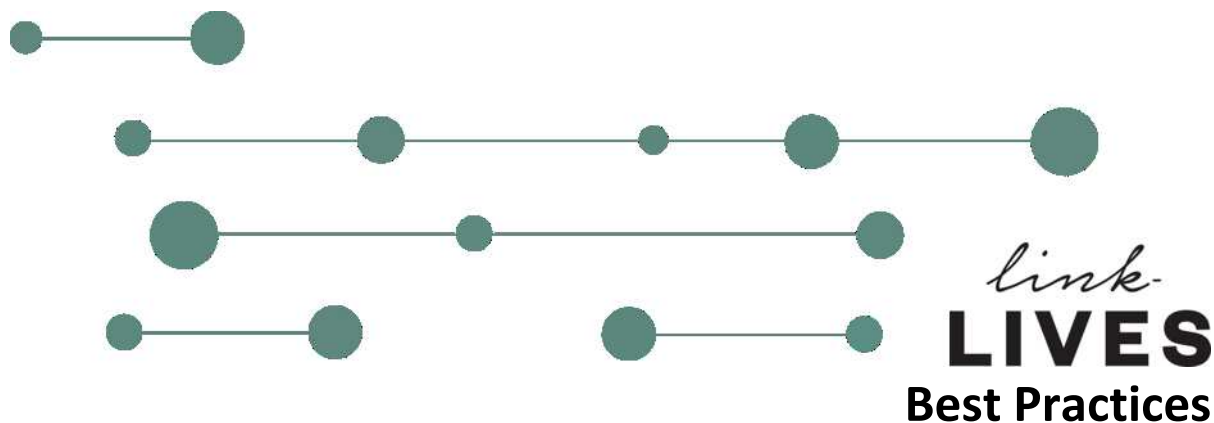
B) Census illustration

The screen grabs below show Johan Jensen's household in the urban area of Odense in 1850 and in 1845. Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The wife and children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	smedemester, huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn



Description	This document outlines the common approach to linking using ALA in the Link-Lives project. It contains universal guidance for all sources as well as specific considerations for individual sources (in tables).
Linking type	Training Data
Sources	<p>Parish record (PR) Baptisms 1813-1917 (nationwide) (specific guidance in blue)</p> <p>PR Marriages 1813-1917 (nationwide) (specific guidance in yellow)</p> <p>PR Burials 1813-1917 (nationwide) (specific guidance in green)</p> <p><i>You will find the following sources in 04 Best Practices_final.pdf:</i></p> <p><i>KBH Burials 1861-1911</i></p> <p><i>Censuses 1845, 1850, 1860, 1880, 1885 (KBH only), 1901</i></p>
Definitions	Source1 is the census we link from, Source2 is the census we link to.
Authors	OR, AT
ALA release	ALA v. 1.4.0 28 June 2021.

What's new?

1. Two person search function added
2. Linking by convenience approach
3. New OBS button definitions to clarify its use
4. Linking forwards for PR Marriages
5. Potential links temporarily not available for PR Burials
6. Detailed descriptions of sources

Contents

From source to ALA.....	3
1. Source types.....	3
A) PR Baptisms 1814-1917	4
B) PR Marriages 1813-1917.....	4
C) PR Burials 1814-1917	5
2. Cleaning & preparing	7
3. Generating potential links.....	7
Principles of linking	7
1. Linking process.....	7
2. Linking decisions	8
3. Linking decisions chart	9
Search strategies.....	12
1. Linking units	12
2. Primary variables	13
A) Household/Family	13
B) Name.....	14
C) Birth place	15
D) Birth year.....	15
E) Civil status & gender	16
3. Secondary variables	17
Conflict checking.....	18
FAQs	18
Appendix	19
1. Quick guide linking decision tree	19
2. Purposes of linking decisions	21
A) Examples	22
B) Census illustration.....	23
3. Census to census details	24
A) Birth places in censuses	24
B) Names in censuses	24

From source to ALA

This section offers some background on the sources, their genesis and how they are presented in ALA.

1. Source types

Parish Records are continuous, dynamic registers containing information on vital events of the parishioners. The core vital events registered are Baptisms, Marriages and Burials – for some periods supplemented by Confirmations, Arrivals and Departures. In Link-Lives generally, and in the ALA-linking specifically, we will concentrate on linking only the core vital events of Baptisms, Marriages and Burials – and this is why this Best Practices document only focuses on them.

Parish records were to be made by the vicar of every parish from ca. 1645-46, but many of the oldest ones have not survived. From most parishes we have parish records from the 18th century onwards, but from ca. 1813 they have been preserved from all parishes. From this year it was decided that parish records had to be kept in two versions, and it was forbidden to have the two versions under the same roof at night, to avoid potential destruction by fire by saving a duplicate. The two versions would generally be written by the vicar (præsten) and the parish clerk (deggen). Another major change in the 1813 reform was that parish records now had to be written in pre-printed books with defined columns for the single pieces of information; in earlier times, the parish records had been written just as the priest wanted, i.e. often in a much more narrative form.



The parish record books of Junget before and after the 1813 reform which introduced the pre-printed, schematized books.

In archival language, the parish records versions have the following names:

- Before 1813: Enesteministerialbog (the only version)
- After 1813: Hovedministerialbog (the vicar's version)
- After 1813: Kontraministerialbog (the parish clerk's version)

For the time after 1813, it will normally be the kontraministerialbog that has been digitized and transcribed. The two versions should be completely identical (the two writers had to “konferere” their books regularly), but occasionally of course there will be differences. This will of course generally have no importance for us, working with the transcriptions and linking on the basis only of those.

Another major change happened in 1892, this time in the content of the books. From this year onwards, the scribes had to include quite a lot more information on the individuals. This will be shown more in detail in the chapters A-C below.

Unfortunately a general problem is that not all information, that might be of value for the linking process, has been transcribed. This goes for the primary variable *marital status* and for the secondary variables *occupation* and *address*. So, even if this information is (sometimes) written in the original source, we cannot use it for linking because it hasn't been transcribed.

A) PR Baptisms 1814-1917

Født					Kvindekøn.		
Nr.	Hvor og Datum.	Barnets fulde Navn.	Fødsels Datum enten i Kirken eller Hjemmet.	Fødselsdatoens År, Måned, Dage og Tidspunkt.	Barnets År, Måned og Tidspunkt.	Overført i det aln. Jord. Reg.	Bemærkninger.
1	1863 Junet	Mette Kristine Christensen	1863	1863	1863	1863	1863
2	1863 Junet	Mette Kristine Christensen	1863	1863	1863	1863	1863

The two first girls baptized in Junget parish 1863.

Født					Kvindekøn.		
Nr.	Fødselsdato og Datum.	Hvor født?	Barnets fulde Navn.	Fødselsdatoens År, Måned, Dage og Tidspunkt.	Dødsdato.	Fødselsdatoens År, Måned og Tidspunkt.	Anmærkninger.
1	1900 14 ^{de} Maj	Junget	Mette Kristine Christensen	1900	1900	1900	1900
2	1900 14 ^{de} Juni	Junget	Mette Kristine Christensen	1900	1900	1900	1900

The two first girls baptized in Junget parish 1900.

B) PR Marriages 1813-1917

Egteskab					Egteskab		
Nr.	Egteskabets År, Måned, Dage og Tidspunkt.	Egteskabets År, Måned, Dage og Tidspunkt.	Egteskabets År, Måned, Dage og Tidspunkt.	Egteskabets År, Måned, Dage og Tidspunkt.	Egteskabets År, Måned, Dage og Tidspunkt.	Egteskabets År, Måned, Dage og Tidspunkt.	Egteskabets År, Måned, Dage og Tidspunkt.
1	1861	1861	1861	1861	1861	1861	1861
2	1861	1861	1861	1861	1861	1861	1861

Two married couples from Junget parish 1861.

Ægte- viede.

199

Nr.	Brudegommens fulde Navn eller fornavn Dødsdato eller ved Eftersyn i Kirkebogen, hans Fødsels- og -dag, Stilling og Opholdssted samt hans Forældres fulde Navn, Stilling og Bopæl. (Brudegommens fulde Navn eller fornavn Dødsdato eller ved Eftersyn i Kirkebogen, hendes Fødsels- og -dag, Stilling og Opholdssted samt hendes Forældres fulde Navn, Stilling og Bopæl.)	Brudebruds fulde Navn eller fornavn Dødsdato eller ved Eftersyn i Kirkebogen, hendes Fødsels- og -dag, Stilling og Opholdssted samt hendes Forældres fulde Navn, Stilling og Bopæl. (Brudebruds fulde Navn eller fornavn Dødsdato eller ved Eftersyn i Kirkebogen, hendes Fødsels- og -dag, Stilling og Opholdssted samt hendes Forældres fulde Navn, Stilling og Bopæl.)	Forlovere fulde Navn, Stilling og Bopæl.	a) Naar blev Tilværing bedt? b) Naar fandt Tilværing Sted? c) Hvis Lyngvej, ej har fundet Sted, angives Hjemmet herfor.	Vedkommens Aar og Datum samt Stedets (Bogst.) hvor den foregik.	a) Af hvilken Prydel-Vedkommens er foretaget. (Hvis den er foretaget af en anden Prydel, angives Hjemmet herfor.) b) Om Vedkommens er foretaget i Kirken eller i Hjemmet. c) I hvilken Tilfælde angives Hjemmet herfor.	Anmærkninger. (Her anføres eventuelt Erklæring om, i hvilken Tilfælde Vedkommens er foretaget, hvilken Erklæring, hvilken Erklæring har i Bogen.)

Nr.	Peter Pedersen Andersen Ungkarl af Kjøbenhavn, født i Kjøbenhavn, den 3 ^{de} December 1834; Søn af Gnd. Peter Pedersen, Smed og hustru Johanne Jacobsen af Lund By, Rindings sogn.	Hedraa Pedersen Pige af Ribe, født i Ribe den 25 ^{de} September 1834; Datter af Gnd. Niels Pedersen og hustru Johanne Jensen af Ribe.	Gnd. Niels Pedersen af Ribe i Gnd. Niels Jensen af Ribe.	a) 17 ^{de} b) 17 ^{de} c) 17 ^{de}	1900 6 ^{de} April i Ribe	af Prydel Niels b. T. Nielsen i Ribe	
-----	--	---	---	---	---	--	--

A married couple from Junget parish 1900.

C) PR Burials 1814-1917

The burial records are generally the vital events from the PRs with least information.

Often there is information only on the deceased person whereas Baptisms and Marriages always include more than one person. This characteristic of the Burials apply especially to adult males: children will often have information about their father/parents, and women will often have information on their husbands (if married/widowed). Adult males, unfortunately, in a large part of the period, have information only on name, age and residence parish – as marital status, occupation and address have not been transcribed.

As the examples from the original source below show, there is a gradually increasing amount of data available during the period. In the beginning, we generally have the following information to link on:

- Name
- Age
- Burial parish
- Husband's name (if a married/widowed woman)
- Father's/parents' names (if a child)

Gradually, the following information will join the ones mentioned above:

- Birth place
- Birth day
- Wife's name (if a married/widowed man)
- Parents' names (all)

After 1892 all this information will be generally available (except maybe for in the largest cities) while before 1892 often only the first batch of information will be registered. Fewer than 30% of the records contain birth places before 1892, for example, but more than 80% from 1892 onwards.

40

Døde. Mandkøn.

No.	Dødsda- gen.	Begravelses- dagen.	Den Dødes Fø- de og Tilnavn.	Stand, Haandtering og Opholdsst.	Alder. almentel. Jærl. Kjørlig.	Bemærkning.
2	17-1-1823	26-1-1823	Rasmus Erichsen	Huus og Muurmester i Junget	44	305
3	19-5-1863	28-5-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75	305
4	19-3-1901	28-3-1901	Hans Kristian Jensen	Indsieder i Junget. Født i Gassum	47	305

Buried men from Junget parish 1823, 1863, and 1900-1901. There is no difference between the registration in 1823 or 1863, even if the registration form now says that the deceased's birth place and father's name should be registered – if obtainable. But the registration of 1900 contains information on birth place, parents' names and spouse's name – data which has all been transcribed.

144

Døde. Mandkøn.

No.	Dødsdagen.	Begravelsesdagen.	Den Dødes Fø- og Tilnavn.	Den Dødes Stand, Haandtering og Opholdsst. samt hans Færdighed og hans Færdighed eller, hvis han er ugift, hans Moders Navn, forsaavidt disse ikke tidligere er registrerede.	Alder.	Den ugift i det aln. Jærl. Kjørlig.	Bemærkning.
2	17-1-1823	26-1-1823	Rasmus Erichsen	Huus og Muurmester i Junget	44	305	
3	19-5-1863	28-5-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75	305	
4	19-3-1901	28-3-1901	Hans Kristian Jensen	Indsieder i Junget. Født i Gassum	47	305	

Døde. Mandkøn. 223

No.	Dødsdagen (År og Dato)	Begravelsesdagen (Kirkens navn, År og Dato)	Den Dødes Fø- og Tilnavn (som vidt muligt efter fornavn, efternavn eller Fødselsstet eller ved Efternavn i Efternavn)	Den Dødes Stand, Haandtering og Opholdsst. samt hans Færdighed og hans Færdighed eller, hvis han er ugift, hans Moders Navn, forsaavidt disse ikke tidligere er registrerede.	Alder. (Når den afledte var over 1 År, angives Alderen med det fulde År.)	Af hvilken Præst Begravelsen er foretaget.	Bemærkninger.
2	21 August 1900	28 August 1900	Rasmus Erichsen	Huus og Muurmester i Junget	44		
3	19-5-1863	28-5-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75		
4	19-3-1901	28-3-1901	Hans Kristian Jensen	Indsieder i Junget. Født i Gassum	47		

Examples of information in the original source (from the above illustrations):

No.	Death date	Bur. date	Name	Occupation and family	Age
2	17-1-1823	26-1-1823	Rasmus Erichsen	Huus og Muurmester i Junget	44 Aar
5	5-4-1863	12-4-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75 År
1	21-3-1901 Junget by	28-3-1901	Hans Kristian Jensen	Indsieder i Junget. Født i Gassum Nørhald Herred; Søn af Hmd. Hans Kr. Jensen og Hustru Abelone Kristensen af Gassum. Gift med Ane Kirstine	47 Aar

				<i>Svenningsen. Sidste fælles Bopæl Junget.</i>	
--	--	--	--	---	--

2. Cleaning & preparing

The data has been prepared in a way that makes sense for both Link-Lives and ALA linkers. The PR data was transcribed in Asia by a company paid by Ancestry, and we must expect that a set of instructions was given to the transcribers directing them on which field to record each piece of information in. As with any historical source, however, be aware that errors, omissions and subjective decisions are inevitable at each stage. For example, for almost all fields we find misplaced or misread information from other fields.

3. Generating potential links

Potential links are generated for each record entry (where possible) by running a rule-based algorithm based on name and age. The most likely 10 are then displayed on the right side of the ALA screen in order of probability score (0 is the closest match). These act as start-point suggestions for linkers. We might find the correct link within this list, or we might not, but we always attempt to challenge them.

Principles of linking

We link according to an agreed set of guidelines that help our team link consistently. The focus is on linking the right ones: we want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links. The process is as follows:

1. Linking process

1. Linking
 - a. Each linking unit is always linked by at least two linkers.
2. Conflict checking
 - a. After each finishes their linking unit, our linking coordinator identifies all the contested links (i.e. those which the linkers disagreed on) and creates a conflict file in Excel for each linker to check.
 - b. Each linker works through the conflict file to check and comment on each contested link.
3. Conflict solving
 - a. The coordinator solves any remaining contested links
 - b. A consolidated training data file is generated and stored.

Golden Linking Rules

1. Only link one person to one person
2. Always challenge a potential link
3. Always search for a competing candidate
4. Always link using primary variables first, before assessing secondary variables
5. Never use unauthorized sources outside of ALA
6. When in doubt, do not link

2. Linking decisions

We link using two sets of variables: primary variables in principle do not vary during life (or, in the case of the name and marital status, if it changes, the change can be expected to follow some relatively well-defined rules). Secondary variables may contain information that gives context to the individual's life, but which generally changes – or can change unpredictably – over time. Link Lives uses the following fields to distinguish primary from secondary:

Primary variables	Secondary variables
Name	Occupation
Birth year	Address
Birth place	Comments
Family/household	Other
Gender	
Marital Status	

In the parish record transcriptions we get very few secondary variables, even when they are in the original source (occupation, for instance, is not transcribed). This means that the following (general) description of how and when to use the primary and secondary variables in searching is less relevant in the PR context than when linking other sources, but we leave them here, as they DO apply at least to the different geographically defined fields (i.e. address possibilities).

- Identify a candidate by searching on primary variables first.
- It may then be relevant to remove one of the criteria, expand the years or remove them entirely, in order to exhaust alternatives.
- Only then should you consider the secondary variables.

The following chart shows the different linking decisions, when to make them and how they are used.¹ See the Appendix for a flowchart diagram to assist in your linking, a fuller list of examples and an overview of the consequences of our linking decisions.

¹ When the linking decisions are processed, Maybes, Links and those with +Secondary are marked 'positive' (shaded green below) and the rest are marked 'negative' (shaded red below). This distinction simply helps to train the machine algorithm.

3. Linking decisions chart

Decision	Definition	How it is commonly used
Link	<p>When you are confident that you have found the correct link.</p> <p><i>Quantitative hint: You should be at least 99% sure.</i></p>	<p>We are usually most sure of links when primary variables all contain satisfying information for a secure link to be established. E.g. there are specific characteristics of name, age, and (for censuses) household and/or birth place.</p> <p>This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.</p>
Maybe	<p>When you are almost sure you have found the correct link, but you want to flag that there is some information missing.</p> <p><i>Quantitative hint: You should be at least 95% sure.</i></p>	<p>We tend to link “maybe” when one of the characteristics is missing but the others are solid. E.g. distinctive name, wrong age.</p> <p>Candidates often include wives who take their husbands’ names, but whose other characteristics are otherwise a match.</p>
<p>+Secondary² can only be selected once you have clicked Maybe, Multiple, Not Found</p>	<p>When you use information secondary variables to move from a “Maybe”, “Multiple” or “Not Found” to a single plausible candidate.</p> <p>Maybe +Secondary One option is made more certain by using secondary information.</p> <p>Multiple +Secondary 2-5 options narrowed to one using secondary information.</p> <p>Not Found +Secondary > 5 options narrowed to one using secondary information.</p> <p><i>Quantitative hint: After using the secondary information, you should be at least 95% sure.</i></p>	<p>Candidates in census-to-census linking often include people with distinctive professions or trades involving some training, e.g. lærer, or occupations that are out of place in the local context.</p> <p>In parish records we might use geographical indicators of residence to narrow a selection to one candidate.</p>

² For linkers who have used ALA versions before, please note there is no longer an ‘Occupation’ decision button.

Decision	Definition	How it is commonly used
Multiple	When you have a well-defined group of no more than 5 plausible candidates, but you cannot choose between them.	e.g. 3 x Maren Madsens from the same parish of the same age and no household support to distinguish between them. e.g. Peter Jens Knudsen died in 1880, and there are 3 Peter Knudsen candidates and 2 Jens Knudsen candidates all in the same age range.
Not found	<p>Not found can be applied to several situations. You do not have to mark which is which. The following is just for explanatory purposes:</p> <p>1) Unfound When you find no plausible candidate, even with the presence of distinctive characteristics.</p> <p>2) Too many When there are too many candidates (>5) to choose from and you lack primary information to narrow the pool (though see +Secondary above).</p> <p>3) Unfindable Where there is missing information that makes it impossible to identify any candidate (though see +Secondary above).</p> <p>4) Presumed dead When, linking forward, you expect this person to be dead (e.g. if linking a PR marriage to the following census, and you find the wife as a widow)</p>	<p>Examples include:</p> <p>1) Distinctive characteristics unfound e.g. Robert Anders Ludviggo Lund but despite multiple search strategies, you did not find him.</p> <p>2) Lots of local possibilities e.g. one Maren Madsen from the right parish, but her age is 4 years out, then half a dozen Maren Madsens born in neighbouring parishes with correct ages.</p> <p>3a) Common given names e.g. Marie Sørensen, wife of Hans Sørensen, no maiden name.</p> <p>3b) Vague places of birth e.g. Hans Hansen, b. 1875, Slesvig.</p> <p>3c) Missing transcription information e.g. Ubeboet, Mangler, Jens B[?], no place of birth, no household support.</p>

Decision	Definition	How it is commonly used
Unborn	If it is probable that the person was not yet born at the time of the census.	Used for people whose birth years are calculated to within a year of a census year and who are missing at the end of a family grouping. If no household support, simply mark Not Found.
OBS	<p>When you find records which you simply cannot link and would like to flag them for later analysis:</p> <ol style="list-style-type: none"> 1) more than one person (e.g. two people listed in one record) 2) non-persons (e.g. empty addresses or fields) 3) duplicates (e.g. the same event registered or transcribed twice) 	<p>In the CPH Burials there are several hundred records which contain more than one deceased person, especially mothers and infants or stillborns but also two children, twins etc. We do NOT, though, expect this to be the case with the PR Burials.</p> <p>Some records identify empty addresses, or the hand-writing in the source was illegible to the transcriber.</p> <p>There are multiple reasons for the presence of duplicates in the PR data. Of course in the perfect world every individual event would exist only one time in the PR dataset, but for various reasons a percentage of them were registered twice, many of them also in the original sources. These are some examples that have been documented:</p> <ul style="list-style-type: none"> • The same event has been registered in more than one parish (i.e. in both the birth and the baptism parishes, or in both the death and burial parishes)³ • Two books, covering the same period and the same parish have been digitized and transcribed • Two different digitizations of the same book have both been transcribed <p>Follow the guidance for duplicates in the User Guide in these cases.</p>

³ In one case a marriage was found four times in the data: in the parish of the betrothal, in the parish of the marriage, and in the parish where the marrying vicar normally worked (twice in his book by mistake).

Search strategies

Refer to the User Guide for full details of search functions, shortcut keys and search terms.

PR Baptisms → census	PR Marriages → census	PR Burials → census
		<p>The information available to link buried people to the previous census varies greatly with chronology, geography, gender, civil status and age. The smallest amount of information is given about adult men in the early period and/or in bigger towns: often we have only name, age and burial parish. The fullest information we get after 1892 outside of the biggest towns: here we normally gain also birth date & place, parents' and spouse's names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend heavily on the specific characteristics of the linking unit and the individual deceased persons.</p>

1. Linking units

We continue to use [Trello](#) to track the linking production process. A [video](#) explaining in more detail how the board works can be found on Teams.

In parish record linking, we have introduced a new process for defining link units. We aim to link all records within set year ranges (e.g. 1860-1880) for a particular parish.

The PR data is organized in entire, transcribed books, and this is also how you'll find them displayed in ALA. The records will appear in the same order as in the original source.

PR Baptisms → census	PR Marriages → census	PR Burials → census
		<p>Your linking unit will be defined as the burials from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link burials from X parish to a given census (e.g. Junget 1860). You will then link records falling</p>

		<p>between 1 Feb of that census to 31 Jan of the next census (e.g. 1/2/1860-31/1/1880).</p> <p>NB:</p> <ul style="list-style-type: none"> You may need to link more than one 'book' to cover the whole link period. Most books contain both men and women: sometimes the genders are mixed, sometimes first we get all the men, then all the women. Both genders must be linked. See User Guide for full details of the new process for selecting and noting start and end records.
--	--	---

For PR records, always link by convenience:

1. Make a linking decision for every main record
2. Only link a non-main if they appear in source 2 (census) in the same household as the main
3. Do not link non-mains if the main is not linked

2. Primary variables

Primary variables in principle do not vary during life (or, in the case of the name or the marital status, if it changes, the change can be expected to follow some relatively well-defined rules).

A) Household/Family

PR Baptisms → census	PR Marriages → census	PR Burials → census
		<p>When the household information is available, it's usually in the form of the names of a spouse or of the parent(s).</p> <p>The name of a spouse – most often a deceased woman's husband – can be a good</p>

		<p>identifier when linking the deceased to the previous census. But remember that we cannot know from the transcribed PR data whether the spouse was alive or had died many years ago – or had been changed between the census and the death.</p> <p>Very often the names of the parents will not be of much help when linking a deceased person to the census immediately before the death, as we must expect many of the deceased individuals to have not been cohabiting with their parents for many years.</p>
--	--	--

B) Name

Generally, you should expect the name to be precise for a good link (apart from spelling variations). If the name does not fit, we are hesitant to link unless we have supporting information to confirm it.

Names in parish records

In parish records, names should generally be expected to be more “correct” than in censuses – especially in the later parish records.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Children’s names at baptisms have a special status, as they represent the person’s legal name, even though this concept of legality was definitely less clear for Danes in the first half of the 19th century than nowadays.</p> <p>Before a new law in 1828 – and sometimes even after – only the child’s first name was normally registered in the parish records. This means that you have to deduce the last name from the father’s name, and depending on local customs this could be both a derived or an inherited patronym – or a family name.</p>		<p>Especially for burials after 1892, you should expect to get the full “legal” name, including maiden name, for each deceased person.</p> <p>But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.</p>

With illegitimate children the last name may even come from the mother.		
---	--	--

Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information in how to interpret and understand birth places in the censuses, see the Appendix.

C) Birth place

Birth place is a very strong identification variable, especially when confirmed by name and age/birth date. If it is registered in the source, it is very important to use it in the linking process.

For information in how to interpret and understand birth places in the censuses, see the Appendix.

PR Baptisms → census	PR Marriages → census	PR Burials → census
		<p>After 1892 it was obligatory to register the birth place of the deceased (if available). This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR burials than in censuses, probably because the birth certificate was used when registering the deceased.</p> <p>Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.</p>

D) Birth year

The birth year in the ALA main screen is, depending on the PR event in question (baptism, marriage or burial), either directly taken from the original source or calculated from a combination of age and event year. In some cases, there is no birth year because of lacking information, but you might as linkers be able to derive it manually from the context of the data – maybe you can deduce a lacking death year for a deceased person from the death dates of the records above and below, and through that calculate a birth year. It is important to check also the original information in the “Full person info” screen if you lack information. It might be there, but in a not-machine readable form.

PR Baptisms → census	PR Marriages → census	PR Burials → census
		The birth year information in the main screen [*Birth year] is a derived field which combines available information from the original fields of

		<ul style="list-style-type: none"> • Birth year • Burial age/Death age • Death year/Burial year <p>If there is not enough information to calculate the birth year automatically, [*Birth year] will be empty. If that is the case, then look through the information in the “Full person info” screen in order to see if there is enough information for you to manually calculate a birth year for linking purposes.</p>
--	--	--

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

<p style="text-align: center;">General age searching recommendations</p> <ol style="list-style-type: none"> 1. Age 0-49: Use calculated birth year “+3” 2. Age 50-64: Use calculated birth year “+5” 3. Age 65-: Use calculated birth year “+10” 4. To search for any age, run a ‘blank age search’ using #
--

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range.

i) **Unborns**

PR Baptisms → census	PR Marriages → census	PR Burials → census
		<p>People who died before or were born after the year of the census you are linking to, are pre-marked as Unborn.</p> <p>This means that if you are linking to the 1860 census, a person who</p> <ul style="list-style-type: none"> • died in 1859 or • was born in 1861 <p>will be pre-marked as Unborn.</p>

E) *Civil status & gender*

These are considered to be primary variables. They are often used to exclude possible matches.

PR Baptisms → census	PR Marriages → census	PR Burials → census
----------------------	-----------------------	---------------------

		<p>Civil status has not been transcribed from the original sources. It can, though, be partially derived from indirect information:</p> <p>Before 1892</p> <ul style="list-style-type: none"> • If a spouse is registered, this means that the deceased was either married or widowed (but we don't know which) • If a deceased woman has no spouse registered, expect her to be unmarried, but don't be too sure (especially do not expect anything in big towns where data is known to be scarce) • Spouses are generally not registered for deceased men. This means that the absence of information tells us nothing about his marital status. <p>After 1892</p> <ul style="list-style-type: none"> • If a spouse is registered, this means that the deceased was either married or widowed (but we don't know which) • If no spouse is registered, this is a rather secure sign that the deceased was unmarried (it was obligatory to register the spouse). This, unfortunately, goes only for rural parishes and minor towns. In bigger towns, data is more scarce.
--	--	--

3. Secondary variables

All variables in ALA that are not listed above as primary variables are considered secondary information. In all sources we have access to extra fields containing information that might be relevant for finding someone and in certain cases to make linking decisions. There are not many of them in the PRs, though. For instance, unfortunately occupation wasn't transcribed from the original source.

The secondary variables available include:

PR Baptisms → census	PR Marriages → census	PR Burials → census
		Burial place contains information usually about the parish but sometimes the cemetery.
		Residence place and death place rarely contain different information to what is already displayed.

Important. Non-primary-variable fields should be used with caution. **Make your decision based on the primary variables first.** Only then should you consider secondary variables. The decision tree in the Appendix can help with this.

Important reminders

1. Match the primary variables (if available) first, then use the secondary variables to potentially help secure a link.
2. Do not use the secondary variables before the primary ones
3. Place of residence at the time of the census is a SECONDARY VARIABLE. Use +Secondary to mark a definite candidate

Conflict checking

Our conflict solver takes the final decision on contested links – those where both linkers disagree. All information provided in the comments section helps, but the conflict solver has helpfully outlined the following to be most useful:

1. If the place of birth is a place name in the parish we are looking for, please write that. It saves me a lot of time.
2. Have you used the reverse search to get to your result? If so it should be clearly stated.
3. How did you search for your result? If you did something special to find it, which you might think the other linker and the solver wouldn't think of, share it!
4. If you have a link and the other person says it's a multiple, and you decided that you still think your link is the best, then say in the comments why you still think it's a link.
5. If you cannot decide, please write that in the comments and why.
6. I am in love with the commenters that put the id of the other multiples in the comments and shall grant them my firstborn 😊 It saves me a ton of work, though it's, in a strict sense, not necessary because I will do a search in ALA myself anyway.

FAQs

- **I'm worried I am linking incorrectly or am making the wrong decision**
Do not be concerned if you are unsure of your linking decisions as we have built in checks at

many stages of our process. The Link Lives approach has adopted the principle of always having two linkers link each unit. Any 'contested' links (those the two linkers do not agree on) are then solved by a third linker (the conflict solver). Interestingly, when each pair of link files are compared, there are on average only 10-15% of links that are not identical (in census->census), which means that 85-90% of the time we are all linking in the same way. In whatever source we are linking, any conflicts are also usually halved after the conflict-checking stage, so we can be confident that this process helps to refine our 'uncontested' decisions.

- **Can I use other sources outside ALA to help me link?**
No, in order to train the algorithm you must never use sources outside ALA apart from those listed by name in 2.Primary Variables/C) Birthplaces.
- **Can't I just accept the potential link with the lowest score rather than searching each time?**
We always challenge the potential links to make sure we've been thorough. Ideally the algorithm is good enough to find each link, but then we wouldn't need human linkers! Eventually we test how good the algorithm is by comparing its links to those made by humans. Humans are good at spotting a subtle spelling error, or age difference, that a more rigid rule-based algorithm might miss.
- **I've found duplicate records. What should I do?**
Refer to the Duplicates section in the User Guide.
- **I've found a fun or interesting record**
Add it to our [Freak Links](#) file or share it in the chat!
- **What do I do with a blank record?**
Sometimes there is no information in a record. Simply mark it with the >1 button for us to deal with later. If you have only partial information (lack of birth year or last name, for instance), check the "Full person info" fields in order to see if the information is actually there, but in the wrong field or in a non-machine readable form.
- **My ALA has shut down with no warning. Why?**
Check if you have left one of your link (.csv) files open. If so, and if you tried to save some linking decisions right before ALA shut down, this is the explanation. Close your links file and restart ALA.

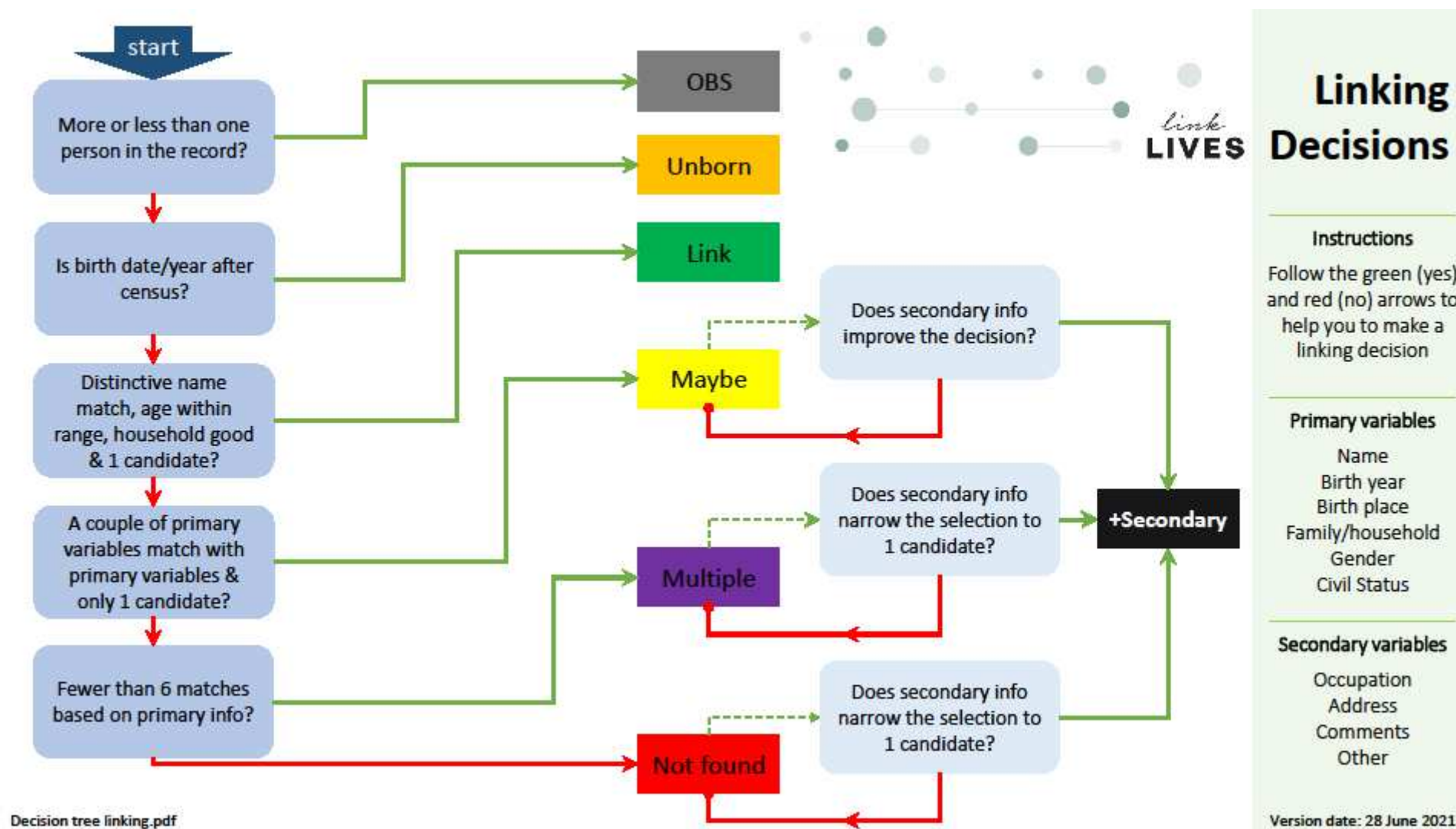
If it was not because of the links-file, please describe in as much detail as possible what you did just before ALA shut down – and try to reproduce the error. Please send a mail with your description to Asbjørn or Atlanta (at@sa.dk / olivia@hum.ku.dk).

- **What kind of comments should I write in my conflict checking file?**
See the section above called Conflict Checking for more guidance. But generally: the more guidance to your decision, the better for the conflict checker.

Appendix

1. Quick guide linking decision tree

This flow diagram is a summary of the descriptions above, and should act as a guide when linking. Follow the green (yes) and red (no) lines to help you to make a decision.



2. Purposes of linking decisions

This table shows the consequences of our linking decisions.

Decision	Purpose/function
Link	<ul style="list-style-type: none"> • Trains the machine that this is a secure link. • Data is used to train the algorithm. • Helps to build life courses. • The positive decisions (ie. maybe, link etc) are usually merged for Machine Learning purposes.
Maybe	<ul style="list-style-type: none"> • Offers the ML team evidence of a link that is missing something. They are then able to identify what that missing information is and use that as a dataset in itself. • The positive decisions are usually merged for ML purposes.
+Secondary can only be selected once you have clicked Maybe, Multiple, Not Found	<ul style="list-style-type: none"> • Because the 'accepted' approach to linking is using only unchanging variables, BRE would like to capture all instances where we use OTHER information in the source to make a link.
Multiple	<ul style="list-style-type: none"> • Trains the machine that there is more than one plausible candidate based on the information available. They then assess what data is missing that is preventing a "link". • Conflict solvers use this to help them investigate a finite potential pool of candidates.
Not found	<ul style="list-style-type: none"> • Trains the machine that there is no link for this candidate based on the information available. <p>Even though it feels like a negative decision, rest assured that there is a good chance that we'll find this person later on and link him/her based on adding information from other sources. You do not need to mark things as "multiple" or "maybe" to show you have made an effort 😊.</p>
Unborn	<ul style="list-style-type: none"> • These are excluded from our training data.
OBS	<ul style="list-style-type: none"> • We analyse this dataset later using more advanced techniques, so all we need to do is flag records for this purpose. • We link on a one-to-one basis, so if there is more or less than one person in a record it cannot be included in our standard training data.

A) Examples

Here are some examples of how to use the different linking decisions in different circumstances.

Source 1 (linking from)	Source 2 (linking to)	Reasoning	Decision
Maren Henriette Poulsen (wife of Rasmus Poulsen) b. 1850 in Græsted.	Maren Henriette Ferdinandsdatter, tjenestepige b. 1850 in Græsted.	No other Maren Henriettes in the parish or amt but with no household support there's a small chance there is more than one Maren Henriette of that age born in the parish.	Maybe
Hans Krøll's wife Christine Krøll, b 1820 Krønge.	Hans Krøll's wife Kirstine Jensdatter, b 1824 Krønge.	We cannot be sure whether he remarried, because we don't have her birth name and the age is slightly differing. We take "Maybe" to flag the doubt, but we're not REALLY worried.	Maybe
Laura Nielsen født Petersen in 1863 in Gilleleje.	There are two Laur* Pe*ers* in Gilleleje: Laure Jensine Kristine Pedersen and Laura Johanne Marie Pedersen, both 2 years out. Another Laura Petersen is the same age but born in the neighbouring parish.	There is a small set of identifiable candidates and we are relatively sure the right link is one of them, but we are missing household information and further middle name identifiers.	Multiple
Petrine Petersen, wife of Laurits Petersen, b. 1861 Gilleleje. The source is also Gilleleje, so it's likely she would be there 20 years before but with her maiden name.	There are 2 Petrine names registered in the parish within her age range and a further 3 Petronella variations.	A lack of information therefore stops us from deciding on one candidate, even though the name is quite unusual.	Not found
Peter Nielsen b 1853 Græsted.	11 people are returned on a search for name, age and Græsted as birth place, some with multiple given names, but no occupation or household support to distinguish between them.	A number of people are identifiable but a lack of further distinguishing information stops us from narrowing the pool to less than 6	Not found
Robert Anders Ludviggo Lund (distinctive name).	Cannot be found despite multiple versions of his name/age/birth place searches.	This is not lack of information but some anomaly in the transcription or source (e.g. he was out of the country, not enumerated, not transcribed, source illegible etc.)	Not found
Marie Hansen (married to Johan	We cannot know what her maiden name is, even though we have a definite place of	We lack information on her surname to even begin to be able to search for her.	Not found

Hansen) b. 1875 in Stege living in Sorø.	birth, because her name is too common.		
Peter Hansen b. 1875 in Årre, tjenestekarl.	There are 18 Peter Hansens in Ribe but none born in Årre. Several are born in neighbouring parishes but none in 1874-6	With no distinguishing characteristics (e.g. unusual name(s)), and no household support, it is not possible to reduce the pool of candidates.	Not found
Burials examples to follow.			

B) Census illustration

The screen grabs below show Johan Jensen's household in the urban area of Odense in 1850 and in 1845. Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	smedemester, huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn

3. Census to census details

A) Birth places in censuses

Generally, you should expect the birth place to be exact for a good link. Some systematic errors have been identified though:

- the further away, the less precise the birth place: a person not born in the parish will have a less specific place of birth (maybe just the amt is listed or even just the country, e.g. Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong probably increases with the distance.
- an oldest child was often born in another parish (before the parents settled down) but in the census is actually registered with the birth place where he/she actually grew up instead.
- parishes within the same “pastorat” are more often confused with each other than others.
- the administrative levels of birth places can vary from census to census (e.g. farm, village, parish, town, county, country)
 - Bear it in mind as a possibility that a birth place is not always recorded as a parish name.
 - A sparse use of contemporary geographic search engines is recommended. Not for every search, but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or if they were part of the same parish.
 - <https://www.danishfamilysearch.dk/>
 - <http://www.krabsen.dk/stednavnebase/>
 - Google maps
- Birth places are often recorded as variants of “her i sognet”, “samme sted” or “ditto”, i.e. text strings which do not include the actual name of the place.
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.
 - Look for the person in the standardized birth parish (Parish Std) in order to account for potential matches which appear simply with ‘ditto’ marks.

i) Special characteristics for birth places in URBAN areas

In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility, causing:

1. the imprecision mentioned above (because of distance) and
2. a tendency to register birth *places* rather than birth *parishes*, probably because of the census enumerator’s lack of knowledge of (and interest in) exact parish names in the surrounding rural areas.

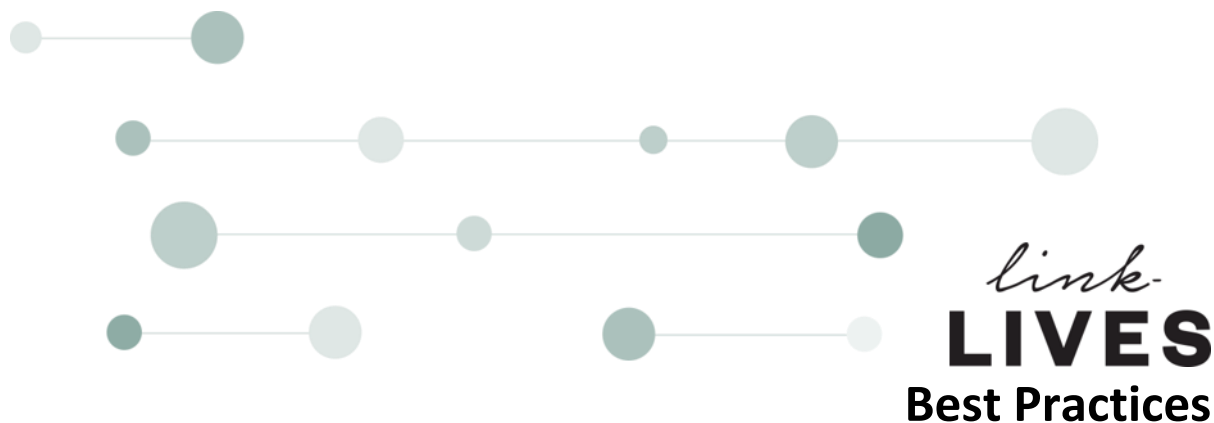
B) Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information in how to interpret and understand birth places in the censuses, see the Appendix.

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the

same person without having matching first names – and without any of the names being wrong!

- Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
- Watch out for common diminutives: Vilhelmine/Mine, Nielsine/Sine etc.
- First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names than men
 - married women taking the husband's last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen's children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription



Description	This document outlines the common approach to linking using ALA in the Link-Lives project. It contains universal guidance for all sources as well as specific considerations for individual sources (in tables).
Linking type	Training Data
Sources	<p>Parish record (PR) Baptisms 1813-1917 (nationwide) (specific guidance in blue)</p> <p>PR Marriages 1813-1917 (nationwide) (specific guidance in yellow)</p> <p>PR Burials 1813-1917 (nationwide) (specific guidance in green)</p> <p><i>You will find the following sources in 04 Best Practices_final.pdf:</i></p> <p><i>KBH Burials 1861-1911</i></p> <p><i>Censuses 1845, 1850, 1860, 1880, 1885 (KBH only), 1901</i></p>
Definitions	Source1 is the census we link from, Source2 is the census we link to.
Authors	OR, AT
ALA release	ALA v.1.4.0 29 June 2021

What's new?

1. Two person search function added
2. Linking by convenience approach
3. New OBS button definitions to clarify its use
4. Linking forwards for PR Baptisms & PR Marriages
5. Potential links temporarily not available (June 2021)
6. Detailed descriptions of sources

Contents

From source to ALA.....	3
1. Source types.....	3
A) PR Baptisms 1813-1917	4
B) PR Marriages 1813-1917.....	5
C) PR Burials 1813-1917	7
2. Cleaning & preparing	9
3. Generating potential links.....	9
Principles of linking	9
1. Linking process.....	9
2. Linking decisions	10
3. Linking decisions chart	11
Search strategies.....	14
1. Linking units	14
2. Primary variables	15
A) Household/Family	15
B) Name.....	16
C) Birth place	17
D) Birth year.....	17
E) Civil status & gender	19
3. Secondary variables	19
Conflict checking.....	21
FAQs	21
Appendix	22
1. Quick guide linking decision tree	22
2. Purposes of linking decisions	24
A) Examples	25
B) Census illustration.....	26
3. Census to census details	27
A) Birth places in censuses	27
B) Names in censuses	27

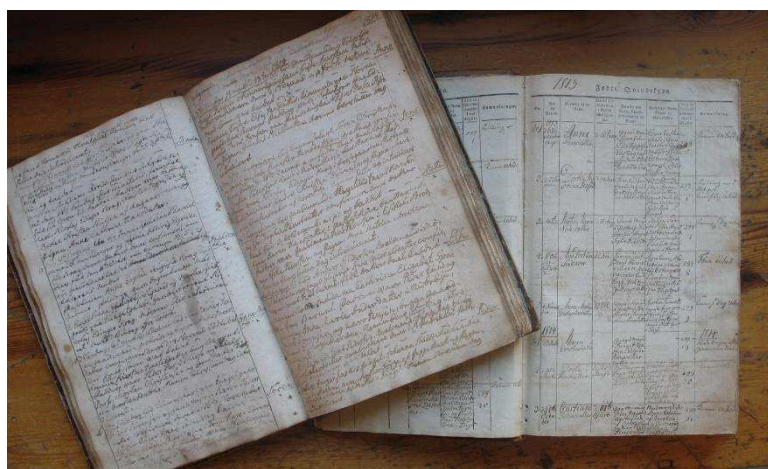
From source to ALA

This section offers some background on the sources, their genesis and how they are presented in ALA.

1. Source types

Parish Records are continuous, dynamic registers containing information on vital events of the parishioners. The core vital events registered are Baptisms, Marriages and Burials – for some periods supplemented by Confirmations, Arrivals and Departures. In Link-Lives generally, and in the ALA-linking specifically, we will concentrate on linking only the core vital events of Baptisms, Marriages and Burials – and this is why this Best Practices document only focuses on them.

Parish records were to be made by the vicar of every parish from ca. 1645-46, but many of the oldest ones have not survived. From most parishes we have parish records from the 18th century onwards, but from ca. 1813 they have been preserved from all parishes. From this year it was decided that parish records had to be kept in two versions, and it was forbidden to have the two versions under the same roof at night, to avoid potential destruction by fire by saving a duplicate. The two versions would generally be written by the vicar (præsten) and the parish clerk (deggen). Another major change in the 1813 reform was that parish records now had to be written in pre-printed books with defined columns for the single pieces of information; in earlier times, the parish records had been written just as the priest wanted, i.e. often in a much more narrative form.



The parish record books of Junget before and after the 1813 reform which introduced the pre-printed, schematized books.

In archival language, the parish records versions have the following names:

- Before 1813: Enesteministerialbog (the only version)
- After 1813: Hovedministerialbog (the vicar's version)
- After 1813: Kontraministerialbog (the parish clerk's version)

For the time after 1813, it will normally be the kontraministerialbog that has been digitized and transcribed. The two versions should be completely identical (the two writers had to “konferere” their books regularly), but occasionally of course there will be differences. This will of course generally have no importance for us, working with the transcriptions and linking on the basis only of those.

Født					Kvindekøn.		
No.	År og Dato.	Barnets fulde Navn.	Dødsdato enten i Kirke eller Hjemmet.	Barnets Navn, Sted, Døbskirke og Bopæl.	Barnets Navn, Sted og Døbskirke.	Overført i det sin. Døds. Reg.	Bemærkninger.
1863	27-6-1822	Niels Nielsen Christensen	27-6-1823	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		
1863	15-3-1863	Mette Katrine Elisabeth Jensen Svane	15-3-1863	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		
1863	17-6-1863	Kristine Iversen	19-7-1863	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds		uægte barn

Født					Kvindekøn.		
No.	År og Dato.	Barnets fulde Navn.	Dødsdato enten i Kirke eller Hjemmet.	Barnets Navn, Sted, Døbskirke og Bopæl.	Barnets Navn, Sted og Døbskirke.	Overført i det sin. Døds. Reg.	Bemærkninger.
1	1900	Mette Katrine Elisabeth Jensen Svane	15-3-1863	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		
2	1900	Kristine Iversen	19-7-1863	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds		uægte barn

Examples of information in the original source (from the above illustrations):

No.	Birth date	Bap. date	Name	Parents	Comments
3	27-6-1822	27-6-1823	Niels Nielsen	Gaard Mand Niels Christensen Bak og Hustrue Maren Nielsdatter af Junget	
1	28-2-1863	15-3-1863	Mette Katrine Elisabeth Jensen Svane	Gårdejer og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	
2	17-6-1863	19-7-1863	Kristine Iversen	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	uægte barn
1	14-5-1900	31-5-1900	Marie Andersen Møjbæk Møjbæk tilføjet i H. t. Bevilling af 28/7-05	Gmd. Niels Andersen (Møjbæk) og Hustru Hansine Nielsen af Junget 37 Aar	

B) PR Marriages 1813-1917

We link PR Marriages *forward* to the nearest following census. This strategy has been chosen in order to maximize the use of the identification potential of the household information inherent in

the marriage record: we can look for a couple instead of two single persons. This approach is particularly important, as the Marriage records for a large part of the period only contain information on name, age and marriage parish – the household support thus dramatically increases the possibility for reliable linking.

As the examples from the original source below show, there is a gradually increasing amount of data available as time goes on. In the beginning, we generally have the following information to link on:

- Name
- Age
- Marriage parish
- Spouse's name & age (household info)

Gradually, the following will join the ones mentioned above:

- Birth place
- Birth day
- Parents' names (all)
- Same data for the spouse (household info)

After 1892 all this information will be generally available (except maybe in the largest cities) while before 1892 often only the first batch of information will be registered. Fewer than 30% of the records contain birth places before 1892, for example, but more than 80% from 1892 onwards.

166

Copulerede.

No.	Brudgommens Navn, Alder, Bøndering og Cykelsteds.	Brudens Navn, Alder, Bøndering og Cykelsteds.	Doct.	Væstet Dagen.	Om registreret siden 1801 hjemme og i hvis Tilfælde Brudgomens Datum.	Har an- ført i det statede Jensf. Regist.	Bemærking.
				1821			
1	Aug. Rasmussen	Aug. Rasmussen	30 år	30 år			
				1825			
1	Aug. Rasmussen	Aug. Rasmussen	30 år	30 år			

Married couples from Junget parish 1821-23, 1861-62, and 1900. The forms are more or less identical between the 1820s and the 1860s, but a bit more information has been entered in the latter records: for identification purposes the more exact registration of age might be helpful, and the same goes for the vaccination information – which has not been transcribed though. The 1900 record contains information on birth date & place and both parents' names – data which has all been transcribed. According to the post-1892 form this data was supposed to be entered according to the spouses' official birth certificates.

67

Copulerede.

No.	Brudgommens Navn, Alder, Bøndering og Cykelsteds.	Brudens Navn, Alder, Bøndering og Cykelsteds.	Doct.	Væstet Dagen.	Om registreret siden 1801 hjemme og i hvis Tilfælde Brudgomens Datum.	Har an- ført i det statede Jensf. Regist.	Bemærking.
1	Aug. Rasmussen	Aug. Rasmussen	30 år	30 år			
2	Aug. Rasmussen	Aug. Rasmussen	30 år	30 år			

Ægte-		viede.		199	
<p>Bredens fulde Navn efter forretnings Duallitet eller ved Eftersyn i Kirkebogen, hendes Fødselsaar og -dag, Stilling og Opholdssted samt hendes Forældres fulde Navn, Stilling og Bopæl.</p> <p>(Bredens betegnelse som Udgift, Bæmsmand eller fraktilt; i de to sidste Tilfælde andres fulde Navn, Stilling og Bopæl samt respektive Datoer for hendes Død og for den Døds eller Bevilgning, ved hvilken Ægteskabet er ophævet og Tilfældet til nyt Ægteskab er meddelet.)</p>	<p>Bredens fulde Navn efter forretnings Duallitet eller ved Eftersyn i Kirkebogen, hendes Fødselsaar og -dag, Stilling og Opholdssted samt hendes Forældres fulde Navn, Stilling og Bopæl.</p> <p>(Bredens betegnelse som Udgift, Bæmsmand eller fraktilt; i de to sidste Tilfælde andres fulde Navn, Stilling og Bopæl samt respektive Datoer for hendes Død og for den Døds eller Bevilgning, ved hvilken Ægteskabet er ophævet og Tilfældet til nyt Ægteskab er meddelet.)</p>	<p>a) Navn blev Tilflytning bestemt?</p> <p>b) Navn fandt Tilflytning Sted?</p> <p>c) Hvis flytning af har fundet Sted, angives Hjemmet (Sognet), hvor den foregik.</p>	<p>Vedkommende Aar og Måned samt Sted (Sognet), hvor den foregik.</p>	<p>a) Af hvilken Præst-Vedkommende er foretaget?</p> <p>b) For saa vidt en fremmed Præst har foretaget Vedkommende, angives Hjemmet (Sognet), hvor den foregik.</p> <p>c) Om Vedkommende er forrettet i Kirken eller i Hjemmet.</p> <p>(I sidste Tilfælde angives Bevilgningens Dato.)</p>	<p>Anmærkninger.</p> <p>(Her andres eventuelt Eftersyn om, i hvilken Tilfælde Vedkommende skulde opføres, hvilken Eftersyn angives her i Bogen.)</p>
<p>2. Peder Pedersen Nautrup, Ungkarl af Krejberg, født i Nautrup, den 3^{de} December 1872; Søn af Gmd. Peder Pedersen Nautrup og Hustru Johanne Kristensen af Knud By, Rødding Sogn.</p>	<p>Kirstine Petersen, Pige af Skove, født i Skove, den 25^{de} September 1881; Datter af Gmd. Niels Hans Peter Petersen og Hustru Johanne Jensen af Skove.</p>	<p>Gmd. Niels Hans a. Skove, 1900</p> <p>Peder Pedersen af Skove, 6^{de} April i Skove</p> <p>Gmd. Niels Hans af Skove, 6^{de} April i Skove</p>	<p>1900</p> <p>6^{de} April i Skove</p>	<p>af Præst Nautrup</p> <p>af Præst Nautrup</p>	

Examples of information in the original source (from the above illustrations):

No.	Groom	Bride	Marriage date and place
1	Ung Karl Jens Sørensen af Væjsmark, gammel 26 Aar, Vaccineret	Ung Pige Anne Katrine Jensdatter af Brøholm, gammel 30 Aar, Vaccineret	24-2-1823 i Kirken
4	Ungkarl Rasmus Jensen af Junget, 28½ Aar, vaccineret 26de Maj 1836 af Tetens	Pigen Bodil Kirstine Villadsdatter af Junget 35 Aar gl.; vaccineret 1827 af Stengel	2-1-1862 i Kirken
2	Peder Pedersen Nautrup, Ungkarl af Krejberg; født i Nautrup, den 3die December 1872; Søn af Gmd. Peder Pedersen Nautrup og Hustru Johanne Kristensen af Knud By, Rødding Sogn.	Kirstine Petersen, Pige af Skove, født i Skove Junget Sogn den 25de September 1881. Datter af Gmd. Niels Hans Peter Petersen og Hustru Johanne Jensen af Skove.	6-4-1900 i Junget

C) PR Burials 1813-1917

The burial records are generally the vital events from the PRs with least information.

Often there is information on the deceased person alone, whereas Baptisms and Marriages always include more than one associated person (e.g. mother, spouse). This characteristic of the Burials applies especially to adult males: children will often have information about their father/parents, and women will often have information on their husbands (if married/widowed). Adult males, unfortunately, for a large part of the period, only have information on name, age and residence parish – marital status, occupation and address have not been transcribed.

As the examples from the original source below show, there is a gradually increasing amount of data available as time goes on. In the beginning, we generally have the following information to link on:

- Name
- Age
- Burial parish
- Husband's name (if a married/widowed woman)
- Father's/parents' names (if a child)

Gradually, the following will join the ones mentioned above:

- Birth place

- Birth day
- Wife's name (if a married/widowed man)
- Parents' names (all)

After 1892 all this information will be generally available (except maybe in the largest cities) while before 1892 often only the first batch of information will be registered. Fewer than 30% of the records contain birth places before 1892, for example, but more than 80% from 1892 onwards.

40

Døde. Mandkøn.

No.	Dødsdags- pr.	Begravelses- stedet.	Den Dødes Fø- dselsnavn.	Stand, Haandtering og Opholdssted.	Alder. almindl. Tross. Konger.	Antal af fødsels- år.	Antal af fødsels- år.
1	17. 12. 1823	Sanr.	Basmur Christen	gæst og hustru i Junget	44	305	1831
3	19. 10. 1823	Sanr.	Anders Christen	gæst og hustru i Junget	25	305	1832
4	19. 10. 1823	Sanr.	Anders Christen	gæst og hustru i Junget	23	305	1833

Buried men from Junget parish 1823, 1863, and 1900-1901. There is no difference between the registration in 1823 or 1863, even if the registration form now says that the deceased's birth place and father's name ought to be registered – if obtainable. But the registration of 1900 contains information on birth place, parents' names and spouse's name – data which has all been transcribed.

144

Døde. Mandkøn.

No.	Dødsdagen.	Begravelsesdagen.	Den Dødes Fødselsnavn.	Den Dødes Stand, Haandtering og Opholdssted samt hans Fæderes eller, hvis han er nægtet, hans Moderes Navn, forsaavidt disse ikke afsløres kan revideres.	Alder.	Antal af fødsels- år.	Antal af fødsels- år.
1	18. 10. 1863	18. 10. 1863	Anders Christen	gæst og hustru i Junget	44	305	1831
2	19. 10. 1863	19. 10. 1863	Anders Christen	gæst og hustru i Junget	25	305	1832
3	19. 10. 1863	19. 10. 1863	Anders Christen	gæst og hustru i Junget	23	305	1833

Døde. Mandkøn.

223

No.	Dødsdagen (År og Dato) og Dødsstedet (Paa Landet: Hjem, Sagst og Herredet. I Købstad: Købstaden og tillige Sagst, hvis der i Hjem er mere end 1 Sagst).	Begravelsesstedet (Kirkens navn, Sagst og Herredet (Købstaden) Navn) og Begravelsesdagen (År og Dato)	Den Afdødes fulde Navn (saa vidt muligt eller forrest Døds- eller Fødselsnavn eller ved Efterfølgende i Kirke- bogen).	Den Afdødes Silling og Bøgel (s): sidste faste Opholdssted, hans eller hendes fæderes eller, hvis han er nægtet, hans Moderes Navn, forsaavidt disse ikke afsløres kan revideres. Hvis den Afdøde var eller vilde at have været gift, an- føres tillige dette samt, saa vidt muligt, Afdødes fulde Navn, Silling og Bøgel.	Alder. (Når den Afdøde var over 1 Aar, angives Alderen med det fulde År.)	Af hvilken Præst Begravelsen er foretaget.	Antagelser.
2	21 August 1900 Geddesvej, Sagst, Junget Købstad	21 August 1900 Købstad	Anders Christen (Hustru)	gæst og hustru i Junget	90 Aar	Anders Christen	
1	21. 10. 1863 Geddesvej, Sagst, Junget Købstad	21. 10. 1863 Købstad	Anders Christen (Hustru)	gæst og hustru i Junget	47 Aar	Anders Christen	

Examples of information in the original source (from the above illustrations):

No.	Death date	Bur. date	Name	Occupation and family	Age
2	17-1-1823	26-1-1823	Rasmus Erichsen	Huus og Muurmester i Junget	44 Aar
5	5-4-1863	12-4-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75 År
1	21-3-1901 Junget by	28-3-1901	Hans Kristian Jensen	Indsidder i Junget. Født i Gassum Nørhald Herred; Søn af Hmd. Hans Kr. Jensen og Hustru Abelone Kristensen af Gassum. Gift med Ane Kirstine Svenningsen. Sidste fælles Bopæl Junget.	47 Aar

2. Cleaning & preparing

The data has been prepared in a way that makes sense for both Link-Lives and ALA linkers. The PR data was transcribed in Asia by a company paid by Ancestry, and we must expect that a set of instructions was given to the transcribers directing them on which field to record each piece of information in. As with any historical source, however, be aware that errors, omissions and subjective decisions are inevitable at each stage. For example, for almost all fields we find misplaced or misread information from other fields.

3. Generating potential links

Potential links are generated for each record entry (where possible) by running a rule-based algorithm based on name and age. The most likely 10 are then displayed on the right side of the ALA screen in order of probability score (0 is the closest match). These act as start-point suggestions for linkers. We might find the correct link within this list, or we might not, but we always attempt to challenge them.

Principles of linking

We link according to an agreed set of guidelines that help our team link consistently. The focus is on linking the right ones: we want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links. The process is as follows:

1. Linking process

1. Linking
 - a. Each linking unit is always linked by at least two linkers.
2. Conflict checking
 - a. After each finishes their linking unit, our linking coordinator identifies all the contested links (i.e. those which the linkers disagreed on) and creates a conflict file in Excel for each linker to check.
 - b. Each linker works through the conflict file to check and comment on each contested link.
3. Conflict solving
 - a. The coordinator solves any remaining contested links
 - b. A consolidated training data file is generated and stored.

Golden Linking Rules

1. Only link one person to one person
2. Always challenge a potential link
3. Always search for a competing candidate
4. Always link using primary variables first, before assessing secondary variables
5. Never use unauthorized sources outside of ALA
6. When in doubt, do not link

2. Linking decisions

We link using two sets of variables: primary variables in principle do not vary during life (or, in the case of the name and marital status, if it changes, the change can be expected to follow some relatively well-defined rules). Secondary variables may contain information that gives context to the individual's life, but which generally changes – or can change unpredictably – over time. Link Lives uses the following fields to distinguish primary from secondary:

Primary variables	Secondary variables
Name	Occupation
Birth year	Address
Birth place	Comments
Family/household	Other
Gender	
Marital Status	

In the parish record transcriptions we get very few secondary variables, even when they are in the original source (occupation, for instance, is not transcribed). This means that the following (general) description of how and when to use the primary and secondary variables in searching is less relevant in the PR context than when linking other sources, but we leave them here, as they DO apply at least to the different geographically defined fields (i.e. address possibilities).

- Identify a candidate by searching on primary variables first.
- It may then be relevant to remove one of the criteria, expand the years or remove them entirely, in order to exhaust alternatives.
- Only then should you consider the secondary variables.

The following chart shows the different linking decisions, when to make them and how they are used.² See the Appendix for a flowchart diagram to assist in your linking, a fuller list of examples and an overview of the consequences of our linking decisions.

² When the linking decisions are processed, Maybes, Links and those with +Secondary are marked 'positive' (shaded green below) and the rest are marked 'negative' (shaded red below). This distinction simply helps to train the machine algorithm.

3. Linking decisions chart

Decision	Definition	How it is commonly used
Link	<p>When you are confident that you have found the correct link.</p> <p><i>Quantitative hint: You should be at least 99% sure.</i></p>	<p>We are usually most sure of links when primary variables all contain satisfying information for a secure link to be established. E.g. there are specific characteristics of name, age, and (for censuses) household and/or birth place.</p> <p>This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.</p>
Maybe	<p>When you are almost sure you have found the correct link, but you want to flag that there is some information missing.</p> <p><i>Quantitative hint: You should be at least 95% sure.</i></p>	<p>We tend to link “maybe” when one of the characteristics is missing but the others are solid. E.g. distinctive name, wrong age.</p> <p>Candidates often include wives who take their husbands’ names, but whose other characteristics are otherwise a match.</p>
<p>+Secondary³ can only be selected once you have clicked Maybe, Multiple, Not Found</p>	<p>When you use information secondary variables to move from a “Maybe”, “Multiple” or “Not Found” to a single plausible candidate.</p> <p>Maybe +Secondary One option is made more certain by using secondary information.</p> <p>Multiple +Secondary 2-5 options narrowed to one using secondary information.</p> <p>Not Found +Secondary > 5 options narrowed to one using secondary information.</p> <p><i>Quantitative hint: After using the secondary information, you should be at least 95% sure.</i></p>	<p>Candidates in census-to-census linking often include people with distinctive professions or trades involving some training, e.g. lærer, or occupations that are out of place in the local context.</p> <p>In parish records we might use geographical indicators of residence to narrow a selection to one candidate.</p>

³ For linkers who have used ALA versions before, please note there is no longer an ‘Occupation’ decision button.

Decision	Definition	How it is commonly used
Multiple	When you have a well-defined group of no more than 5 plausible candidates, but you cannot choose between them.	e.g. 3 x Maren Madsens from the same parish of the same age and no household support to distinguish between them. e.g. Peter Jens Knudsen died in 1880, and there are 3 Peter Knudsen candidates and 2 Jens Knudsen candidates all in the same age range.
Not found	<p>Not found can be applied to several situations. You do not have to mark which is which. The following is just for explanatory purposes:</p> <p>1) Unfound When you find no plausible candidate, even with the presence of distinctive characteristics.</p> <p>2) Too many When there are too many candidates (>5) to choose from and you lack primary information to narrow the pool (though see +Secondary above).</p> <p>3) Unfindable Where there is missing information that makes it impossible to identify any candidate (though see +Secondary above).</p> <p>4) Presumed dead When, linking forward, you expect this person to be dead (e.g. if linking a PR marriage to the following census, and you find the wife as a widow)</p>	<p>Examples include:</p> <p>1) Distinctive characteristics unfound e.g. Robert Anders Ludviggo Lund but despite multiple search strategies, you did not find him.</p> <p>2) Lots of local possibilities e.g. one Maren Madsen from the right parish, but her age is 4 years out, then half a dozen Maren Madsens born in neighbouring parishes with correct ages.</p> <p>3a) Common given names e.g. Marie Sørensen, wife of Hans Sørensen, no maiden name.</p> <p>3b) Vague places of birth e.g. Hans Hansen, b. 1875, Slesvig.</p> <p>3c) Missing transcription information e.g. Ubeboet, Mangler, Jens B[?], no place of birth, no household support.</p>

Decision	Definition	How it is commonly used
Unborn	If it is probable that the person was not yet born at the time of the census.	Used for people whose birth years are calculated to within a year of a census year and who are missing at the end of a family grouping. If no household support, simply mark Not Found.
OBS	<p>When you find records which you simply cannot link and would like to flag them for later analysis:</p> <ol style="list-style-type: none"> 1) more than one person (e.g. two people listed in one record) 2) non-persons (e.g. empty addresses or fields) 3) duplicates (e.g. the same event registered or transcribed twice) 	<p>In the CPH Burials there are several hundred records which contain more than one deceased person, especially mothers and infants or stillborns but also two children, twins etc. We do NOT, though, expect this to be the case with the PR Burials.</p> <p>Some records identify empty addresses, or the hand-writing in the source was illegible to the transcriber.</p> <p>There are multiple reasons for the presence of duplicates in the PR data. Of course in the perfect world every individual event would exist only one time in the PR dataset, but for various reasons a percentage of them were registered twice, many of them also in the original sources. These are some examples that have been documented:</p> <ul style="list-style-type: none"> • The same event has been registered in more than one parish (i.e. in both the birth and the baptism parishes, or in both the death and burial parishes)⁴ • Two books, covering the same period and the same parish have been digitized and transcribed • Two different digitizations of the same book have both been transcribed <p>Follow the guidance for duplicates in the User Guide in these cases.</p>

⁴ In one case a marriage was found four times in the data: in the parish of the betrothal, in the parish of the marriage, and in the parish where the marrying vicar normally worked (twice in his book by mistake).

Search strategies

Refer to the User Guide for full details of search functions, shortcut keys and search terms.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The information available to link baptisms is more or less the same throughout the period in question.</p> <p>The most important linking variables are name and age of the child combined with the household information inherent in the baptism: we know who the child's parents are.</p>	<p>The information available to link married people to the subsequent census varies with chronology and geography. But generally it's of huge importance to take the household information inherent in the marriage record into account when linking.</p> <p>The fullest individual information we get after 1892 is outside of the biggest towns: here we normally gain also birth date & place, and parents' names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend on the specific characteristics of the linking unit.</p>	<p>The information available to link buried people to the previous census varies greatly with chronology, geography, gender, civil status and age. The smallest amount of information is given about adult men in the early period and/or in bigger towns: often we have only name, age and burial parish. The fullest information we get after 1892 outside of the biggest towns: here we normally gain also birth date & place, parents' and spouse's names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend heavily on the specific characteristics of the linking unit and the individual deceased persons.</p>

1. Linking units

We continue to use [Trello](#) to track the linking production process. A [video](#) explaining in more detail how the board works can be found on Teams.

In parish record linking, we have introduced a new process for defining link units. We aim to link all records within set year ranges (e.g. 1860-1880) for a particular parish.

The PR data is organized in entire, transcribed books, and this is also how you'll find them displayed in ALA. The records will appear in the same order as in the original source.

PR Baptisms → census	PR Marriages → census	PR Burials → census
Your linking unit will be defined as the baptisms from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link baptisms from X parish to a given census (e.g. Junget 1860). You will then link	Your linking unit will be defined as the marriages from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link marriages from X parish to a given census (e.g. Junget 1860). You will then link	Your linking unit will be defined as the burials from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link burials from X parish to a given census (e.g. Junget 1860). You will then link records falling

records falling between 1 Feb of the year of the previous census to 31 Jan of the year of the census in question (e.g. 1/2/1850-31/1/1860).

NB:

- You may need to link more than one 'book' to cover the whole link period.
- Most books contain both boys and girls: sometimes the genders are mixed, sometimes first we get all the boys, then all the girls. Both genders must be linked.

See User Guide for full details of the new process for selecting and noting start and end records.

records falling between 1 Feb of the year of the previous census to 31 Jan of the year of the census in question (e.g. 1/2/1850-31/1/1860).

NB:

- You may need to link more than one 'book' to cover the whole link period.
- **See User Guide for full details of the new process** for selecting and noting start and end records.

between 1 Feb of that census to 31 Jan of the next census (e.g. 1/2/1860-31/1/1880).

NB:

- You may need to link more than one 'book' to cover the whole link period.
- Most books contain both men and women: sometimes the genders are mixed, sometimes first we get all the men, then all the women. Both genders must be linked.
- **See User Guide for full details of the new process** for selecting and noting start and end records.

For PR records, always link by convenience:

1. Make a linking decision for every main record
2. Only link a non-main if they appear in source 2 (census) in the same household as the main
3. Do not link non-mains if the main is not linked

2. Primary variables

Primary variables in principle do not vary during life (or, in the case of the name or the marital status, if it changes, the change can be expected to follow some relatively well-defined rules).

A) Household/Family

PR Baptisms → census	PR Marriages → census	PR Burials → census
Generally, it's of huge importance to take the household information inherent in the baptism record into account when linking: we know who the child's parents are, and the child will – depending on the year range in	Generally, it's of huge importance to take the household information inherent in the marriage record into account when linking.	When the household information is available, it's usually in the form of the names of a spouse or of the parent(s).
		The name of a spouse – most often a deceased woman's

<p>question – most often live with the parents in the next census.</p> <p>Be aware though, that the child might be illegitimate, and that you can therefore not 100% expect the parents to be married at the time of the census: it is possible to be unmarried and have children – and we cannot distinguish between legitimate and illegitimate children from the transcribed data. The child might also live in an institutions or with foster parents.</p>	<p>Be aware though, of the possibility that</p> <ul style="list-style-type: none"> • one/both of the spouses died before the census • the couple was separated/divorced before the census • the couple did not live together at the time of the census (but were still listed with “married” as marital status) 	<p>husband – can be a good identifier when linking the deceased to the previous census. But remember that we cannot know from the transcribed PR data whether the spouse was alive or had died many years ago – or had been changed between the census and the death.</p> <p>Very often the names of the parents will not be of much help when linking a deceased person to the census immediately before the death, as we must expect many of the deceased individuals to have not been cohabiting with their parents for many years.</p>
--	--	--

B) Name

Generally, you should expect the name to be precise for a good link (apart from spelling variations). If the name does not fit, we are hesitant to link unless we have supporting information to confirm it.

Names in parish records

In parish records, names should generally be expected to be more “correct” than in censuses – especially in the later parish records.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Children’s names at baptisms have a special status, as they represent the person’s legal name, even though this concept of legality was definitely less clear for Danes in the first half of the 19th century than nowadays.</p> <p>Before a new law in 1828 – and sometimes even after – it was often only the child’s first name that was registered in the parish records. In those cases you have to deduce the last name from the father’s name, and depending on local customs this could be both a derived or an inherited patronym – or a family name.</p>	<p>Especially for marriages after 1892, you should expect to get the full “legal” name, including maiden name, for each of the spouses (it says in the registration form that the record should be based on the birth certificate).</p> <p>But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.</p>	<p>Especially for burials after 1892, you should expect to get the full “legal” name, including maiden name, for each deceased person.</p> <p>But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.</p>

With illegitimate children the last name may even come from the mother.

Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information in how to interpret and understand birth places in the censuses, see the Appendix.

C) Birth place

Birth place is a very strong identification variable, especially when confirmed by name and age/birth date. If it is registered in the source, it is very important to use it in the linking process.

For information in how to interpret and understand birth places in the censuses, see the Appendix.

PR Baptisms → census	PR Marriages → census	PR Burials → census
The child's birth place is generally the parish where the baptism was recorded – though sometimes there is a birth parish and a baptism parish.	After 1892 it was obligatory to register the birth place of the spouses – based on the information from their birth certificates. This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR Marriages than in censuses, probably because the birth certificate was actually used.	After 1892 it was obligatory to register the birth place of the deceased (if available). This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR burials than in censuses, probably because the birth certificate was used when registering the deceased.
The birth parish of the parents is never recorded.	Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.	Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.

D) Birth year

The birth year in the ALA main screen is, depending on the PR event in question (baptism, marriage or burial), either directly taken from the original source or calculated from a combination of age and event year. In some cases, there is no birth year because of lacking information, but you might as linkers be able to derive it manually from the context of the data – maybe you can deduce a lacking death year for a deceased person from the death dates of the records above and below, and through that calculate a birth year. It is important to check also the original information in the “Full person info” screen if you lack information. It might be there, but in a not-machine readable form.

PR Baptisms → census	PR Marriages → census	PR Burials → census
The birth year of the child is, except for a few odd cases, equal to the baptism year.	The birth year information in the main screen [*Birth year] is a derived field which combines	The birth year information in the main screen [*Birth year] is a derived field which combines

<p>The newer the records are, the more often they will contain an age for the mother (rarely the father). This will give a calculated birth year for the mother.</p>	<p>available information from the original fields of</p> <ul style="list-style-type: none"> • Birth year • Marriage age • Marriage year <p>If there is not enough information to calculate the birth year automatically, [*Birth year] will be empty. If that is the case, then look through the information in the "Full person info" screen in order to see if there is enough information for you to manually calculate a birth year for linking purposes.</p>	<p>available information from the original fields of</p> <ul style="list-style-type: none"> • Birth year • Burial age/Death age • Death year/Burial year <p>If there is not enough information to calculate the birth year automatically, [*Birth year] will be empty. If that is the case, then look through the information in the "Full person info" screen in order to see if there is enough information for you to manually calculate a birth year for linking purposes.</p>
--	--	---

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

General age searching recommendations

1. Age 0-49: Use calculated birth year "+3"
2. Age 50-64: Use calculated birth year "+5"
3. Age 65-: Use calculated birth year "+10"
4. To search for any age, run a 'blank age search' using #

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range.

i) **Unborns**

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>This is not a relevant possibility, as we are linking forwards.</p>	<p>This is not a relevant possibility, as we are linking forwards.</p>	<p>People who died before or were born after the year of the census you are linking to, are pre-marked as Unborn.</p> <p>This means that if you are linking to the 1860 census, a person who</p> <ul style="list-style-type: none"> • died in 1859 or • was born in 1861 <p>will be pre-marked as Unborn.</p>

E) Civil status & gender

These are considered to be primary variables. They are often used to exclude possible matches.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The child will only very rarely have a marital status other than “unmarried” in the next census.</p> <p>Be aware that you cannot expect the parents to be necessarily married, together or the same in the following census: the child might be born outside wedlock, fostered etc.</p>	<p>The spouses’ marital statuses haven’t been transcribed from the original sources, but as long as we’re linking forwards, this doesn’t influence our linking possibilities.</p> <p>When linking forward to the following census you’d generally need very strong evidence to support linking from a marriage to an unmarried person. All other marital statuses – married, widowed, separated/divorced – are possible matches.</p> <p>Marriage records are probably those with the least gender confusion in the data.</p>	<p>Civil status has not been transcribed from the original sources. It can, though, be partially derived from indirect information:</p> <p>Before 1892</p> <ul style="list-style-type: none"> • If a spouse is registered, this means that the deceased was either married or widowed (but we don’t know which) • If a deceased woman has no spouse registered, expect her to be unmarried, but don’t be too sure (especially do not expect anything in big towns where data is known to be scarce) • Spouses are generally not registered for deceased men. This means that the absence of information tells us nothing about his marital status. <p>After 1892</p> <ul style="list-style-type: none"> • If a spouse is registered, this means that the deceased was either married or widowed (but we don’t know which) • If no spouse is registered, this is a rather secure sign that the deceased was unmarried (it was obligatory to register the spouse). This, unfortunately, goes only for rural parishes and minor towns. In bigger towns, data is more scarce.

3. Secondary variables

All variables in ALA that are not listed above as primary variables are considered secondary information. We use the +Secondary button in ALA when we use anything other than primary variables to make our linking decision.

In all sources we have access to extra fields containing information that might be relevant for finding someone and in certain cases to make linking decisions. There are not many of them in the PRs, though. For instance, unfortunately occupation wasn't transcribed from the original source.

The secondary variables available include:

PR Baptisms → census	PR Marriages → census	PR Burials → census
The baptism parish. This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision – this is fine, but just remember to mark it +Secondary!	The marriage parish. This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision – this is fine, but just remember to mark it +Secondary!	<p>This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision – this is fine, but just remember to mark it +Secondary!</p> <p>Burial place contains information usually about the parish but sometimes the cemetery. This is seen to be largely irrelevant for linking purposes.</p> <p>Residence place and death place rarely contain different information to what is already displayed.</p>

Important. Non-primary-variable fields should be used with caution. **Make your decision based on the primary variables first.** Only then should you consider secondary variables. The decision tree in the Appendix can help with this.

Important reminders

1. Match the primary variables (if available) first, then use the secondary variables to potentially help secure a link
2. Do not use the secondary variables before the primary ones
3. Place of residence at the time of the census is a SECONDARY VARIABLE. Use +Secondary to mark a definite candidate

Conflict checking

Our conflict solver takes the final decision on contested links – those where both linkers disagree. All information provided in the comments section helps, but the conflict solver has helpfully outlined the following to be most useful:

1. If the place of birth is a place name in the parish we are looking for, please write that. It saves me a lot of time.
2. Have you used the reverse search to get to your result? If so it should be clearly stated.
3. How did you search for your result? If you did something special to find it, which you might think the other linker and the solver wouldn't think of, share it!
4. If you have a link and the other person says it's a multiple, and you decided that you still think your link is the best, then say in the comments why you still think it's a link.
5. If you cannot decide, please write that in the comments and why.
6. I am in love with the commenters that put the id of the other multiples in the comments and shall grant them my firstborn ☺ It saves me a ton of work, though it's, in a strict sense, not necessary because I will do a search in ALA myself anyway.

FAQs

- **I'm worried I am linking incorrectly or am making the wrong decision**
Do not be concerned if you are unsure of your linking decisions as we have built in checks at many stages of our process. The Link Lives approach has adopted the principle of always having two linkers link each unit. Any 'contested' links (those the two linkers do not agree on) are then solved by a third linker (the conflict solver). Interestingly, when each pair of link files are compared, there are on average only 10-15% of links that are not identical (in census->census), which means that 85-90% of the time we are all linking in the same way. In whatever source we are linking, any conflicts are also usually halved after the conflict-checking stage, so we can be confident that this process helps to refine our 'uncontested' decisions.
- **Can I use other sources outside ALA to help me link?**
No, in order to train the algorithm you must never use sources outside ALA apart from those listed by name in 2.Primary Variables/C) Birthplaces.
- **Can't I just accept the potential link with the lowest score rather than searching each time?**
We always challenge the potential links to make sure we've been thorough. Ideally the algorithm is good enough to find each link, but then we wouldn't need human linkers! Eventually we test how good the algorithm is by comparing its links to those made by humans. Humans are good at spotting a subtle spelling error, or age difference, that a more rigid rule-based algorithm might miss.
- **I've found duplicate records. What should I do?**
Refer to the Duplicates section in the User Guide.
- **I've found a fun or interesting record**
Add it to our [Freak Links](#) file or share it in the chat!
- **What do I do with a blank record?**
Sometimes there is no information in a record. Simply mark it with the >1 button for us to deal with later. If you have only partial information (lack of birth year or last name, for

instance), check the “Full person info” fields in order to see if the information is actually there, but in the wrong field or in a non-machine readable form.

- **My ALA has shut down with no warning. Why?**

Check if you have left one of your link (.csv) files open. If so, and if you tried to save some linking decisions right before ALA shut down, this is the explanation. Close your links file and restart ALA.

If it was not because of the links-file, please describe in as much detail as possible what you did just before ALA shut down – and try to reproduce the error. Please send a mail with your description to Asbjørn or Atlanta (at@sa.dk / olivia@hum.ku.dk).

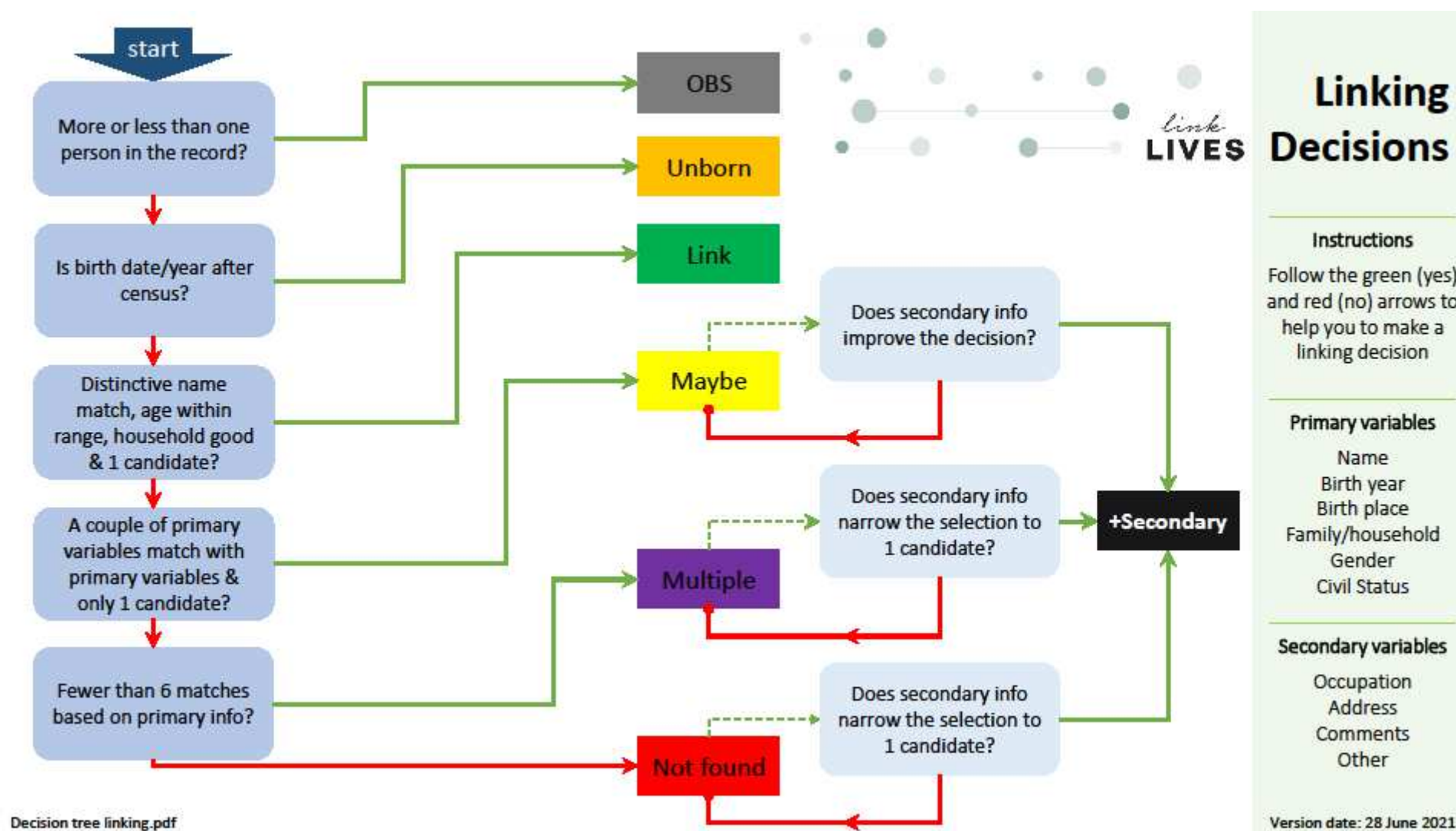
- **What kind of comments should I write in my conflict checking file?**

See the section above called Conflict Checking for more guidance. But generally: the more guidance to your decision, the better for the conflict checker.

Appendix

1. Quick guide linking decision tree

This flow diagram is a summary of the descriptions above, and should act as a guide when linking. Follow the green (yes) and red (no) lines to help you to make a decision.



2. Purposes of linking decisions

This table shows the consequences of our linking decisions.

Decision	Purpose/function
Link	<ul style="list-style-type: none"> • Trains the machine that this is a secure link. • Data is used to train the algorithm. • Helps to build life courses. • The positive decisions (ie. maybe, link etc) are usually merged for Machine Learning purposes.
Maybe	<ul style="list-style-type: none"> • Offers the ML team evidence of a link that is missing something. They are then able to identify what that missing information is and use that as a dataset in itself. • The positive decisions are usually merged for ML purposes.
+Secondary can only be selected once you have clicked Maybe, Multiple, Not Found	<ul style="list-style-type: none"> • Because the 'accepted' approach to linking is using only unchanging variables, BRE would like to capture all instances where we use OTHER information in the source to make a link.
Multiple	<ul style="list-style-type: none"> • Trains the machine that there is more than one plausible candidate based on the information available. They then assess what data is missing that is preventing a "link". • Conflict solvers use this to help them investigate a finite potential pool of candidates.
Not found	<ul style="list-style-type: none"> • Trains the machine that there is no link for this candidate based on the information available. <p>Even though it feels like a negative decision, rest assured that there is a good chance that we'll find this person later on and link him/her based on adding information from other sources. You do not need to mark things as "multiple" or "maybe" to show you have made an effort 😊.</p>
Unborn	<ul style="list-style-type: none"> • These are excluded from our training data.
OBS	<ul style="list-style-type: none"> • We analyse this dataset later using more advanced techniques, so all we need to do is flag records for this purpose. • We link on a one-to-one basis, so if there is more or less than one person in a record it cannot be included in our standard training data.

A) Examples

Here are some examples of how to use the different linking decisions in different circumstances.

Source 1 (linking from)	Source 2 (linking to)	Reasoning	Decision
Maren Henriette Poulsen (wife of Rasmus Poulsen) b. 1850 in Græsted.	Maren Henriette Ferdinandsdatter, tjenestepige b. 1850 in Græsted.	No other Maren Henriettes in the parish or amt but with no household support there's a small chance there is more than one Maren Henriette of that age born in the parish.	Maybe
Hans Krøll's wife Christine Krøll, b 1820 Krønge.	Hans Krøll's wife Kirstine Jensdatter, b 1824 Krønge.	We cannot be sure whether he remarried, because we don't have her birth name and the age is slightly differing. We take "Maybe" to flag the doubt, but we're not REALLY worried.	Maybe
Laura Nielsen født Petersen in 1863 in Gilleleje.	There are two Laur* Pe*ers* in Gilleleje: Laure Jensine Kristine Pedersen and Laura Johanne Marie Pedersen, both 2 years out. Another Laura Petersen is the same age but born in the neighbouring parish.	There is a small set of identifiable candidates and we are relatively sure the right link is one of them, but we are missing household information and further middle name identifiers.	Multiple
Petrine Petersen, wife of Laurits Petersen, b. 1861 Gilleleje. The source is also Gilleleje, so it's likely she would be there 20 years before but with her maiden name.	There are 2 Petrine names registered in the parish within her age range and a further 3 Petronella variations.	A lack of information therefore stops us from deciding on one candidate, even though the name is quite unusual.	Not found
Peter Nielsen b 1853 Græsted.	11 people are returned on a search for name, age and Græsted as birth place, some with multiple given names, but no occupation or household support to distinguish between them.	A number of people are identifiable but a lack of further distinguishing information stops us from narrowing the pool to less than 6	Not found
Robert Anders Ludviggo Lund (distinctive name).	Cannot be found despite multiple versions of his name/age/birth place searches.	This is not lack of information but some anomaly in the transcription or source (e.g. he was out of the country, not enumerated, not transcribed, source illegible etc.)	Not found
Marie Hansen (married to Johan	We cannot know what her maiden name is, even though we have a definite place of	We lack information on her surname to even begin to be able to search for her.	Not found

Hansen) b. 1875 in Stege living in Sorø.	birth, because her name is too common.		
Peter Hansen b. 1875 in Årre, tjenestekarl.	There are 18 Peter Hansens in Ribe but none born in Årre. Several are born in neighbouring parishes but none in 1874-6	With no distinguishing characteristics (e.g. unusual name(s)), and no household support, it is not possible to reduce the pool of candidates.	Not found
Burials examples to follow.			

B) Census illustration

The screen grabs below show Johan Jensen's household in the urban area of Odense in 1850 and in 1845. Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	smedemester, huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn

3. Census to census details

A) Birth places in censuses

Generally, you should expect the birth place to be exact for a good link. Some systematic errors have been identified though:

- the further away, the less precise the birth place: a person not born in the parish will have a less specific place of birth (maybe just the amt is listed or even just the country, e.g. Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong probably increases with the distance.
- an oldest child was often born in another parish (before the parents settled down) but in the census is actually registered with the birth place where he/she actually grew up instead.
- parishes within the same “pastorat” are more often confused with each other than others.
- the administrative levels of birth places can vary from census to census (e.g. farm, village, parish, town, county, country)
 - Bear it in mind as a possibility that a birth place is not always recorded as a parish name.
 - A sparse use of contemporary geographic search engines is recommended. Not for every search, but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or if they were part of the same parish.
 - <https://www.danishfamilysearch.dk/>
 - <http://www.krabsen.dk/stednavnebase/>
 - Google maps
- Birth places are often recorded as variants of “her i sognet”, “samme sted” or “ditto”, i.e. text strings which do not include the actual name of the place.
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.
 - Look for the person in the standardized birth parish (Parish Std) in order to account for potential matches which appear simply with ‘ditto’ marks.

i) Special characteristics for birth places in URBAN areas

In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility, causing:

1. the imprecision mentioned above (because of distance) and
2. a tendency to register birth *places* rather than birth *parishes*, probably because of the census enumerator’s lack of knowledge of (and interest in) exact parish names in the surrounding rural areas.

B) Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information in how to interpret and understand birth places in the censuses, see the Appendix.

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the

same person without having matching first names – and without any of the names being wrong!

- Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
- Watch out for common diminutives: Vilhelmine/Mine, Nielsine/Sine etc.
- First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names than men
 - married women taking the husband's last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen's children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription



Description	This document outlines the common approach to linking using ALA in the Link-Lives project. It contains universal guidance for all sources as well as specific considerations for individual sources (colour specific).
Linking type	Training Data
Source 1	Parish record (PR) Baptisms 1813-1917 (nationwide) (specific guidance in blue) PR Marriages 1813-1917 (nationwide) (specific guidance in yellow) PR Burials 1813-1917 (nationwide) (specific guidance in green)
Source 2	1845, 1850, 1860, 1880, 1901 censuses <i>You will find the following sources in 04 Best Practices_final.pdf:</i> <i>KBH Burials 1861-1911</i> <i>Censuses 1845, 1850, 1860, 1880, 1885 (KBH only), 1901</i>
Definitions	Source1 is the census we link from, Source2 is the census we link to.
Authors	OR, AT
ALA release	ALA v.1.6 13 Oct 2021

What's new?

1. Linking forwards for PR Baptisms & PR Marriages
2. New linking by convenience guidelines
3. New conflict checking guidelines
4. New OBS button definitions to clarify its use
5. New FAQs

Contents

From source to ALA.....	3
1. Source types.....	3
A) PR Baptisms 1813-1917	4
B) PR Marriages 1813-1917.....	6
C) PR Burials 1813-1917	7
2. Cleaning & preparing	9
3. Generating potential links.....	9
Principles of linking	10
1. Linking process.....	10
2. Linking by convenience	10
3. Linking decisions	11
4. Linking decisions chart	12
Search strategies.....	14
1. Linking units	14
2. Primary variables	15
A) Household/Family	15
B) Name.....	16
C) Birth place	16
D) Birth year.....	17
E) Civil status & gender	18
3. Secondary variables	19
Conflict checking.....	21
FAQs	21
Appendix	22
1. Quick guide linking decision tree	22
1. Purposes of linking decisions	24
A) Examples	25
B) Census illustration.....	26
2. Census to census details	27
A) Birth places in censuses	27
B) Names in censuses	27

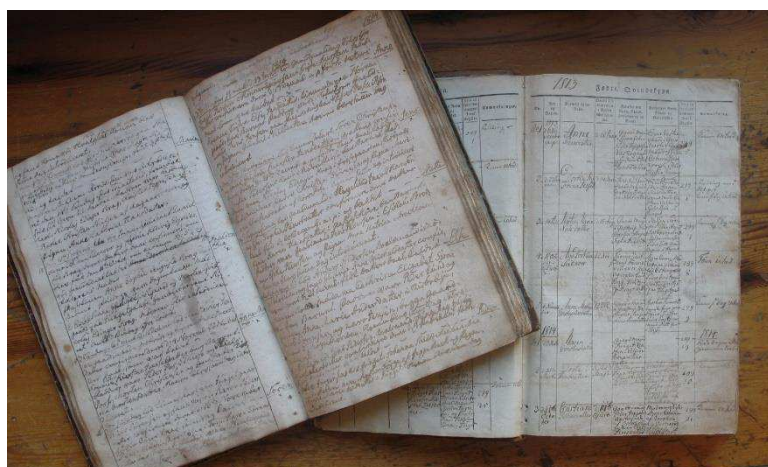
From source to ALA

This section offers some background on the sources, their genesis and how they are presented in ALA.

1. Source types

Parish Records are continuous, dynamic registers containing information on vital events of the parishioners. The core vital events registered are Baptisms, Marriages and Burials – for some periods supplemented by Confirmations, Arrivals and Departures. In Link-Lives generally, and in the ALA-linking specifically, we will concentrate on linking only the core vital events of Baptisms, Marriages and Burials – and this is why this Best Practices document only focuses on them.

Parish records were to be made by the vicar of every parish from ca. 1645-46, but many of the oldest ones have not survived. From most parishes we have parish records from the 18th century onwards, but from ca. 1813 they have been preserved from all parishes. From this year it was decided that parish records had to be kept in two versions, and it was forbidden to have the two versions under the same roof at night, to avoid potential destruction by fire by saving a duplicate. The two versions would generally be written by the vicar (præsten) and the parish clerk (deggen). Another major change in the 1813 reform was that parish records now had to be written in pre-printed books with defined columns for the single pieces of information; in earlier times, the parish records had been written just as the priest wanted, i.e. often in a much more narrative form.



The parish record books of Junget before and after the 1813 reform which introduced the pre-printed, schematized books.

In archival language, the parish records versions have the following names:

- Before 1813: Enesteministerialbog (the only version)
- After 1813: Hovedministerialbog (the vicar's version)
- After 1813: Kontraministerialbog (the parish clerk's version)

For the time after 1813, it will normally be the kontraministerialbog that has been digitized and transcribed. The two versions should be completely identical (the two writers had to “konferere” their books regularly), but occasionally of course there will be differences. This will of course generally have no importance for us, working with the transcriptions and linking on the basis only of those.

Another major change happened in 1892, this time in the content of the books. From this year onwards, the scribes had to include quite a lot more information on the individuals. This will be shown more in detail in the chapters A-C below.

Unfortunately a general problem is that not all information, that might be of value for the linking process, has been transcribed. This goes for the primary variable *marital status* and for the secondary variables *occupation* and *address*. In PR Baptisms, the mother's age is not available after 1892 in the data. So, even if this information is (sometimes) written in the original source, we cannot use it for linking because it hasn't been transcribed.

A) PR Baptisms 1813-1917

We link PR Baptisms *forward* to the next census. This strategy has of course been chosen because it is not possible to link a baptized child to an earlier census. As the examples from the original source below show, there is not a big difference in the amount of data available to link on during the period:

- The child's name
- The child's birth/baptism date
- The child's birth parish
- The parents' names (but not their marital status)

As the century progresses, the following information will join the ones mentioned above:

- Mother's age (up to 1892)

No.	Dato	Forældres fulde Navn	Døds Dato, eller Dødsårsag, eller Dødssted	Forældres Navn, Stand, Bopæl	Forældres Navn, Stand, Bopæl	Overført til det almindelige Register	Bemærkning
3	1822 Juni 27	Niels Nielsen	27 Jun	Christen Nielsen	Christen Nielsen	305	Indt. i den 4. Aug.
15	1823 Janv. 20	Christen Nielsen	20 Jan	Christen Nielsen	Christen Nielsen	305	Indt. i den 15. Febr.

Baptisms from Junget parish 1822-23, 1863, and 1900. The forms are more or less identical between the 1820s and the 1860s, while the 1900 records contain information on the mother's age. According to the post-1892 form, the full names of the parents were supposed to be entered according to their official birth or marriage certificates. Note that the records of illegitimate children generally have the same transcribed data as the legitimate ones¹ – making it impossible to distinguish between the two types in the transcription. Note also the name change recorded in the 1900 baptism – an illustration

of the fact that we're dealing with the official, legal personal name here.

¹ The presence of the unmarried mother's age in the 1863 baptism is not an expression of a consistent difference between the recording of married and unmarried mothers.

Født					Kvindekøn.		
Nr.	År og Dato.	Barnets fulde Navn.	Dødsdato enten i Kirke eller Hjem.	Barnets Navn, Stand, Paastrækning og Bopæl.	Barnets Navn, Stand og Leveforhold.	Overført i det alm. Døds. Reg.	Bemærkninger.
1863	27-6-1822	Niels Nielsen Christensen	27-6-1822	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		
1863	15-3-1863	Mette Katrine Elisabeth Jensen	15-3-1863	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		
1863	17-6-1863	Kristine Iversen	17-6-1863	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		

Født					Kvindekøn.		
Nr.	År og Dato.	Barnets fulde Navn.	Dødsdato enten i Kirke eller Hjem.	Barnets Navn, Stand, Paastrækning og Bopæl.	Barnets Navn, Stand og Leveforhold.	Overført i det alm. Døds. Reg.	Bemærkninger.
1	1900	Mette Katrine Elisabeth Jensen	15-3-1863	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		
2	1900	Kristine Iversen	17-6-1863	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Gårdmand og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm		

Examples of information in the original source (from the above illustrations). Items in bold will NOT have been transcribed:

No.	Birth date	Bap. date	Name	Parents	Comments
3	27-6-1822	27-6-1822	Niels Nielsen	Gaard Mand Niels Christensen Bak og Hustrue Maren Nielsdatter af Junget	
1	28-2-1863	15-3-1863	Mette Katrine Elisabeth Jensen Svane	Gårdejer og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	
2	17-6-1863	19-7-1863	Kristine Iversen	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	uægte barn
1	14-5-1900	31-5-1900	Marie Andersen Møjbæk Møjbæk tilføjet i H. t. Bevilling af 28/7-05	Gmd. Niels Andersen (Møjbæk) og Hustru Hansine Nielsen af Junget 37 Aar	

B) PR Marriages 1813-1917

We link PR Marriages *forward* to the nearest following census. This strategy has been chosen in order to maximize the use of the identification potential of the household information inherent in the marriage record: we can look for a couple instead of two single persons. This approach is particularly important, as the Marriage records for a large part of the period only contain information on name, age and marriage parish – the household support thus dramatically increases the possibility for reliable linking.

As the examples from the original source below show, there is a gradually increasing amount of data available as time goes on. In the beginning, we generally have the following information to link on:

- Name
- Age
- Marriage parish
- Spouse's name & age (household info)

Gradually, the following will join the ones mentioned above:

- Birth place
- Birth day
- Parents' names (all)
- Same data for the spouse

After 1892 all this information will be generally available (except maybe in the largest cities) while before 1892 often only the first batch of information will be registered. Fewer than 15% of the records contain birth places before 1892, for example, but more than 80% from 1892 onwards.

No.	Betrobte	Betrobte	Dag	Dag	Dag	Dag	Bemærkning.
1	Mig Hans	Mig Hans	1821	1821	1821	1821	304
2	Mig Hans	Mig Hans	1823	1823	1823	1823	305

Married couples from Junget parish 1821-23, 1861-62, and 1900. The forms are more or less identical between the 1820s and the 1860s, but a bit more information has been entered in the latter records: for identification purposes the more exact registration of age might be helpful, and the same goes for the vaccination information – which has not been transcribed though. The 1900 record contains information on birth date & place and both parents' names – data which has all been transcribed. According to the post-1892 form this data was supposed to be entered according to the spouses' official birth certificates.

- Name
- Age
- Burial parish
- Husband's name (if a married/widowed woman)
- Father's/parents' names (if a child)

Gradually, the following will join the ones mentioned above:

- Birth place
- Birth day
- Wife's name (if a married/widowed man)
- Parents' names (all)

After 1892 all this information will be generally available (except maybe in the largest cities) while before 1892 often only the first batch of information will be registered.

40

Døde. Mandkøn.

No.	Dødsda- gen.	Begravet- sedagen.	Den Døde For- og Efternavn.	Stand, Haandtering og Opholdssted.	Deres al- der i det første Registr.	Nummering.
1823						
2	17 ^{de} Janr.	22 ^{de} Janr.	Rasmus Christen Jensen og Minne	44	30 ⁵	1831
3	19 ^{de} Septbr.	24 ^{de} Septbr.	Jens Lauritz Christen Jensen og Minne	25	30 ⁵	1832
4	19 ^{de} Octobr.	24 ^{de} Octobr.	Jens Lauritz Christen Jensen og Minne	30 ⁵	23	1833

Buried men from Junget parish 1823, 1863, and 1900-1901. There is no difference between the registration in 1823 or 1863, even if the registration form now says that the deceased's birth place and father's name ought to be registered – if obtainable. But the registration of 1900 contains information on birth place, parents' names and spouse's name – data which has all been transcribed.

146.

Døde. Mandkøn.

No.	Dødsda- gen.	Begravet- sedagen.	Den Døde For- og Efternavn.	Den Døde Stand, Haandtering og Opholdssted samt hans Færdighed og hans Færdighed eller, hvis han er nylig faldet, hans Moder Navn, forsaavidt disse ikke tidligere er registrerede.	Deres al- der i det første Reg.	Nummering.
3	16 ^{de} Januar	21 ^{de} Januar	Jens Lauritz Jensen	Deres alder og deres færdighed	30 ⁵	1831
7	16 ^{de} Januar	21 ^{de} Januar	Kristen Christen Jensen	Deres alder og deres færdighed	30 ⁵	1832
8	24 ^{de} Janr.	29 ^{de} Janr.	Kristen Christen Jensen	Deres alder og deres færdighed	30 ⁵	1833
15	16 ^{de} Apr.	21 ^{de} Apr.	Kristen Christen Jensen	Deres alder og deres færdighed	30 ⁵	1834
6	16 ^{de} Apr.	21 ^{de} Apr.	Kristen Christen Jensen	Deres alder og deres færdighed	30 ⁵	1835

Døde				Mandkøn.				223
No.	Dødsdagen (Aar og Dagen) og Dødsstedet (Paa Landet: Byen, Sognet og Herredet, i Købstad: Købstaden og tillige Sognet, hvis der i Byen er mere end 1 Sogn).	Begravelsesstedet (Kirkegaardens, Sognets og Herredets (Købstadens) Navn) og Begravelsesdagen (Aar og Dagen).	Den afdødes fulde Navn (som vidt muligt efter forrest Døds- eller Fødselssted eller ved Efternavn i Kirke- bogen).	Den afdødes Silling og Bopæl (s): sidste faste Opholdssted, betegnet paa Landet ved Byen, Sognet og Herredet, i Køb- stad ved Købstaden, Gades og Havnens navn) samt, for saa vidt Oplysning derom kan erhverdes, tillige den afdødes Fødselssted og Fødselsdagen (ved søgte fulde Navn) fulde Navn, Silling og Bopæl. Hvis den afdøde var eller vidst at have været gift, an- føres tillige dette samt, saa vidt muligt, Aftenslættens fulde Navn og dens sidste fulde Bopæl.	Alder. (Saa den afdøde var over 1 Aar, angives Alderen med det fulde Aar.)	Af hvilken Præst Begravelsen er foretaget.	Anmærkninger.	
2.	21 August 1900 Geddeløvs By, Junget Sogn i Nord Hmd.	Thorm. Rindgaard 11 ^{de} 27 August	Lars Rindgaard (Hustru)	Endnu ingen af Aftenslættens fulde i Junget gift med Ane Jakobsdatter. Sidste fælles Bopæl i Thorm.	90 Aar	Lars Rindgaard		
5.	1901 2-4 Skalsk. Junget By og Sogn i Nord Hmd.	Junget Rindgaard 11 ^{de} 25 ^{de} April	Hans Kristian Jensen	Indsiddet i Junget. Født i Gassum, Nørhald Herred; Søn af Rind. Hans Jensen og Hustru. Abelone Kristensen af Gassum. Gift med Ane Kirstine Svenningsen. Sidste fælles Bopæl Junget.	47 Aar	Lars Rindgaard		

Examples of information in the original source (from the above illustrations). Items in bold will NOT have been transcribed:

No.	Death date	Bur. date	Name	Occupation and family	Age
2	17-1-1823	26-1-1823	Rasmus Erichsen	Huus og Muurmester i Junget	44 Aar
5	5-4-1863	12-4-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75 År
1	21-3-1901 Junget by	28-3-1901	Hans Kristian Jensen	Indsiddet i Junget. Født i Gassum Nørhald Herred; Søn af Hmd. Hans Kr. Jensen og Hustru Abelone Kristensen af Gassum. Gift med Ane Kirstine Svenningsen. Sidste fælles Bopæl Junget.	47 Aar

2. Cleaning & preparing

The data has been prepared in a way that makes sense for both Link-Lives and ALA linkers. The PR data was transcribed in Asia by a company paid by Ancestry, and we must expect that a set of instructions was given to the transcribers directing them on which field to record each piece of information in. As with any historical source, however, be aware that errors, omissions and subjective decisions are inevitable at each stage. For example, for almost all fields we find misplaced or misread information from other fields.

3. Generating potential links

Potential links are generated for each record entry (where possible) by running a rule-based algorithm based on gender and age. The most likely 10 are then displayed on the right side of the ALA screen in order of probability score (0 is the closest match). These act as start-point suggestions for linkers. We might find the correct link within this list, or we might not, but we always attempt to challenge them.

Principles of linking

We link according to an agreed set of guidelines that help our team link consistently. The focus is on linking the right ones: we want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links. The process is as follows:

1. Linking process

1. Linking
 - a. Each linking unit is always linked by at least two linkers.
2. Conflict checking
 - a. After each finishes their linking unit, our linking coordinator identifies all the contested links (i.e. those which the linkers disagreed on) and creates a conflict file in Excel for each linker to check.
 - b. Each linker works through the conflict file to check and comment on each contested link.
3. Conflict solving
 - a. The coordinator solves any remaining contested links
 - b. A consolidated training data file is generated and stored.

Golden Linking Rules

1. Only link one person to one person
2. Always challenge a potential link
3. Always search for a competing candidate
4. Always link using primary variables first, before assessing secondary variables
5. Never use unauthorized sources outside of ALA
6. When in doubt, do not link

2. Linking by convenience

Linking by convenience means searching individually for, then making a decision on, a person with a key role in a Source1 record, and only linking non-key people that appear in the record if they are present in the same census household in Source2. Key persons are defined as follows:

PR Baptisms → census	PR Marriages → census	PR Burials → census
Key persons: <i>Main (child being baptized)</i> <i>Mother</i> <i>Father</i>	Key persons: <i>Bride</i> <i>Groom</i>	Key persons: <i>Deceased</i>
Non-key-persons: <i>All others</i>	Non-key-persons: <i>All others</i>	Non-key-persons: <i>All others</i>

For PR records, always link by convenience:

1. Individually search (and make a link decision) for everyone listed as key persons
2. Only make a link decision for a non key person if they appear in the census in the same household as the key person. Otherwise leave blank.
2. Do not individually search for a non key person
3. Use the same linking decision for the non key person as for the key person
4. Do not link non key persons if the key person is not linked

3. Linking decisions

We link using two sets of variables: primary variables in principle do not vary during life (or, in the case of the name and marital status, if it changes, the change can be expected to follow some relatively well-defined rules). Secondary variables may contain information that gives context to the individual's life, but which generally changes – or can change unpredictably – over time. Link Lives uses the following fields to distinguish primary from secondary:

Primary variables	Secondary variables
Name	Occupation
Birth year	Address
Birth place	Comments
Family/household	Other
Gender	
Marital Status	

In the parish record transcriptions we get very few secondary variables, even when they are in the original source (occupation, for instance, is not transcribed). This means that the following (general) description of how and when to use the primary and secondary variables in searching is less relevant in the PR context than when linking other sources, but we leave them here, as they DO apply at least to the different geographically defined fields (i.e. address possibilities).

- Identify a candidate by searching on primary variables first.
- It may then be relevant to remove one of the criteria, expand the years or remove them entirely, in order to exhaust alternatives.
- Only then should you consider the secondary variables.

The following chart shows the different linking decisions, when to make them and how they are used.² See the Appendix for a flowchart diagram to assist in your linking, a fuller list of examples and an overview of the consequences of our linking decisions.

² When the linking decisions are processed, Maybes, Links and those with +Secondary are marked 'positive' (shaded green below) and the rest are marked 'negative' (shaded red below). This distinction simply helps to train the machine algorithm.

4. Linking decisions chart

Decision	Definition	How it is commonly used
Link	<p>When you are confident that you have found the correct link.</p> <p><i>Quantitative hint: You should be at least 99% sure.</i></p>	<p>We are usually most sure of links when primary variables all contain satisfying information for a secure link to be established. E.g. there are specific characteristics of name, age, and (for censuses) household and/or birth place.</p> <p>This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.</p>
Maybe	<p>When you are almost sure you have found the correct link, but you want to flag that there is some information missing.</p> <p><i>Quantitative hint: You should be at least 95% sure.</i></p>	<p>We tend to link “maybe” when one of the characteristics is missing but the others are solid. E.g. distinctive name, wrong age.</p> <p>Candidates often include wives who take their husbands’ names, but whose other characteristics are otherwise a match.</p>
<p>+Secondary³ can only be selected once you have clicked Maybe, Multiple, Not Found</p>	<p>When you use information secondary variables to move from a “Maybe”, “Multiple” or “Not Found” to a single plausible candidate.</p> <p>Maybe +Secondary One option is made more certain by using secondary information.</p> <p>Multiple +Secondary 2-5 options narrowed to one using secondary information.</p> <p>Not Found +Secondary > 5 options narrowed to one using secondary information.</p> <p><i>Quantitative hint: After using the secondary information, you should be at least 95% sure.</i></p>	<p>Candidates in census-to-census linking often include people with distinctive professions or trades involving some training, e.g. lærer, or occupations that are out of place in the local context.</p> <p>In parish records we might use geographical indicators of residence to narrow a selection to one candidate.</p>

³ For linkers who have used ALA versions before, please note there is no longer an ‘Occupation’ decision button.

Decision	Definition	How it is commonly used
Multiple	When you have a well-defined group of no more than 5 plausible candidates, but you cannot choose between them.	e.g. 3 x Maren Madsens from the same parish of the same age and no household support to distinguish between them. e.g. Peter Jens Knudsen died in 1880, and there are 3 Peter Knudsen candidates and 2 Jens Knudsen candidates all in the same age range.
Not found	<p>Not found can be applied to several situations. You do not have to mark which is which. The following is just for explanatory purposes:</p> <p>1) Unfound When you find no plausible candidate, even with the presence of distinctive characteristics.</p> <p>2) Too many When there are too many candidates (>5) to choose from and you lack primary information to narrow the pool (though see +Secondary above).</p> <p>3) Unfindable Where there is missing information that makes it impossible to identify any candidate (though see +Secondary above).</p> <p>4) Presumed dead When, linking forward, you expect this person to be dead (e.g. if linking a PR marriage to the following census, and you find the wife as a widow)</p>	<p>Examples include:</p> <p>1) Distinctive characteristics unfound e.g. Robert Anders Ludviggo Lund but despite multiple search strategies, you did not find him.</p> <p>2) Lots of local possibilities e.g. one Maren Madsen from the right parish, but her age is 4 years out, then half a dozen Maren Madsens born in neighbouring parishes with correct ages.</p> <p>3a) Common given names e.g. Marie Sørensen, wife of Hans Sørensen, no maiden name.</p> <p>3b) Vague places of birth e.g. Hans Hansen, b. 1875, Slesvig.</p> <p>3c) Missing transcription information e.g. Ubeboet, Mangler, Jens B[?], no place of birth, no household support.</p>
Unborn	If it is probable that the person was not yet born at the time of the census.	Used for people whose birth years are calculated to within a year of a census year and who are missing at the end of a family grouping. If no household support, simply mark Not Found.
OBS	When you find records which you simply cannot link because they contain no persons or more than one person. However, we do not expect to find this in the parish records.	

Search strategies

Refer to the User Guide for full details of search functions, shortcut keys and search terms.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The information available to link baptisms to the subsequent census is more or less the same throughout the period in question.</p> <p>The most important linking variables are name and age of the child combined with the household information inherent in the baptism: we know who the child's parents are.</p>	<p>The information available to link married people to the subsequent census varies with chronology and geography. But generally it's of huge importance to take the household information inherent in the marriage record into account when linking.</p> <p>The fullest individual information we get after 1892 is outside of the biggest towns: here we normally gain also birth date & place, and parents' names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend on the specific characteristics of the linking unit.</p>	<p>The information available to link buried people to the previous census varies greatly with chronology, geography, gender, civil status and age. The smallest amount of information is given about adult men in the early period and/or in bigger towns: often we have only name, age and burial parish. The fullest information we get after 1892 outside of the biggest towns: here we normally gain also birth date & place, parents' and spouse's names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend heavily on the specific characteristics of the linking unit and the individual deceased persons.</p>

1. Linking units

We continue to use [Trello](#) to track the linking production process. A [video](#) explaining in more detail how the board works can be found on Teams.

In parish record linking, we have introduced a new process for defining link units. We aim to link all records within set year ranges (e.g. 1860-1880) for a particular parish.

The PR data is organized in entire, transcribed books, and this is also how you'll find them displayed in ALA. The records will appear in the same order as in the original source.

PR Baptisms → census	PR Marriages → census	PR Burials → census
Your linking unit will be defined as the baptisms from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link baptisms from X parish to a given census (e.g. Junget 1860). You will then link	Your linking unit will be defined as the marriages from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link marriages from X parish to a given census (e.g. Junget 1860). You will then link	Your linking unit will be defined as the burials from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link burials from X parish to a given census (e.g. Junget 1860). You will then link records falling

records falling between 1 Feb of the year of the previous census to 31 Jan of the year of the census in question (e.g. 1/2/1850-31/1/1860).

NB:

- You may need to link more than one 'book' to cover the whole link period.
- Most books contain both boys and girls: sometimes the genders are mixed, sometimes first we get all the boys, then all the girls. Both genders must be linked.

See User Guide for full details of the new process for selecting and noting start and end records.

records falling between 1 Feb of the year of the previous census to 31 Jan of the year of the census in question (e.g. 1/2/1850-31/1/1860).

NB:

- You may need to link more than one 'book' to cover the whole link period.
- **See User Guide for full details of the new process** for selecting and noting start and end records.

between 1 Feb of that census to 31 Jan of the next census (e.g. 1/2/1860-31/1/1880).

NB:

- You may need to link more than one 'book' to cover the whole link period.
- Most books contain both men and women: sometimes the genders are mixed, sometimes first we get all the men, then all the women. Both genders must be linked.
- **See User Guide for full details of the new process** for selecting and noting start and end records.

2. Primary variables

Primary variables in principle do not vary during life (or, in the case of the name or the marital status, if it changes, the change can be expected to follow some relatively well-defined rules).

A) Household/Family

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Generally, it's of huge importance to take the household information inherent in the baptism record into account when linking: we know who the child's parents are, and the child will – depending on the year range in question – most often live with the parents in the next census.</p> <p>Be aware though, that the child might be illegitimate, and that you can therefore not 100% expect the parents to be married at the time of the census: it is possible to be unmarried and have children – and we cannot distinguish between legitimate and illegitimate children from the transcribed data. The child</p>	<p>Generally, it's of huge importance to take the household information inherent in the marriage record into account when linking.</p> <p>Be aware though, of the possibility that</p> <ul style="list-style-type: none"> • one/both of the spouses died before the census • the couple was separated/divorced before the census • the couple did not live together at the time of the census (but were still listed with "married" as marital status) 	<p>When the household information is available, it's usually in the form of the names of a spouse or of the parent(s).</p> <p>The name of a spouse – most often a deceased woman's husband – can be a good identifier when linking the deceased to the previous census. But remember that we cannot know from the transcribed PR data whether the spouse was alive or had died many years ago – or had been changed between the census and the death.</p> <p>Very often the names of the parents will not be of much help when linking a deceased person to the census</p>

might also live in an institution or with foster parents.

immediately before the death, as we must expect many of the deceased individuals to have not been cohabiting with their parents for many years.

B) Name

Generally, you should expect the name to be precise for a good link (apart from spelling variations). If the name does not fit, we are hesitant to link unless we have supporting information to confirm it.

Names in parish records

In parish records, names should generally be expected to be more “correct” than in censuses – especially in the later parish records.

PR Baptisms → census	PR Marriages → census	PR Burials → census
Children’s names at baptisms have a special status, as they represent the person’s legal name, even though this concept of legality was definitely less clear for Danes in the first half of the 19 th century than nowadays.	Especially for marriages after 1892, you should expect to get the full “legal” name, including maiden name, for each of the spouses (it says in the registration form that the record should be based on the birth certificate).	Especially for burials after 1892, you should expect to get the full “legal” name, including maiden name, for each deceased person.
Before a new law in 1828 – and sometimes even after – it was often only the child’s first name that was registered in the parish records. In those cases you have to deduce the last name from the father’s name, and depending on local customs this could be both a derived or an inherited patronym – or a family name.	But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.	But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.
With illegitimate children the last name may even come from the mother.		

Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information on how to interpret and understand birth places in the censuses, see the Appendix.

C) Birth place

Birth place is a very strong identification variable, especially when confirmed by name and age/birth date. If it is registered in the source, it is very important to use it in the linking process.

For information on how to interpret and understand birth places in the censuses, see the Appendix.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The child's birth place is generally the parish where the baptism was recorded – though sometimes there is a birth parish and a baptism parish.</p> <p>The birth parish of the parents is never recorded.</p>	<p>After 1892 it was obligatory to register the birth place of the spouses – based on the information from their birth certificates. This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR Marriages than in censuses, probably because the birth certificate was actually used.</p> <p>Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.</p>	<p>After 1892 it was obligatory to register the birth place of the deceased (if available). This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR burials than in censuses, probably because the birth certificate was used when registering the deceased.</p> <p>Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.</p>

D) Birth year

The birth year in the ALA main screen is, depending on the PR event in question (baptism, marriage or burial), either directly taken from the original source or calculated from a combination of age and event year. In some cases, there is no birth year because of lacking information, but you might as linkers be able to derive it manually from the context of the data – maybe you can deduce a lacking death year for a deceased person from the death dates of the records above and below, and through that calculate a birth year. It is important to check also the original information in the “Full person info” screen if you lack information. It might be there, but in a not-machine readable form.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The birth year of the child is, except for a few odd cases, equal to the baptism year.</p> <p>The later records in the period up until 1892 will contain an age for the mother (never the father). This will give a calculated birth year for the mother on which to search.</p> <p>Baptisms often occurred up to a year after birth, so a search of +/- 2 years is a reasonable date ranges to search on.</p>	<p>The birth year information in the main screen displays both original [Birth Year and Age] and calculated [*Birth year] data. The *BirthYear field calculates a birth year from the marriage age and the marriage year.</p>	<p>The birth year information in the main screen [*Birth year] is a derived field which combines available information from the original fields of</p> <ul style="list-style-type: none"> • Birth year • Burial age/Death age • Death year/Burial year <p>If there is not enough information to calculate the birth year automatically, [*Birth year] will be empty. If that is the case, then look through the information in the “Full person info” screen in order to see if there is enough information for you to</p>

		manually calculate a birth year for linking purposes.
--	--	---

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking. However, when linking to the 1901 census, the birth date is more likely to be an exact match.

General age searching recommendations

1. Age 0-49: Use calculated birth year "+3"
2. Age 50-64: Use calculated birth year "+5"
3. Age 65-: Use calculated birth year "+10"
4. To search for any age, run a 'blank age search' using #

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend the year range. The closer we get to the 20th century, though, birth information is generally more trustworthy.

i) Unborns

PR Baptisms → census	PR Marriages → census	PR Burials → census
This is not a relevant possibility, as we are linking forwards.	This is not a relevant possibility, as we are linking forwards.	<p>People who died before or were born after the year of the census you are linking to, are pre-marked as Unborn.</p> <p>This means that if you are linking to the 1860 census, a person who</p> <ul style="list-style-type: none"> • died in 1859 or • was born in 1861 <p>will be pre-marked as Unborn.</p>

E) Civil status & gender

These are considered to be primary variables. They are often used to exclude possible matches.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The child will only very rarely have a marital status other than "unmarried" in the next census.</p> <p>Be aware that you cannot expect the parents to be necessarily married, together or the same in the following census: the child might be born outside wedlock, fostered etc.</p>	<p>The spouses' marital statuses haven't been transcribed from the original sources, but as long as we're linking forwards, this doesn't influence our linking possibilities.</p> <p>When linking forward to the following census you'd generally need very strong evidence to support linking</p>	<p>Civil status has not been transcribed from the original sources. It can, though, be partially derived from indirect information:</p> <p>Before 1892</p> <ul style="list-style-type: none"> • If a spouse is registered, this means that the deceased was either

	<p>from a marriage to an unmarried person. All other marital statuses – married, widowed, separated/divorced – are possible matches.</p> <p>Marriage records are probably those with the least gender confusion in the data.</p> <p>People other than bride + groom show up in the marriages. Until the parents-in-law become regularly registered (in the 20th century), we CANNOT trust the roles assigned in the data.</p>	<p>married or widowed (but we don't know which)</p> <ul style="list-style-type: none"> • If a deceased woman has no spouse registered, expect her to be unmarried, but don't be too sure (especially do not expect anything in big towns where data is known to be scarce) • Spouses are generally not registered for deceased men. This means that the absence of information tells us nothing about his marital status. • After 1892. If a spouse is registered, this means that the deceased was either married or widowed (but we don't know which) • If no spouse is registered, this is a rather secure sign that the deceased was unmarried (it was obligatory to register the spouse). This, unfortunately, goes only for rural parishes and minor towns. In bigger towns, data is more scarce <p>People other than bride + groom show up in the marriages. Until the parents-in-law become regularly registered (in the 20th century), we CANNOT trust the roles assigned in the data.</p>
--	--	---

3. Secondary variables

All variables in ALA that are not listed above as primary variables are considered secondary information. We use the +Secondary button in ALA when we use anything other than primary variables to make our linking decision.

In all sources we have access to extra fields containing information that might be relevant for finding someone and in certain cases to make linking decisions. There are not many of them in the PRs, though. For instance, unfortunately occupation wasn't transcribed from the original source.

The secondary variables available include:

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The baptism parish. This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable.</p> <p>It is important to know the place of 'birth registration', otherwise we risk missing people. We put greatest weight on the parish of baptism followed by place of actual birth. If nothing appears under BaptismParish in a record, you should assume the child was born in the parish of baptism. If something is displayed in BaptismParish, you should search on BOTH BaptismParish AND Birth Parish in the census.</p>	<p>The marriage parish. This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision – this is fine, but just remember to mark it +Secondary!</p>	<p>This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision – this is fine, but just remember to mark it +Secondary!</p> <p>Burial place contains information usually about the parish but sometimes the cemetery. This is seen to be largely irrelevant for linking purposes.</p> <p>Residence place and death place rarely contain different information to what is already displayed.</p>

Important. Non-primary-variable fields should be used with caution. **Make your decision based on the primary variables first.** Only then should you consider secondary variables. The decision tree in the Appendix can help with this.

Important reminders

1. Match the primary variables (if available) first, then use the secondary variables to potentially help secure a link
2. Do not use the secondary variables before the primary ones
3. Search for and make a link decision on every key person
4. Place of residence at the time of the census is a SECONDARY VARIABLE. Use +Secondary to mark a definite candidate

Conflict checking

A conflict is when your linking decision differs from that of another linker. After linking a link unit, you will be invited to conflict-check those link decisions on which you and the other linker disagreed in an Excel file. Your linker number depends on which order you linked the unit – check Trello.

- a) Address each case in the Excel spreadsheet and see whether you stand by your own link decision or whether you agree more with the other one.
- b) You will see linker1's decisions on the left side of the sheet (shaded blue) and linker2's on the right (shaded green).
- c) Use ALA to look up each case if necessary.
- d) Write 'linker1' or 'linker2' in the column headed 'mere_sandsynlig' according to your assessment. You can also write 'none' if you now believe neither is the correct decision.
- e) In the column headed 'kommentar', we ask you to write a short description as to why you made that choice.
- f) Append the Excel file name with your initials (e.g. _OR) then upload to the Trello folder you downloaded it from. Mark the Trello card done.

When conflict checking parish record linking units (ie. that contain key and non-key persons) you only need to make a decision and give an explanation for the key person. The non-main decisions will be taken care of by a python script.

Once both conflict files are complete, our conflict solver then takes the final decision on contested links – those where both linkers disagree, even after conflict checking. This is done on your comments alone, so please be specific when writing these.

FAQs

- **I'm worried I am linking incorrectly or am making the wrong decision**
Do not be concerned if you are unsure of your linking decisions as we have built in checks at many stages of our process. The Link Lives approach has adopted the principle of always having two linkers link each unit. Any 'contested' links (those the two linkers do not agree on) are then solved by a third linker (the conflict solver). Interestingly, when each pair of link files are compared, there are on average only 10-15% of links that are not identical (in census->census), which means that 85-90% of the time we are all linking in the same way. In whatever source we are linking, any conflicts are also usually halved after the conflict-checking stage, so we can be confident that this process helps to refine our 'uncontested' decisions.
- **Can I use other sources outside ALA to help me link?**
No, in order to train the algorithm you must never use sources outside ALA apart from those listed by name in 2. Primary Variables/C) Birthplaces.
- **Can't I just accept the potential link with the lowest score rather than searching each time?**
We always challenge the potential links to make sure we've been thorough. Ideally the algorithm is good enough to find each link, but then we wouldn't need human linkers! Eventually we test how good the algorithm is by comparing its links to those made by humans. Humans are good at spotting a subtle spelling error, or age difference, that a more rigid rule-based algorithm might miss.

- **I've found duplicate records. What should I do?**
Refer to the Duplicates section in the User Guide.
- **I've found a fun or interesting record**
Add it to our [Freak Links](#) file or share it in the chat!
- **What do I do with a blank record?**
Sometimes there is no information in a record. Simply mark it with the OBS button for us to deal with later. If you have only partial information (lack of birth year or last name, for instance), check the "Full person info" fields in order to see if the information is actually there, but in the wrong field or in a non-machine readable form.
- **My ALA has shut down with no warning. Why?**
Check if you have left one of your link (.csv) files open. If so, and if you tried to save some linking decisions right before ALA shut down, this is the explanation. Close your links file and restart ALA.

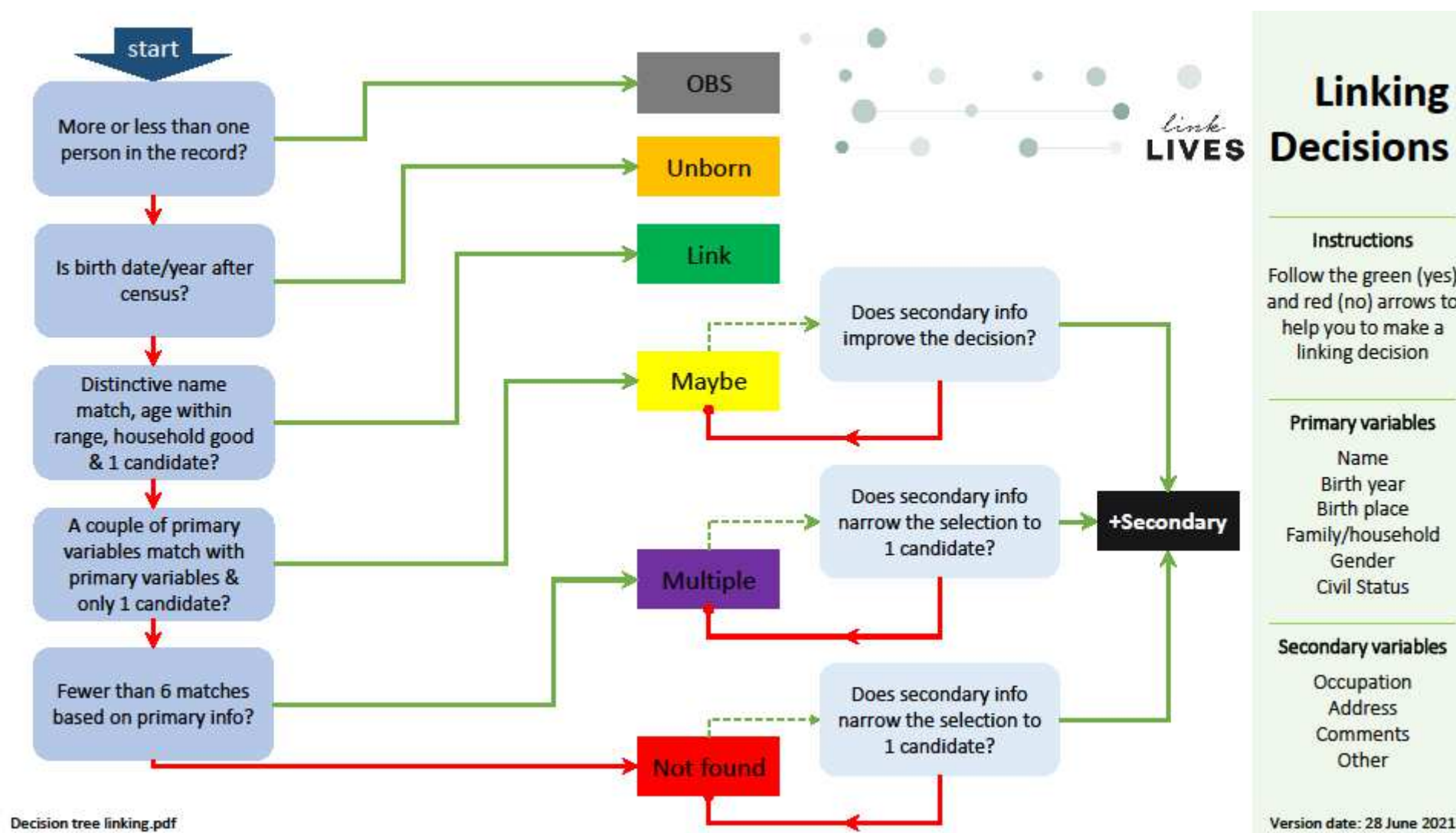
If it was not because of the links-file, please write an email describing in as much detail as possible what you did just before ALA shut down. Find the following file in your ALA/data folder: error.log. Attach this to your mail and send it to Olivia (olivia@hum.ku.dk).

- **What kind of comments should I write in my conflict checking file?**
See the section above called Conflict Checking for more guidance. But generally: the more guidance to your decision, the better for the conflict solver.
- **The non-main person has been listed as a spouse/father/opposite sex to what was expected. How do I treat that?**
This is most likely a transcription error, where the names of the relatives for the deceased were put in the wrong column (father, spouse, mother etc) by the transcriber. You should trust the raw data (names, ages, etc.) rather than the structural information (e.g. roles which often must be derived from the transcribers understanding of the context...).

Appendix

1. Quick guide linking decision tree

This flow diagram is a summary of the descriptions above, and should act as a guide when linking. Follow the green (yes) and red (no) lines to help you to make a decision.



1. Purposes of linking decisions

This table shows the consequences of our linking decisions.

Decision	Purpose/function
Link	<ul style="list-style-type: none"> • Trains the machine that this is a secure link. • Data is used to train the algorithm. • Helps to build life courses. • The positive decisions (ie. maybe, link etc) are usually merged for Machine Learning purposes.
Maybe	<ul style="list-style-type: none"> • Offers the ML team evidence of a link that is missing something. They are then able to identify what that missing information is and use that as a dataset in itself. • The positive decisions are usually merged for ML purposes.
+Secondary can only be selected once you have clicked Maybe, Multiple, Not Found	<ul style="list-style-type: none"> • Because the 'accepted' approach to linking is using only unchanging variables, BRE would like to capture all instances where we use OTHER information in the source to make a link.
Multiple	<ul style="list-style-type: none"> • Trains the machine that there is more than one plausible candidate based on the information available. They then assess what data is missing that is preventing a "link". • Conflict solvers use this to help them investigate a finite potential pool of candidates.
Not found	<ul style="list-style-type: none"> • Trains the machine that there is no link for this candidate based on the information available. <p>Even though it feels like a negative decision, rest assured that there is a good chance that we'll find this person later on and link him/her based on adding information from other sources. You do not need to mark things as "multiple" or "maybe" to show you have made an effort 😊.</p>
Unborn	<ul style="list-style-type: none"> • These are excluded from our training data.
OBS	<ul style="list-style-type: none"> • We analyse this dataset later using more advanced techniques, so all we need to do is flag records for this purpose. • We link on a one-to-one basis, so if there is more or less than one person in a record it cannot be included in our standard training data.

A) Examples

Here are some examples of how to use the different linking decisions in different circumstances.

Source 1 (linking from)	Source 2 (linking to)	Reasoning	Decision
Maren Henriette Poulsen (wife of Rasmus Poulsen) b. 1850 in Græsted.	Maren Henriette Ferdinandsdatter, tjenestepige b. 1850 in Græsted.	No other Maren Henriettes in the parish or amt but with no household support there's a small chance there is more than one Maren Henriette of that age born in the parish.	Maybe
Hans Krøll's wife Christine Krøll, b 1820 Krønge.	Hans Krøll's wife Kirstine Jensdatter, b 1824 Krønge.	We cannot be sure whether he remarried, because we don't have her birth name and the age is slightly differing. We take "Maybe" to flag the doubt, but we're not REALLY worried.	Maybe
Laura Nielsen født Petersen in 1863 in Gilleleje.	There are two Laur* Pe*ers* in Gilleleje: Laure Jensine Kristine Pedersen and Laura Johanne Marie Pedersen, both 2 years out. Another Laura Petersen is the same age but born in the neighbouring parish.	There is a small set of identifiable candidates and we are relatively sure the right link is one of them, but we are missing household information and further middle name identifiers.	Multiple
Petrine Petersen, wife of Laurits Petersen, b. 1861 Gilleleje. The source is also Gilleleje, so it's likely she would be there 20 years before but with her maiden name.	There are 2 Petrine names registered in the parish within her age range and a further 3 Petronella variations.	A lack of information therefore stops us from deciding on one candidate, even though the name is quite unusual.	Not found
Peter Nielsen b 1853 Græsted.	11 people are returned on a search for name, age and Græsted as birth place, some with multiple given names, but no occupation or household support to distinguish between them.	A number of people are identifiable but a lack of further distinguishing information stops us from narrowing the pool to less than 6	Not found
Robert Anders Ludviggo Lund (distinctive name).	Cannot be found despite multiple versions of his name/age/birth place searches.	This is not lack of information but some anomaly in the transcription or source (e.g. he was out of the country, not enumerated, not transcribed, source illegible etc.)	Not found
Marie Hansen (married to Johan	We cannot know what her maiden name is, even though we have a definite place of	We lack information on her surname to even begin to be able to search for her.	Not found

Hansen) b. 1875 in Stege living in Sorø.	birth, because her name is too common.		
Peter Hansen b. 1875 in Årre, tjenestekarl.	There are 18 Peter Hansens in Ribe but none born in Årre. Several are born in neighbouring parishes but none in 1874-6	With no distinguishing characteristics (e.g. unusual name(s)), and no household support, it is not possible to reduce the pool of candidates.	Not found
Burials examples to follow.			

B) Census illustration

The screen grabs below show Johan Jensen's household in the urban area of Odense in 1850 and in 1845. Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	smedemester, huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn

2. Census to census details

A) Birth places in censuses

Generally, you should expect the birth place to be exact for a good link. Some systematic errors have been identified though:

- the further away, the less precise the birth place: a person not born in the parish will have a less specific place of birth (maybe just the amt is listed or even just the country, e.g. Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong probably increases with the distance.
- an oldest child was often born in another parish (before the parents settled down) but in the census is actually registered with the birth place where he/she actually grew up instead.
- parishes within the same “pastorat” are more often confused with each other than others.
- the administrative levels of birth places can vary from census to census (e.g. farm, village, parish, town, county, country)
 - Bear it in mind as a possibility that a birth place is not always recorded as a parish name.
 - A sparse use of contemporary geographic search engines is recommended. Not for every search, but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or if they were part of the same parish.
 - <https://www.danishfamilysearch.dk/>
 - <http://www.krabsen.dk/stednavnebase/>
 - Google maps
- Birth places are often recorded as variants of “her i sognet”, “samme sted” or “ditto”, i.e. text strings which do not include the actual name of the place.
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.
 - Look for the person in the standardized birth parish (Parish Std) in order to account for potential matches which appear simply with ‘ditto’ marks.

i) Special characteristics for birth places in URBAN areas

In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility, causing:

1. the imprecision mentioned above (because of distance) and
2. a tendency to register birth *places* rather than birth *parishes*, probably because of the census enumerator’s lack of knowledge of (and interest in) exact parish names in the surrounding rural areas.

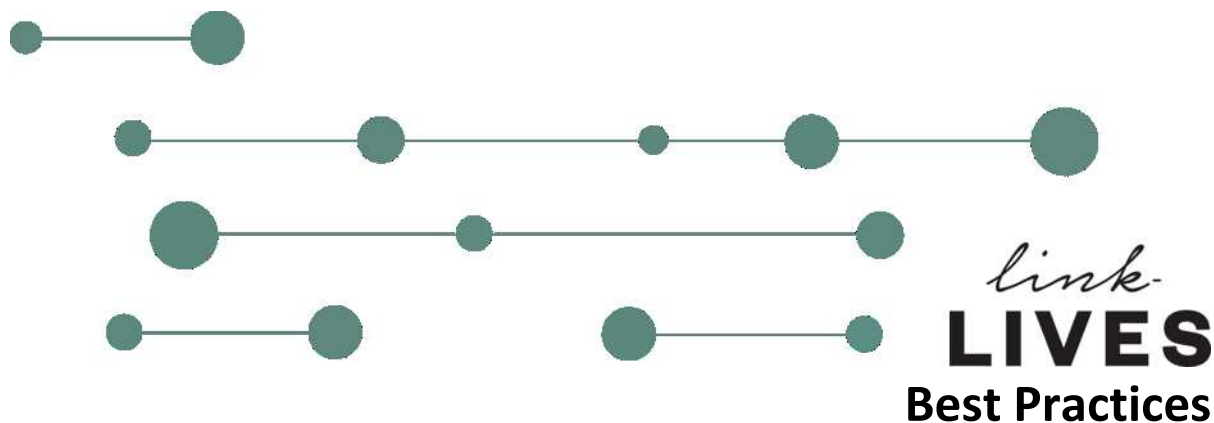
B) Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information in how to interpret and understand birth places in the censuses, see the Appendix.

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the

same person without having matching first names – and without any of the names being wrong!

- Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
- Watch out for common diminutives: Vilhelmine/Mine, Nielsine/Sine etc.
- First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names than men
 - married women taking the husband's last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen's children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription



Description	This document outlines the common approach to linking using ALA in the Link-Lives project. It contains universal guidance for all sources as well as specific considerations for individual sources (colour specific).
Linking type	Training Data
Source 1	Parish record (PR) Baptisms 1813-1917 (nationwide) (specific guidance in blue) PR Marriages 1813-1917 (nationwide) (specific guidance in yellow) PR Burials 1813-1917 (nationwide) (specific guidance in green)
Source 2	1845, 1850, 1860, 1880, 1901 censuses <i>You will find the following sources in 04 Best Practices_final.pdf:</i> <i>KBH Burials 1861-1911</i> <i>Censuses 1845, 1850, 1860, 1880, 1885 (KBH only), 1901</i>
Definitions	Source1 is the census we link from, Source2 is the census we link to.
Authors	OR, AT
ALA release	ALA v.1.8 1 Dec 2021

What's new?

1. New search function suggests unambiguous candidates (replaces potentials)
2. New search field enables birth day, month and year searching
3. Data fixed so we can link to 1901 census from all sources
4. Household proximity search crash fixed
5. Guidance on assessing birth year and age values for PR Marriages

Contents

Principles of linking	3
1. Linking process	3
2. Linking by convenience	3
3. Linking decisions	4
4. Linking decisions chart	5
Search strategies	7
1. Linking units	7
2. Primary variables	8
A) Household/Family	8
B) Name	9
C) Birth place	10
D) Birth year	10
E) Civil status & gender	12
3. Secondary variables	13
Conflict checking	14
FAQs	15
Quick guide linking decision tree	16
Appendix	18
From source to ALA	18
1. Source types	18
F) PR Baptisms 1813-1917	19
G) PR Marriages 1813-1917	21
H) PR Burials 1813-1917	22
2. Cleaning & preparing	24
3. Generating potential links	24
4. Purposes of linking decisions	24
A) Examples	26
B) Census illustration	27
Source-specific information	28
C) Census to census details	28

Principles of linking

We link according to an agreed set of guidelines that help our team link consistently. The focus is on linking the right ones: we want to get as many links as we can, but we do not want to link people if we do not have enough evidence to say that they are correct links. The process is as follows:

1. Linking process

1. Linking
 - a. Each linking unit is always linked by at least two linkers.
2. Conflict checking
 - a. After each finishes their linking unit, our linking coordinator identifies all the contested links (i.e. those which the linkers disagreed on) and creates a conflict file in Excel for each linker to check.
 - b. Each linker works through the conflict file to check and comment on each contested link.
3. Conflict solving
 - a. The coordinator solves any remaining contested links
 - b. A consolidated training data file is generated and stored.

Golden Linking Rules

1. Only link one person to one person
2. Always challenge a potential link
3. Always search for a competing candidate
4. Always link using primary variables first, before assessing secondary variables
5. Never use unauthorized sources outside of ALA
6. When in doubt, do not link

2. Linking by convenience

Linking by convenience means searching individually for, then making a decision on, a person with a key role in a Source1 record, and only linking non-key people that appear in the record if they are present in the same census household in Source2. Key persons are defined as follows:

PR Baptisms → census	PR Marriages → census	PR Burials → census
Key persons: <i>Main (child being baptized)</i> <i>Mother</i> <i>Father</i>	Key persons: <i>Bride</i> <i>Groom</i>	Key persons: <i>Deceased</i>
Non-key-persons: <i>All others</i>	Non-key-persons: <i>All others</i>	Non-key-persons: <i>All others</i>

For PR records, always link by convenience:

1. Individually search (and make a link decision) for everyone listed as key persons
2. Only make a link decision for a non key person if they appear in the census in the same household as the key person. Otherwise leave blank.
2. Do not individually search for a non key person
3. Use the same linking decision for the non key person as for the key person
4. Do not link non key persons if the key person is not linked
5. In cases where a non-main appears in the same household as a main but is assigned the wrong role (e.g. spouse when it's a father etc), do not link. Ignore, ie. treat as if that person was not found in the household.

3. Linking decisions

We link using two sets of variables: primary variables in principle do not vary during life (or, in the case of the name and marital status, if it changes, the change can be expected to follow some relatively well-defined rules). Secondary variables may contain information that gives context to the individual's life, but which generally changes – or can change unpredictably – over time. Link Lives uses the following fields to distinguish primary from secondary:

Primary variables	Secondary variables
Name	Occupation
Birth year	Address
Birth place	Comments
Family/household	Other
Gender	
Marital Status	

In the parish record transcriptions we get very few secondary variables, even when they are in the original source (occupation, for instance, is not transcribed). This means that the following (general) description of how and when to use the primary and secondary variables in searching is less relevant in the PR context than when linking other sources, but we leave them here, as they DO apply at least to the different geographically defined fields (i.e. address possibilities).

- Identify a candidate by searching on primary variables first.
- It may then be relevant to remove one of the criteria, expand the years or remove them entirely, in order to exhaust alternatives.
- Only then should you consider the secondary variables.
- The following chart shows the different linking decisions, when to make them and how they are used.¹ See the Appendix for a flowchart diagram to assist in your linking, a fuller list of examples and an overview of the consequences of our linking decisions.

¹ When the linking decisions are processed, Maybes, Links and those with +Secondary are marked 'positive' (shaded green below) and the rest are marked 'negative' (shaded red below). This distinction simply helps to train the machine algorithm.

4. Linking decisions chart

Decision	Definition	How it is commonly used
Link	<p>When you are confident that you have found the correct link.</p> <p><i>Quantitative hint: You should be at least 99% sure.</i></p>	<p>We are usually most sure of links when primary variables all contain satisfying information for a secure link to be established. E.g. there are specific characteristics of name, age, and (for censuses) household and/or birth place.</p> <p>This rule of thumb can – in rare cases – be overruled if some of the other variables contain information that is sufficiently convincing.</p>
Maybe	<p>When you are almost sure you have found the correct link, but you want to flag that there is some information missing.</p> <p><i>Quantitative hint: You should be at least 95% sure.</i></p>	<p>We tend to link “maybe” when one of the characteristics is missing but the others are solid. E.g. distinctive name, wrong age.</p> <p>Candidates often include wives who take their husbands’ names, but whose other characteristics are otherwise a match.</p>
<p>+Secondary² can only be selected once you have clicked Maybe, Multiple, Not Found</p>	<p>When you use information secondary variables to move from a “Maybe”, “Multiple” or “Not Found” to a single plausible candidate.</p> <p>Maybe +Secondary One option is made more certain by using secondary information.</p> <p>Multiple +Secondary 2-5 options narrowed to one using secondary information.</p> <p>Not Found +Secondary > 5 options narrowed to one using secondary information.</p> <p><i>Quantitative hint: After using the secondary information, you should be at least 95% sure.</i></p>	<p>Candidates in census-to-census linking often include people with distinctive professions or trades involving some training, e.g. lærer, or occupations that are out of place in the local context.</p> <p>In parish records we might use geographical indicators of residence to narrow a selection to one candidate.</p> <p>If you link someone, then later in the unit find out more information about that person, you can go back and change your mind. But you must mark the decision +Secondary. This shows that the link decision was made based on further information.</p>

² For linkers who have used ALA versions before, please note there is no longer an ‘Occupation’ decision button.

Decision	Definition	How it is commonly used
Multiple	When you have a well-defined group of no more than 5 plausible candidates, but you cannot choose between them.	e.g. 3 x Maren Madsens from the same parish of the same age and no household support to distinguish between them. e.g. Peter Jens Knudsen died in 1880, and there are 3 Peter Knudsen candidates and 2 Jens Knudsen candidates all in the same age range.
Not found	<p>Not found can be applied to several situations. You do not have to mark which is which. The following is just for explanatory purposes:</p> <p>1) Unfound When you find no plausible candidate, even with the presence of distinctive characteristics.</p> <p>2) Too many When there are too many candidates (>5) to choose from and you lack primary information to narrow the pool (though see +Secondary above).</p> <p>3) Unfindable Where there is missing information that makes it impossible to identify any candidate (though see +Secondary above).</p> <p>4) Presumed dead When, linking forward, you expect this person to be dead (e.g. if linking a PR marriage to the following census, and you find the wife as a widow. Or a child baptized but with no name given.)</p>	<p>Examples include:</p> <p>1) Distinctive characteristics unfound e.g. Robert Anders Ludviggo Lund but despite multiple search strategies, you did not find him.</p> <p>2) Lots of local possibilities e.g. one Maren Madsen from the right parish, but her age is 4 years out, then half a dozen Maren Madsens born in neighbouring parishes with correct ages.</p> <p>3a) Common given names e.g. Marie Sørensen, wife of Hans Sørensen, no maiden name.</p> <p>3b) Vague places of birth e.g. Hans Hansen, b. 1875, Slesvig.</p> <p>3c) Missing transcription information e.g. Ubeboet, Mangler, Jens B[?], no place of birth, no household support.</p>
Unborn	If it is probable that the person was not yet born at the time of the census.	Used for people whose birth years are calculated to within a year of a census year and who are missing at the end of a family grouping. If no household support, simply mark Not Found.
OBS	When you find records which you simply cannot link because they contain no persons or more than one person.	

Search strategies

Refer to the User Guide for full details of search functions, shortcut keys and search terms.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The information available to link baptisms to the subsequent census is more or less the same throughout the period in question.</p> <p>The most important linking variables are name and age of the child combined with the household information inherent in the baptism: we know who the child's parents are.</p> <p>The order of searching for baptisms (sometimes using 2 person searching) is shown below. Continue down the list until you have made a link decision for all key people.</p> <ol style="list-style-type: none"> 1) child + one of the parents (usually start with the most unusual name) 2) child + the other parent 3) child alone (most useful when there is a big gap between baptism and census but can be useful for all time differences) 4) parent 1 + parent 2 / parent 2 + parent 1 5) parent 1 6) parent 2 	<p>The information available to link married people to the subsequent census varies with chronology and geography. But generally it's of huge importance to take the household information inherent in the marriage record into account when linking.</p> <p>The fullest individual information we get after 1892 outside of the biggest towns: here we normally gain also birth date & place, and parents' names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend on the specific characteristics of the linking unit.</p>	<p>The information available to link buried people to the previous census varies greatly with chronology, geography, gender, civil status and age. The smallest amount of information is given about adult men in the early period and/or in bigger towns: often we have only name, age and burial parish. The fullest information we get after 1892 outside of the biggest towns: here we normally gain also birth date & place, parents' and spouse's names.</p> <p>The search strategy employed and the amount of securely identifiable links therefore depend heavily on the specific characteristics of the linking unit and the individual deceased persons.</p>

1. Linking units

We continue to use [Trello](#) to track the linking production process. A [video](#) explaining in more detail how the board works can be found on Teams.

In parish record linking, we have introduced a new process for defining link units. We aim to link all records within set year ranges (e.g. 1860-1880) for a particular parish.

The PR data is organized in entire, transcribed books, and this is also how you'll find them displayed in ALA. The records will appear in the same order as in the original source.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Your linking unit will be defined as the baptisms from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link baptisms from X parish to a given census (e.g. Junget 1860). You will then link records falling between 1 Feb of the year of the previous census to 31 Jan of the year of the census in question (e.g. 1/2/1850-31/1/1860).</p> <p>NB:</p> <ul style="list-style-type: none"> You may need to link more than one 'book' to cover the whole link period. Most books contain both boys and girls: sometimes the genders are mixed, sometimes first we get all the boys, then all the girls. Both genders must be linked. <p>See User Guide for full details of this process for selecting and noting start and end records.</p>	<p>Your linking unit will be defined as the marriages from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link marriages from X parish to a given census (e.g. Junget 1860). You will then link records falling between 1 Feb of the year of the previous census to 31 Jan of the year of the census in question (e.g. 1/2/1850-31/1/1860).</p> <p>NB:</p> <ul style="list-style-type: none"> You may need to link more than one 'book' to cover the whole link period. See User Guide for full details of this process for selecting and noting start and end records. 	<p>Your linking unit will be defined as the burials from a specific parish and from a specific year range which will correspond to the time between two censuses. For instance, you'll be told to link burials from X parish to a given census (e.g. Junget 1860). You will then link records falling between 1 Feb of that census to 31 Jan of the next census (e.g. 1/2/1860-31/1/1880).</p> <p>NB:</p> <ul style="list-style-type: none"> You may need to link more than one 'book' to cover the whole link period. Most books contain both men and women: sometimes the genders are mixed, sometimes first we get all the men, then all the women. Both genders must be linked. See User Guide for full details of this process for selecting and noting start and end records.

2. Primary variables

Primary variables in principle do not vary during life (or, in the case of the name or the marital status, if it changes, the change can be expected to follow some relatively well-defined rules).

A) Household/Family

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Generally, it's of huge importance to take the household information inherent in the baptism record into account when linking: we know who the child's parents are, and the child will – depending on the year range in question – most often live with the parents in the next census.</p>	<p>Generally, it's of huge importance to take the household information inherent in the marriage record into account when linking.</p> <p>Be aware though, of the possibility that</p> <ul style="list-style-type: none"> one/both of the spouses died before the census 	<p>When the household information is available, it's usually in the form of the names of a spouse or of the parent(s).</p> <p>The name of a spouse – most often a deceased woman's husband – can be a good identifier when linking the deceased to the previous census. But remember that we</p>

Be aware though, that the child might be illegitimate, and that you can therefore not 100% expect the parents to be married at the time of the census: it is possible to be unmarried and have children – and we cannot distinguish between legitimate and illegitimate children from the transcribed data. The child might also live in an institution or with foster parents.	<ul style="list-style-type: none"> the couple was separated/divorced before the census the couple did not live together at the time of the census (but were still listed with “married” as marital status) 	<p>cannot know from the transcribed PR data whether the spouse was alive or had died many years ago – or had been changed between the census and the death.</p> <p>Very often the names of the parents will not be of much help when linking a deceased person to the census immediately before the death, as we must expect many of the deceased individuals to have not been cohabiting with their parents for many years.</p>
---	--	--

B) Name

Generally, you should expect the name to be precise for a good link (apart from spelling variations). If the name does not fit, we are hesitant to link unless we have supporting information to confirm it.

Names in parish records

In parish records, names should generally be expected to be more “correct” than in censuses – especially in the later parish records.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Children’s names at baptisms have a special status, as they represent the person’s legal name, even though this concept of legality was definitely less clear for Danes in the first half of the 19th century than nowadays.</p> <p>Before a new law in 1828 – and sometimes even after – it was often only the child’s first name that was registered in the parish records. In those cases you have to deduce the last name from the father’s name, and depending on local customs this could be both a derived or an inherited patronym – or a family name.</p> <p>With illegitimate children the last name may even come from the mother.</p> <p>For København parishes</p>	<p>Especially for marriages after 1892, you should expect to get the full “legal” name, including maiden name, for each of the spouses (it says in the registration form that the record should be based on the birth certificate).</p> <p>But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.</p>	<p>Especially for burials after 1892, you should expect to get the full “legal” name, including maiden name, for each deceased person.</p> <p>But even before 1892, you should count quite a lot on the names being full – i.e. all first names are registered, and rarely in an abbreviated form.</p>

Try searching only with the child's (often many) first names – without the last name. And to remember that mothers more frequently use the fathers' last name than their own birth name.

Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information on how to interpret and understand birth places in the censuses, see the Appendix.

C) Birth place

Birth place is a very strong identification variable, especially when confirmed by name and age/birth date. If it is registered in the source, it is very important to use it in the linking process.

For information on how to interpret and understand birth places in the censuses, see the Appendix.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The child's birth place is generally the parish where the baptism was recorded.</p> <p>Use the linking unit parish as a proxy for place of birth UNLESS another place is listed in the place of birth field.</p> <p>The birth parish of the parents is rarely recorded.</p>	<p>After 1892 it was obligatory to register the birth place of the spouses – based on the information from their birth certificates. This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR Marriages than in censuses, probably because the birth certificate was actually used.</p> <p>Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.</p>	<p>After 1892 it was obligatory to register the birth place of the deceased (if available). This means that birth place is registered far more coherently after 1892 than before. Experience suggests that post-1892 birth places are often more reliable in PR burials than in censuses, probably because the birth certificate was used when registering the deceased.</p> <p>Experience also suggests that birth places registered before 1892 might depend more on the local registration culture and less on formal rules – and may therefore be less secure.</p>

D) Birth year

The birth year in the ALA main screen is, depending on the PR event in question (baptism, marriage or burial), either directly taken from the original source or calculated from a combination of age and event year. In some cases, there is no birth year because of lacking information, but you might as linkers be able to derive it manually from the context of the data – maybe you can deduce a lacking death year for a deceased person from the death dates of the records above and below, and through that calculate a birth year. It is important to check also the original information in the “Full person info” screen if you lack information. It might be there, but in a not-machine readable form.

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The birth year of the child is, except for a few odd cases, equal to the baptism year.</p> <p>The later records in the period up until 1892 will contain an age for the mother (never the father). This will give a calculated birth year for the mother on which to search.</p> <p>Baptisms often occurred up to a year after birth, so a search of +/- 2 years is a reasonable date ranges to search on.</p>	<p>The birth year information in the main screen displays both original [Birth Year] and [Age] and calculated [*Birth year] data. The *BirthYear field calculates a birth year from the marriage age and the marriage year. If there is no age, [*Birth year] is filled with the content from [Birth year].</p> <p>If you see a difference between the two birth year columns, we suggest you give extra weight to the earliest date when searching.</p> <p>This is because some of the transcribed data contains dates that were not actually birth dates but have got mixed in, incorrectly.</p>	<p>The birth year information in the main screen [*Birth year] is a derived field which combines available information from the original fields of</p> <ul style="list-style-type: none"> • Birth year • Burial age/Death age • Death year/Burial year <p>If there is not enough information to calculate the birth year automatically, [*Birth year] will be empty. If that is the case, then look through the information in the "Full person info" screen in order to see if there is enough information for you to manually calculate a birth year for linking purposes.</p>

Generally, you should expect the birth year to be within a 2 year range for a good link. A difference of 1 year cannot be given any weight when linking.

When linking to the 1901 census, the birth date is more likely to be an exact match. Use the following search formats:

- To search for 13 June 1868
Search format: 13-06-1868
- To search for is June 1868
Search format: 00-06-1868
- To search x years either side of the year
Search format: 13-06-1868+3
- To search x units either side of the day or month
Search format: 13+3-06-1868
Search format: 13-06+3-1868

General age searching recommendations

1. Age 0-49: Use calculated birth year "+3"
2. Age 50-64: Use calculated birth year "+5"
3. Age 65-: Use calculated birth year "+10"
4. To search for any age, run a 'blank age search' using #

These are general recommendations which should be used flexibly. If, for instance, a person has a very rare name and is not found within the recommended year range, it would be clever to extend

the year range. The closer we get to the 20th century, though, birth information is generally more trustworthy.

i) **Unborns / deceased**

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>Only children recorded as 'dødfødt' should be marked as unborn.</p> <p>Just because a child does not have a name, does not make it dødfødt. In empty-name cases, mark Not Found.</p> <p>If you make a link decision on a baptism and then later find that a child with the same name and parents appears in a later baptism, you can probably assume that the first child died, and the parents named the next child the same. In this case, simply cancel your decision on the first child, mark it Not Found (NOT unborn) and then make the link decision as normal on the secondly-baptised child.</p>	<p>This is not a relevant possibility, as we are linking forwards.</p>	<p>People who died before or were born after the year of the census you are linking to, are pre-marked as Unborn.</p> <p>For any child listed as 'dødfødt' use the Unborn decision button.</p> <p>This means that if you are linking to the 1860 census, a person who</p> <ul style="list-style-type: none"> died in 1859 or was born in 1861 <p>will be pre-marked as Unborn.</p>

E) *Civil status & gender*

These are considered to be primary variables. They are often used to exclude possible matches. Be aware that women can be searched both as 'Kvinde' (K) or 'Female' (F).

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The child will only very rarely have a marital status other than "unmarried" in the next census.</p> <p>Be aware that you cannot expect the parents to be necessarily married, together or the same in the following census: the child might be born outside wedlock, fostered etc.</p>	<p>The spouses' marital statuses haven't been transcribed from the original sources, but as long as we're linking forwards, this doesn't influence our linking possibilities.</p> <p>When linking forward to the following census you'd generally need very strong evidence to support linking from a marriage to an unmarried person. All other marital statuses – married, widowed,</p>	<p>Civil status has not been transcribed from the original sources. It can, though, be partially derived from indirect information:</p> <p>Before 1892</p> <ul style="list-style-type: none"> If a spouse is registered, this means that the deceased was either married or widowed (but we don't know which) If a deceased woman has no spouse registered, expect her to be unmarried, but don't be too sure (especially do not expect anything in big towns

	<p>separated/divorced – are possible matches.</p> <p>Marriage records are probably those with the least gender confusion in the data.</p> <p>People other than bride + groom show up in the marriages. Until the parents-in-law become regularly registered (in the 20th century), we CANNOT trust the roles assigned in the data.</p>	<p>where data is known to be scarce)</p> <ul style="list-style-type: none"> Spouses are generally not registered for deceased men. This means that the absence of information tells us nothing about his marital status. After 1892. If a spouse is registered, this means that the deceased was either married or widowed (but we don't know which) If no spouse is registered, this is a rather secure sign that the deceased was unmarried (it was obligatory to register the spouse). This, unfortunately, goes only for rural parishes and minor towns. In bigger towns, data is more scarce <p>People other than bride + groom show up in the marriages. Until the parents-in-law become regularly registered (in the 20th century), we CANNOT trust the roles assigned in the data.</p>
--	---	---

3. Secondary variables

All variables in ALA that are not listed above as primary variables are considered secondary information. We use the +Secondary button in ALA when we use anything other than primary variables to make our linking decision.

In all sources we have access to extra fields containing information that might be relevant for finding someone and in certain cases to make linking decisions. There are not many of them in the PRs, though. For instance, unfortunately occupation wasn't transcribed from the original source.

The secondary variables available include:

PR Baptisms → census	PR Marriages → census	PR Burials → census
<p>The baptism parish. This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable.</p> <p>It is important to know the place of 'birth registration',</p>	<p>The marriage parish. This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision –</p>	<p>This is the parish you loaded as your linking unit (e.g. Junget). Since people move around, we do not consider residence to be a primary variable. Using your judgement you might decide that this is a key determining factor in making a decision – this is fine, but just</p>

otherwise we risk missing people. We put greatest weight on the parish of baptism followed by place of actual birth. If nothing appears under BaptismParish in a record, you should assume the child was born in the parish of baptism. If something is displayed in BaptismParish, you should search on BOTH BaptismParish AND Birth Parish in the census.

this is fine, but just remember to mark it +Secondary!

remember to mark it +Secondary!

Burial place contains information usually about the parish but sometimes the cemetery. This is seen to be largely irrelevant for linking purposes.

Residence place and death place rarely contain different information to what is already displayed.

Important. Non-primary-variable fields should be used with caution. **Make your decision based on the primary variables first.** Only then should you consider secondary variables. The decision tree in the Appendix can help with this.

Important reminders

1. Match the primary variables (if available) first, then use the secondary variables to potentially help secure a link
2. Do not use the secondary variables before the primary ones
3. Search for and make a link decision on every key person
4. Place of residence at the time of the census is a SECONDARY VARIABLE. Use +Secondary to mark a definite candidate

Conflict checking

A conflict is when your linking decision differs from that of another linker. After linking a link unit, you will be invited to conflict-check those link decisions on which you and the other linker disagreed in an Excel file. Your linker number depends on which order you linked the unit – check Trello.

- a) Address each case in the Excel spreadsheet and see whether you stand by your own link decision or whether you agree more with the other one.
- b) You will see linker1's decisions on the left side of the sheet (shaded blue) and linker2's on the right (shaded green).
- c) Use ALA to look up each case if necessary.
- d) Write 'linker1' or 'linker2' in the column headed 'mere_sandsynlig' according to your assessment. You can also write 'none' if you now believe neither is the correct decision.
- e) In the column headed 'kommentar', we ask you to write a short description as to why you made that choice.
- f) Append the Excel file name with your initials (e.g. _OR) then upload to the Trello folder you downloaded it from. Mark the Trello card done.

When conflict checking parish record linking units (ie. that contain key and non-key persons) you only need to make a decision and give an explanation for the key person. The non-main decisions will be taken care of by a python script.

Once both conflict files are complete, our conflict solver then takes the final decision on contested links – those where both linkers disagree, even after conflict checking. This is done on your comments alone, so please be specific when writing these.

FAQs

- **I'm worried I am linking incorrectly or am making the wrong decision**

Do not be concerned if you are unsure of your linking decisions as we have built in checks at many stages of our process. The Link Lives approach has adopted the principle of always having two linkers link each unit. Any 'contested' links (those the two linkers do not agree on) are then solved by a third linker (the conflict solver). Interestingly, when each pair of link files are compared, there are on average only 10-15% of links that are not identical (in census->census), which means that 85-90% of the time we are all linking in the same way. In whatever source we are linking, any conflicts are also usually halved after the conflict-checking stage, so we can be confident that this process helps to refine our 'uncontested' decisions.

- **Can I use other sources outside ALA to help me link?**

No, in order to train the algorithm you must never use sources outside ALA apart from those listed by name in 2. Primary Variables/C) Birthplaces.

- **Can't I just accept the potential link with the lowest score rather than searching each time?**

We always challenge the potential links to make sure we've been thorough. Ideally the algorithm is good enough to find each link, but then we wouldn't need human linkers! Eventually we test how good the algorithm is by comparing its links to those made by humans. Humans are good at spotting a subtle spelling error, or age difference, that a more rigid rule-based algorithm might miss.

- **I've found duplicate records. What should I do?**

If you find duplicates in the source2:

- Assign the link to the record with the lowest id-number
- Mark the higher id-number with the button **OBS**
- Write the case in the Excel file [Duplicate Cases in ALA.xlsx](#) in Teams, including the linking unit details (year and parish) for each source. Choose the correct tab (ark) for the source you're using.

If you find a duplicate in source1, simply make a link decision as normal.

- **I've found a fun or interesting record**

Add it to our [Freak Links](#) file or share it in the chat!

- **What do I do with a blank record?**

Sometimes there is no information in a record. Simply mark it with the OBS button for us to deal with later. If you have only partial information (lack of birth year or last name, for instance), check the "Full person info" fields in order to see if the information is actually there, but in the wrong field or in a non-machine readable form.

- **My ALA has shut down with no warning. Why?**

Check if you have left one of your link (.csv) files open. If so, and if you tried to save some linking decisions right before ALA shut down, this is the explanation. Close your links file and restart ALA.

If it was not because of the links-file, please write an email describing in as much detail as possible what you did just before ALA shut down. Find the following file in your ALA/data folder: error.log. Attach this to your mail and send it to Nicolai (nrm@sa.dk).

- **What kind of comments should I write in my conflict checking file?**

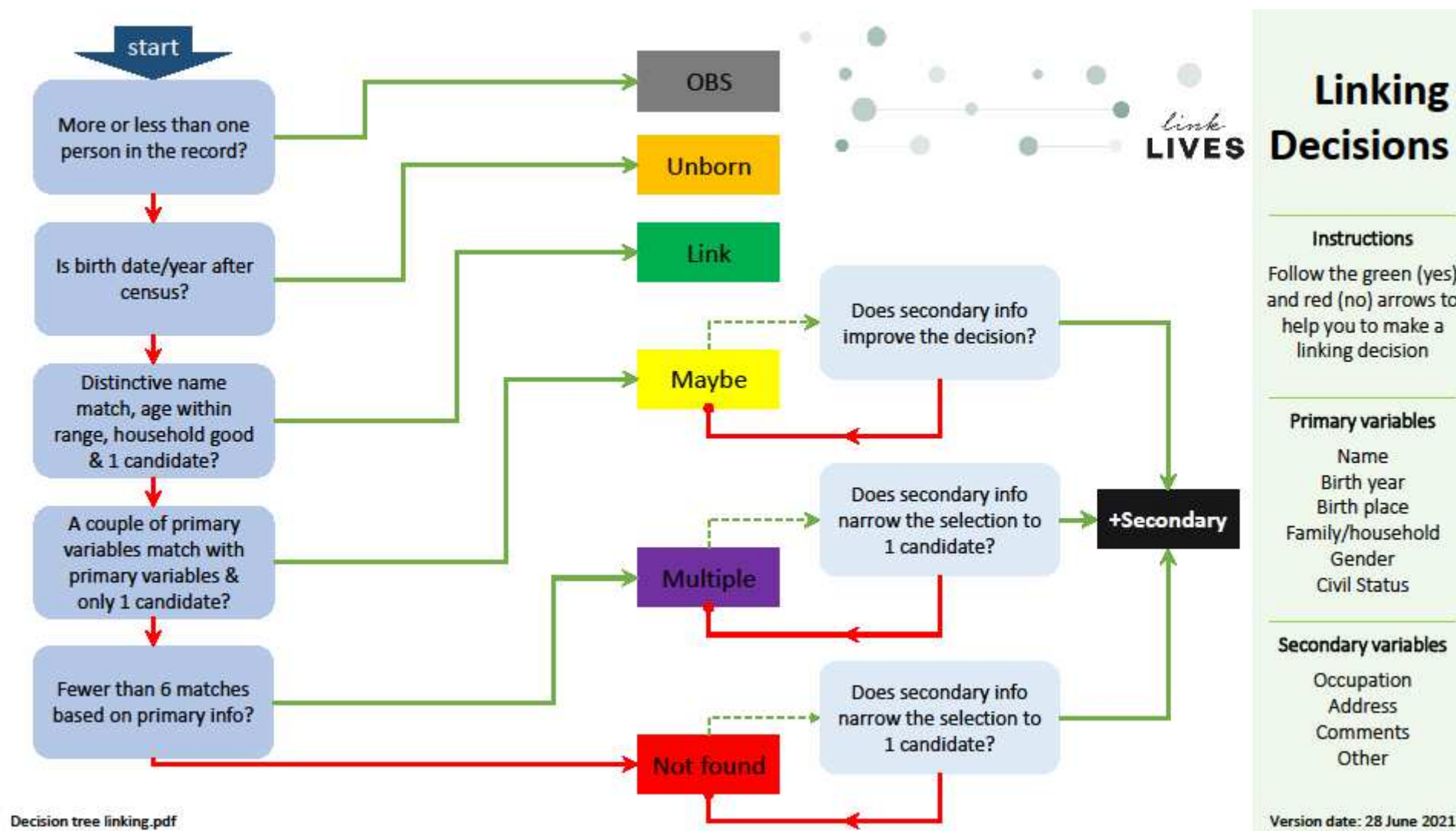
See the section above called Conflict Checking for more guidance. But generally: the more guidance to your decision, the better for the conflict solver.

- **The non-main person has been listed as a spouse/father/opposite sex to what was expected. How do I treat that?**

This is most likely a transcription error, where the names of the relatives for the deceased were put in the wrong column (father, spouse, mother etc) by the transcriber. You should trust the raw data (names, ages, etc.) rather than the structural information (e.g. roles which often must be derived from the transcribers understanding of the context...).

Quick guide linking decision tree

This flow diagram is a summary of the descriptions above, and should act as a guide when linking. Follow the green (yes) and red (no) lines to help you to make a decision.



Appendix

The appendix gives more context to our sources.

From source to ALA

This section offers some background on the sources, their genesis and how they are presented in ALA.

1. Source types

Parish Records are continuous, dynamic registers containing information on vital events of the parishioners. The core vital events registered are Baptisms, Marriages and Burials – for some periods supplemented by Confirmations, Arrivals and Departures. In Link-Lives generally, and in the ALA-linking specifically, we will concentrate on linking only the core vital events of Baptisms, Marriages and Burials – and this is why this Best Practices document only focuses on them.

Parish records were to be made by the vicar of every parish from ca. 1645-46, but many of the oldest ones have not survived. From most parishes we have parish records from the 18th century onwards, but from ca. 1813 they have been preserved from all parishes. From this year it was decided that parish records had to be kept in two versions, and it was forbidden to have the two versions under the same roof at night, to avoid potential destruction by fire by saving a duplicate. The two versions would generally be written by the vicar (præsten) and the parish clerk (deggen). Another major change in the 1813 reform was that parish records now had to be written in pre-printed books with defined columns for the single pieces of information; in earlier times, the parish records had been written just as the priest wanted, i.e. often in a much more narrative form.



The parish record books of Junget before and after the 1813 reform which introduced the pre-printed, schematized books.

In archival language, the parish records versions have the following names:

- Before 1813: Enesteministerialbog (the only version)
- After 1813: Hovedministerialbog (the vicar's version)
- After 1813: Kontraministerialbog (the parish clerk's version)

For the time after 1813, it will normally be the kontraministerialbog that has been digitized and transcribed. The two versions should be completely identical (the two writers had to "konferere" their books regularly), but occasionally of course there will be differences. This will of course

generally have no importance for us, working with the transcriptions and linking on the basis only of those.

Another major change happened in 1892, this time in the content of the books. From this year onwards, the scribes had to include quite a lot more information on the individuals. This will be shown more in detail in the chapters A-C below.

Unfortunately a general problem is that not all information, that might be of value for the linking process, has been transcribed. This goes for the primary variable *marital status* and for the secondary variables *occupation* and *address*. In PR Baptisms, the mother's age is not available after 1892 in the data. So, even if this information is (sometimes) written in the original source, we cannot use it for linking because it hasn't been transcribed.

F) PR Baptisms 1813-1917

We link PR Baptisms *forward* to the next census. This strategy has of course been chosen because it is not possible to link a baptized child to an earlier census. As the examples from the original source below show, there is not a big difference in the amount of data available to link on during the period:

- The child's name
- The child's birth/baptism date
- The child's birth parish
- The parents' names (but not their marital status)

As the century progresses, the following information will join the ones mentioned above:

- Mother's age (up to 1892)

No.	Nar on Datum	Barndt fulde Navn.	Døds D. tum, enten i huse eller hjemme me.	Forældres Navn, Stand, Gaardsteds og Bopæl.	Barndt fulde Navn Stand og Opholdssted.	For an- ført i det almind. Frent. Register.	Andet Lign.
3	27 Janv.	Nils Nilsen	27 Jan	Christian Nilsen Borgermester i Junget	Christian Nilsen Borgermester i Junget	305	Indt. i den 4. Aug.
12	20 Janv.	Christian Nilsen	20 Jan	Christian Nilsen Borgermester i Junget	Christian Nilsen Borgermester i Junget	305	Indt. i den 4. Aug.
13	20 Janv.	Nils Nilsen	20 Jan	Christian Nilsen Borgermester i Junget	Christian Nilsen Borgermester i Junget	305	Indt. i den 4. Aug.

Baptisms from Junget parish 1822-23, 1863, and 1900. The forms are more or less identical between the 1820s and the 1860s, while the 1900 records contain information on the mother's age. According to the post-1892 form, the full names of the parents were supposed to be entered according to their official birth or marriage certificates. Note that the records of illegitimate children generally have the same transcribed data as the legitimate ones³ – making it impossible to distinguish between the two types in the transcription. Note also the name change recorded in the 1900 baptism – an illustration

of the fact that we're dealing with the official, legal personal name here.

³ The presence of the unmarried mother's age in the 1863 baptism is not an expression of a consistent difference between the recording of married and unmarried mothers.

Fødte					Kvindekøn.		
No.	År og Dato.	Barnets fulde Navn.	Tidstid Datoen er indført i Kirken eller Bogen.	Barnets Navn, Sted, Døbtid og Bøgel.	Barnets Navn, Sted og Døbtid.	Overført i det sin. Bøgel. Bøgel.	Bemærkninger.
1	1863 27-6-1822	Mette Katrine Christensen	1822-6-27	Gårdejer og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	1822-6-27	Indført i Bogen.
2	17-6-1863	Kristine Iversen	1863-6-17	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	1863-6-17	uægte barn

Fødte Kvindekøn.

89

Nr.	Fødselsår og Dato.	Hvor født? (Paa Landet: Byen, Sognet og Herredet. Har Gården eller Huset været her, har det været i Kirken? Kosterne og tilgængelige, hvis der i Byen er mere end i Sogn, samt deres alder Gæder og Husmænd.)	Barnets fulde Navn.	Parternes fulde Navn efter Forættelse af Døds- eller i alt Fald Væbnings- eller ved Eftersyn i Kirkebogen; deres Stilling og Bøgel; samt for Moderens Væbnings- og Forættelse af hendes Alder med det fulde År. (Barnets egne Fødsels- eller døds- eller ved Væbnings- eller Eftersyn i Kirkebogen.)	Daadagen.	Hvem der har foretaget Daadagen, og om den er foretaget i Kirken eller hjemme. Ved Hjemmehav indføres tilgængelige Datoen for Fremstillingen i Kirken.	Fødselsår fulde Navn, Stilling og Bøgel.	Bemærkninger. (Her indføres blandt andet, hvad der vedkommer Reklamation af Barnets Fødselsstedet.)
1.	1900 14-5-1900 Maj	Junget By og Sogn Niels Møller	Marie Andersen Møjbæk	Gmd. Niels Andersen Møjbæk og Hustru Hansine Nielsen af Junget 37 Aar.	31-5-1900 Maj	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	1. Niels Andersen & Hustru Hansine Nielsen af Junget 37 Aar.	Junget-Thorsen
2.	1900 14-5-1900 Juni	Junget By og Sogn Niels Møller	Anna Johanne Andersen	Gmd. Niels Andersen og Hustru Marie Møjbæk af Junget 35 Aar.	31-5-1900 Maj	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	1. Niels Andersen & Hustru Hansine Nielsen af Junget 37 Aar.	Dr. —

Examples of information in the original source (from the above illustrations). Items in bold will NOT have been transcribed:

No.	Birth date	Bap. date	Name	Parents	Comments
3	27-6-1822	27-6-1822	Niels Nielsen	Gaard Mand Niels Christensen Bak og Hustrue Maren Nielsdatter af Junget	
1	28-2-1863	15-3-1863	Mette Katrine Elisabeth Jensen Svane	Gårdejer og Sognefoged Jens Peter Svane og Hustru Mette Marie Rasmussen af Brokholm	
2	17-6-1863	19-7-1863	Kristine Iversen	Ugift fruentimmer Johanne Kristensdatter (26 Aar) hos gårdmand Niels Møller i Junget, ved udlagt barnefader Ungkarl Henrik Iversen, tjenende sammesteds	uægte barn
1	14-5-1900	31-5-1900	Marie Andersen Møjbæk Møjbæk tilføjet i H. t. Bevilling af 28/7-05	Gmd. Niels Andersen (Møjbæk) og Hustru Hansine Nielsen af Junget 37 Aar	

G) PR Marriages 1813-1917

We link PR Marriages *forward* to the nearest following census. This strategy has been chosen in order to maximize the use of the identification potential of the household information inherent in the marriage record: we can look for a couple instead of two single persons. This approach is particularly important, as the Marriage records for a large part of the period only contain information on name, age and marriage parish – the household support thus dramatically increases the possibility for reliable linking.

As the examples from the original source below show, there is a gradually increasing amount of data available as time goes on. In the beginning, we generally have the following information to link on:

- Name
- Age
- Marriage parish
- Spouse's name & age (household info)

Gradually, the following will join the ones mentioned above:

- Birth place
- Birth day
- Parents' names (all)
- Same data for the spouse

After 1892 all this information will be generally available (except maybe in the largest cities) while before 1892 often only the first batch of information will be registered. Fewer than 15% of the records contain birth places before 1892, for example, but more than 80% from 1892 onwards.

No.	Brudgommens Navn, Alder, Bønder- tering og Opholdssted.	Brudens Navn, Alder, Bønder- tering og Opholdssted.	Hjem Forlovere	Døstid Dag.	Om forfatteren Styrken eller Hjælperens Høje Tidspunkt Døstidspunkt Dag.	For an- førelsen af den første Børn. Regist.	Bemærking.
				1821			
1	Mig. Hans Christen Nielsen af Junget	Mig. Anne Christen Nielsen af Junget	Hans Christen Nielsen af Junget			304	
				1822		35	
						36	
				1823			
1	Mig. Hans Christen Nielsen af Junget	Mig. Anne Christen Nielsen af Junget	Hans Christen Nielsen af Junget			305	
						27	
						28	

Married couples from Junget parish 1821-23, 1861-62, and 1900. The forms are more or less identical between the 1820s and the 1860s, but a bit more information has been entered in the latter records: for identification purposes the more exact registration of age might be helpful, and the same goes for the vaccination information – which has not been transcribed though. The 1900 record contains information on birth date & place and both parents' names – data which has all been transcribed. According to the post-1892 form this data was supposed to be entered according to the spouses' official birth certificates.

67		Copu. lerede.							
No.	Budgennens Navn, Alder, Gaashtering og Dydsstæd.	Budgens Navn, Alder, Gaashtering og Dydsstæd.	Daen Betsenst. etc.	Bettes-Dagen.	Da Bette er foretaget i Huset eller i Kirken, og i Huset i Huset.	Over skriftet i det alm. Huset.	Over skriftet i det alm. Huset.	Anmærkninger.	
1	Ungkarl Karl Jens Sørensen af Væjsmark, gammel 26 Aar, Vaccineret	Ung Pige Anne Katrine Jensdatter af Brokholm, gammel 30 Aar, Vaccineret						24-2-1823 i Kirken	
4	Ungkarl Rasmus Jensen af Junget, 28½ Aar, vaccineret 26de Maj 1836 af Tetens	Pigen Bodil Kirstine Villadsdatter af Junget 35 Aar gl.; vaccineret 1827 af Stengel						2-1-1862 i Kirken	
2	Peder Pedersen Nautrup, Ungkarl af Krejberg; født i Nautrup, den 3die December 1872; Søn af Gmd. Peder Pedersen Nautrup og Hustru Johanne Kristensen af Knud By, Rødding Sogn.	Kirstine Petersen, Pige af Skove, født i Skove Junget Sogn den 25de September 1881. Datter af Gmd. Niels Hans Peter Petersen og Hustru Johanne Jensen af Skove.						6-4-1900 i Junget	

Examples of information in the original source (from the above illustrations). Items in bold will NOT have been transcribed:

No.	Groom	Bride	Marriage date and place
1	Ung Karl Jens Sørensen af Væjsmark, gammel 26 Aar, Vaccineret	Ung Pige Anne Katrine Jensdatter af Brokholm, gammel 30 Aar, Vaccineret	24-2-1823 i Kirken
4	Ungkarl Rasmus Jensen af Junget, 28½ Aar, vaccineret 26de Maj 1836 af Tetens	Pigen Bodil Kirstine Villadsdatter af Junget 35 Aar gl.; vaccineret 1827 af Stengel	2-1-1862 i Kirken
2	Peder Pedersen Nautrup, Ungkarl af Krejberg; født i Nautrup, den 3die December 1872; Søn af Gmd. Peder Pedersen Nautrup og Hustru Johanne Kristensen af Knud By, Rødding Sogn.	Kirstine Petersen, Pige af Skove, født i Skove Junget Sogn den 25de September 1881. Datter af Gmd. Niels Hans Peter Petersen og Hustru Johanne Jensen af Skove.	6-4-1900 i Junget

H) PR Burials 1813-1917

The burial records are generally the vital events from the PRs with least information.

Often there is information on the deceased person alone, whereas Baptisms and Marriages always include more than one associated person (e.g. mother, spouse). This characteristic of the Burials applies especially to adult males: children will often have information about their father/parents, and women will often have information on their husbands (if married/widowed). Adult males, unfortunately, for a large part of the period, only have information on name, age and residence parish – marital status, occupation and address have not been transcribed.

As the examples from the original source below show, there is a gradually increasing amount of data available as time goes on. In the beginning, we generally have the following information to link on:

- Name
- Age
- Burial parish
- Husband's name (if a married/widowed woman)
- Father's/parents' names (if a child)

Gradually, the following will join the ones mentioned above:

- Birth place
- Birth day
- Wife's name (if a married/widowed man)
- Parents' names (all)

After 1892 all this information will be generally available (except maybe in the largest cities) while before 1892 often only the first batch of information will be registered.

40

Døde. Mandkøn.

No.	Dødsda- gen.	Begravet- stedet.	Den Dødes Fø- dselsdag og Tidspunkt.	Stad, Hædersting og Ejendomsforhold.	Alder. Antal Aar.	Antal Dage.	Bemærkninger.
1823							
2	17 ^{de} Janr.	22 ^{de} Janr.	22 ^{de} Janr.	22 ^{de} Janr.	44	305	1831
3	19 ^{de} Sept.	22 ^{de} Sept.	22 ^{de} Sept.	22 ^{de} Sept.	25	305	1832
4	19 ^{de} Oct.	22 ^{de} Oct.	22 ^{de} Oct.	22 ^{de} Oct.	23	305	1833

Buried men from Junget parish 1823, 1863, and 1900-1901. There is no difference between the registration in 1823 or 1863, even if the registration form now says that the deceased's birth place and father's name ought to be registered – if obtainable. But the registration of 1900 contains information on birth place, parents' names and spouse's name – data which has all been transcribed.

144.

Døde. Mandkøn.

No.	Dødsda- gen.	Begravet- stedet.	Den Dødes Fø- dselsdag og Tidspunkt.	Den Dødes Stad, Hædersting og Ejendomsforhold samt hans Fæderes eller, hvis han er ugift, hans Moders Navn, forsaavidt disse ikke afsløjtes have været.	Alder. Antal Aar.	Antal Dage.	Bemærkninger.
2	16 ^{de} Janr.	22 ^{de} Janr.	22 ^{de} Janr.	22 ^{de} Janr.	84	107	71
7	16 ^{de} Janr.	22 ^{de} Janr.	22 ^{de} Janr.	22 ^{de} Janr.	84	107	71
8	24 ^{de} Febr.	29 ^{de} Febr.	29 ^{de} Febr.	29 ^{de} Febr.	84	107	71
15	16 ^{de} Apr.	22 ^{de} Apr.	22 ^{de} Apr.	22 ^{de} Apr.	84	107	71
6	24 ^{de} Febr.	29 ^{de} Febr.	29 ^{de} Febr.	29 ^{de} Febr.	84	107	71

Døde				Mandkøn.				223
No.	Dødsdagen (Aar og Dato) og Dødsstedet (Paa Landet: Byen, Sognet og Herredet. I Købstad: Købstaden og tillige Sognet, hvis der i Byen er mere end 1 Sogn).	Begravelsesstedet (Kirkegaardens, Sognets og Herredets (Købstadens) Navn) og Begravelsesdagen (Aar og Dato).	Den afdødes fulde Navn (som vidt muligt efter hverst Brude- eller Fødselsattest eller ved Efterlysning i Hids- bogen).	Den afdødes Bølling og Bopæl (a) sidste fælle Opholdssted, hanset paa Landet ved Byen, Sognet og Herredet, i Køb- stad ved Købstaden, Sognet og Herredets navn; her end vidt Oplysning: hvem han erholdes, tillige den afdødes Fødsels- og Fødselsdato (ved søgte fælle Mødernes) fulde Navn, Bølling og Bopæl. Hvis den afdøde var eller vidste at have været gift, op- føres tillige dette samt, som vidt muligt, Afdødsfællens fulde Navn og dens sidste fælles Bopæl.	Alder. (Nær den afdøde var over 1 Aar, angives Alderen med det fulde Aar.)	Af hvilken Præst Begravelsen er foretaget.	Anmærkninger.	
2	21 August 1900 Geddeløse By, Junget Sogn i Nord Herred	Thorvald Thorsgaard 1784 27 August	Lambert Nielsen (Lambert)	Endnu ingen af Afdødsfællens; født i Jylland gift med Ane Jacobsdatter fulde fælles Bopæl i Thorsgaard.	90 Aar	Sognepræst	Langfeldt	
1	1901 23 Marts Junget By og Sogn i Nord Herred	Junget Thorsgaard 1901 25 Marts	Hans Kristian Jensen	Indsiddet i Junget; født i Gassum, i Thorsgaard Herred; Søn af Hmd. Hans Jensen og Hustru Ane Kristensen af Gassum. Gift med Ane Kirstine Svenningsen. Sidste fælles Bopæl Junget.	47 Aar	Sognepræst	Langfeldt	

Examples of information in the original source (from the above illustrations). Items in bold will NOT have been transcribed:

No.	Death date	Bur. date	Name	Occupation and family	Age
2	17-1-1823	26-1-1823	Rasmus Erichsen	Huus og Muurmester i Junget	44 Aar
5	5-4-1863	12-4-1863	Kristen Knudsen	Fattiglem, i Kost i Nørretise	75 År
1	21-3-1901 Junget by	28-3-1901	Hans Kristian Jensen	Indsiddet i Junget. Født i Gassum Nørhald Herred; Søn af Hmd. Hans Kr. Jensen og Hustru Abelone Kristensen af Gassum. Gift med Ane Kirstine Svenningsen. Sidste fælles Bopæl Junget.	47 Aar

2. Cleaning & preparing

The data has been prepared in a way that makes sense for both Link-Lives and ALA linkers. The PR data was transcribed in Asia by a company paid by Ancestry, and we must expect that a set of instructions was given to the transcribers directing them on which field to record each piece of information in. As with any historical source, however, be aware that errors, omissions and subjective decisions are inevitable at each stage. For example, for almost all fields we find misplaced or misread information from other fields.

3. Generating potential links

Potential links are generated for each record entry (where possible) by running a rule-based algorithm based on gender and age. The most likely 10 are then displayed on the right side of the ALA screen in order of probability score (0 is the closest match). These act as start-point suggestions for linkers. We might find the correct link within this list, or we might not, but we always attempt to challenge them.

4. Purposes of linking decisions

This table shows the consequences of our linking decisions.

Decision	Purpose/function
Link	<ul style="list-style-type: none"> • Trains the machine that this is a secure link. • Data is used to train the algorithm. • Helps to build life courses. • The positive decisions (ie. maybe, link etc) are usually merged for Machine Learning purposes.
Maybe	<ul style="list-style-type: none"> • Offers the ML team evidence of a link that is missing something. They are then able to identify what that missing information is and use that as a dataset in itself. • The positive decisions are usually merged for ML purposes.
+Secondary can only be selected once you have clicked Maybe, Multiple, Not Found	<ul style="list-style-type: none"> • Because the 'accepted' approach to linking is using only unchanging variables, BRE would like to capture all instances where we use OTHER information in the source to make a link.
Multiple	<ul style="list-style-type: none"> • Trains the machine that there is more than one plausible candidate based on the information available. They then assess what data is missing that is preventing a "link". • Conflict solvers use this to help them investigate a finite potential pool of candidates.
Not found	<ul style="list-style-type: none"> • Trains the machine that there is no link for this candidate based on the information available. <p>Even though it feels like a negative decision, rest assured that there is a good chance that we'll find this person later on and link him/her based on adding information from other sources. You do not need to mark things as "multiple" or "maybe" to show you have made an effort ☺.</p>
Unborn	<ul style="list-style-type: none"> • These are excluded from our training data.
OBS	<ul style="list-style-type: none"> • We analyse this dataset later using more advanced techniques, so all we need to do is flag records for this purpose. • We link on a one-to-one basis, so if there is more or less than one person in a record it cannot be included in our standard training data.

A) Examples

Here are some examples of how to use the different linking decisions in different circumstances.

Source 1 (linking from)	Source 2 (linking to)	Reasoning	Decision
Maren Henriette Poulsen (wife of Rasmus Poulsen) b. 1850 in Græsted.	Maren Henriette Ferdinandsdatter, tjenestepige b. 1850 in Græsted.	No other Maren Henriettes in the parish or amt but with no household support there's a small chance there is more than one Maren Henriette of that age born in the parish.	Maybe
Hans Krøll's wife Christine Krøll, b 1820 Krønge.	Hans Krøll's wife Kirstine Jensdatter, b 1824 Krønge.	We cannot be sure whether he remarried, because we don't have her birth name and the age is slightly differing. We take "Maybe" to flag the doubt, but we're not REALLY worried.	Maybe
Laura Nielsen født Petersen in 1863 in Gilleleje.	There are two Laur* Pe*ers* in Gilleleje: Laure Jensine Kristine Pedersen and Laura Johanne Marie Pedersen, both 2 years out. Another Laura Petersen is the same age but born in the neighbouring parish.	There is a small set of identifiable candidates and we are relatively sure the right link is one of them, but we are missing household information and further middle name identifiers.	Multiple
Petrine Petersen, wife of Laurits Petersen, b. 1861 Gilleleje. The source is also Gilleleje, so it's likely she would be there 20 years before but with her maiden name.	There are 2 Petrine names registered in the parish within her age range and a further 3 Petronella variations.	A lack of information therefore stops us from deciding on one candidate, even though the name is quite unusual.	Not found
Peter Nielsen b 1853 Græsted.	11 people are returned on a search for name, age and Græsted as birth place, some with multiple given names, but no occupation or household support to distinguish between them.	A number of people are identifiable but a lack of further distinguishing information stops us from narrowing the pool to less than 6	Not found
Robert Anders Ludviggo Lund (distinctive name).	Cannot be found despite multiple versions of his name/age/birth place searches.	This is not lack of information but some anomaly in the transcription or source (e.g. he was out of the country, not enumerated, not transcribed, source illegible etc.)	Not found
Marie Hansen (married to Johan	We cannot know what her maiden name is, even though we have a definite place of	We lack information on her surname to even begin to be able to search for her.	Not found

Hansen) b. 1875 in Stege living in Sorø.	birth, because her name is too common.		
Peter Hansen b. 1875 in Årre, tjenestekarl.	There are 18 Peter Hansens in Ribe but none born in Årre. Several are born in neighbouring parishes but none in 1874-6	With no distinguishing characteristics (e.g. unusual name(s)), and no household support, it is not possible to reduce the pool of candidates.	Not found
Burials examples to follow.			

B) Census illustration

The screen grabs below show Johan Jensen's household in the urban area of Odense in 1850 and in 1845. Note that

- The wife's birth place is not registered as a parish in 1850 (but as a place which is situated in a rural parish outside Odense)
- The wife's birth name is seen only in 1850
- The children have no last names in 1845
- The children's (and the wife's) multiple first names are thoroughly registered in 1845, but not in 1850
- The daughter Elise Kirstine from 1836 is registered with just the diminutive "Lise" in 1850

Linking 1850 to 1845 odense købstad, odense, odense 1850					
Id	Link-id	Name	Birth place	Age / Byear / M.st. / L.i.i.	HH pos.
518733	592708	Johan Jensen	Odense	39 1811 Gift (2023)	smedemester, huusfader
518734	592709	Johanne fød Henrichsen	Stige	45 1805 Gift (2024)	hans kone
518622	592710	Sofie Jensen	Odense	16 1834 Ugift (2025)	deres børn
518623	592711	Lise Jensen	do [Odense]	14 1836 Ugift (2026)	deres børn
518629	592712	Christian Jensen	do [Odense]	12 1838 Ugift (2027)	deres børn
518630	592713	Peder Jensen	do [Odense]	10 1840 Ugift (2028)	deres børn
537858	592714	Viggo Jensen	do [Odense]	8 1842 Ugift (2029)	deres børn
540476	-1	Carl Jensen	do [Odense]	1 1849 Ugift (2030)	deres børn

1845					
Id	Link-id	Name	Birth place	Age / Byear / M.st.	HH pos.
592708	518733	Johan Jensen	Odense	35 1810 Gift (2506)	Smedemester
592709	518734	Johanne Marie Jensen	Stige i Lumbye sogn	40 1805 Gift (2507)	Hans kone
592710	518622	Sophie Bartholdine	Odense	11 1834 Gift (2508)	Deres barn
592711	518623	Elise Kirstine	Odense	9 1836 - (2509)	Deres barn
592712	518629	Hans Christian	Odense	7 1838 - (2510)	Deres barn
592713	518630	Jens Peter Martin	Odense	4 1841 - (2511)	Deres barn
592714	537858	Johan Vigo	Odense	2 1843 - (2512)	Deres barn

Source-specific information

C) Census to census details

Birth places in censuses

Generally, you should expect the birth place to be exact for a good link. Some systematic errors have been identified though:

- the further away, the less precise the birth place: a person not born in the parish will have a less specific place of birth (maybe just the amt is listed or even just the country, e.g. Sweden) than someone local, for whom a more specific location can be provided. Also, the risk of the birth place registration being actually wrong probably increases with the distance.
- an oldest child was often born in another parish (before the parents settled down) but in the census is actually registered with the birth place where he/she actually grew up instead.
- parishes within the same “pastorat” are more often confused with each other than others.
- the administrative levels of birth places can vary from census to census (e.g. farm, village, parish, town, county, country)
 - Bear it in mind as a possibility that a birth place is not always recorded as a parish name.
 - A sparse use of contemporary geographic search engines is recommended. Not for every search, but for cases where there is only one possible candidate but the place does not match. Check if the places are very close by, or if they were part of the same parish.
 - <https://www.danishfamilysearch.dk/>
 - <http://www.krabsen.dk/stednavnebase/>
 - Google maps
- Birth places are often recorded as variants of “her i sognet”, “samme sted” or “ditto”, i.e. text strings which do not include the actual name of the place.
 - Look for the person in the birth parish (as Residence parish) in order to account for potential matches with “her i sognet” as birth place.
 - Look for the person in the standardized birth parish (Parish Std) in order to account for potential matches which appear simply with ‘ditto’ marks.

i) Special characteristics for birth places in URBAN areas

In urban areas, birth places seem to be generally registered in a less precise manner – this is probably due to a greater mobility, causing:

1. the imprecision mentioned above (because of distance) and
2. a tendency to register birth *places* rather than birth *parishes*, probably because of the census enumerator’s lack of knowledge of (and interest in) exact parish names in the surrounding rural areas.

Names in censuses

Names change more often between census periods in the urban areas than in rural areas, a fact which should be taken into consideration when searching for names. For information in how to interpret and understand birth places in the censuses, see the Appendix.

- First names
 - multiple first names are very normal in urban areas, but not all census enumerators were that enthusiastic about registering all names of a person: Laura Concordia Kristine Jensen can become Laura or Concordia or Kristine or Laura Concordia etc. This also means that Laura Jensen (1860) and Concordia Jensen (1850) can be the

same person without having matching first names – and without any of the names being wrong!

- Diminutives are often registered, perhaps because of the generally greater fluctuality of names in urban areas: Laura Concordia Kristine might become Stine in the next census
- Watch out for common diminutives: Vilhelmine/Mine, Nielsine/Sine etc.
- First names can even be abbreviated: Laura Concordia Kirstine Jensen could become L. C. K. Jensen – or simply L. Jensen.
- Last names
 - children and married women are more often registered without last names than men
 - married women taking the husband's last name (Maren Jensdatter to Maren Lumby)
 - children changing from real to inherited patronyms (Niels Hansen's children from Nielsen to Hansen)
 - girls changing from female to male patronym (Hansdatter to Hansen)
 - families changing from patronyms to family names (Jensen to Lumby)
 - as the urban population more often carried special and sometimes foreign family names (Schultz, Suhr, von Snöbbelhöft) there is considerable room for different spellings and errors in transcription

A guide to conflict solving

Link-Lives

Version number: 1

Updated: Feb 2022

A guide to conflict solving

Based on AMNY's Guide to Conflict Solving (last edited 24 Aug 2021).

Find conflict

Trello:

Go to the linking trelloboard and look at the **'conflictfiles uploaded'** header. Find a linking-unit that:

- Is a unit not linked by you
- Has the tags for **'file1'** and **'file2'** on it.

Note down which linker is linker1 and which is linker2 (You are going to need this in a second.)

Put the pink tag 'Solving' with your name on it on the linking unit you are gonna solve.

Teams:

Now go to teams and conflict resolution and find the relevant folder for which source that has been linked.

Find the two linkers conflicts, download each of them, one at the time and open them.

Prepare conflictfile for solving

Linker 1's conflictfile:

First write **_L1** at the end of **'mere sandsynligt'** and **'kommentar'**

Then you copy-paste linker 2's **'mere sandsynligt'** and **'kommentar'** column into linker 1's document.

And write **_L2** on linker 2's columns.

Then you create 4 new columns:

linker1_linker2_review_agreement

mere sandsynlig_CS

type_CS

id2_CS

Linker1's conflictfile should end up looking like this:

	V	W	X	Y	Z	AA	AB	AC	AD	AE
	type_agreement	mere sandsynlig_L1	kommentar_L1	mere sandsynlig_L2	kommentar_L2	linker1 linker2 review agreement	mere sandsynlig_CS	type_CS	id2_CS	
04-link-multiple	Linker2		Der er to med samme navn, som	Linker2	There is both linker1s candidate and a 'w	SAND	Linker2			
21-not found-maybe	Linker2		jeg overvejede, at Jørgensen ku	Linker1	Even though the name and age match, th	FALSK	Linker2			
25-not found-secondary_maybe	None		jeg synes, der er for mange muli	None	This should be a not found+secondary be	SAND	CS	secondary_not found	174991	
27-not found-secondary_not_foun	Linker2		Kan godt se, at bopæl kan give e	Linker2	The secondary characteristics (address at	SAND	Linker2			
30-same-secondary_maybe-sec	None		None	None	There are 3 rasmus rasnussen within the	SAND	CS	secondary_multiple maybe	249930	
04-link-multiple	None		Skulle måske have skrevet mayb	Linker1	The last name seems to be written in an o	FALSK	CS		154950	
20-not found-link	Linker2		Kan ikke forstå, jeg har overset h	Linker2	The name is so unique that I feel confide	SAND	Linker2			
21-not found-maybe	Linker2		Der er ikke så mange gentlere	Linker2	I think the spelling variants of the last na	SAND	Linker2			
30-same-secondary_maybe-sec	None		Erneste lærer Ole Jørgensen	Linker1	There are more than 5 jørgen (jensens, wh	FALSK	Linker2			
62-secondary_maybe-not found	Linker1		Erneste lærer Ole Jørgensen	None	There are 6 ole (x) jørgensen within the ag	FALSK	CS	secondary_not found	288403	
10-different-maybe-link	Linker2		Mændene stilling	Linker2	My candidate is a month for first and last i	SAND	Linker2			

For **linker1_linker2_review_agreement**

In the first cell of "linker1_linker2_review_agreement" you write: = [click the first mere_sandsynlig_L1 cell] = [click the first mere_sansynlig_L2 cell] and press enter

This should give you either a SANDT or FALSK

Drag the cell down the rest of the column. This gives you SANDT or FALSK for the whole column.

All the SANDT you can immediately copy into "mere_sandsynlig_CS"

mere_sandsynlig_CS

Solving

When making a decision you read the comments of each linker, and review the information. If by reading these, you immediately know what the right decision is, then you can write which linker has the right decision in "**mere_sandsynlig_CS**"

If you need more information then use ALA.

If in doubt consult the best practice, and if still in doubt, write Atlanta or Olivia.

A rule of thumb is that if the doubt persists then go with the conservative option.

And:

If a linker has no comments then be extra careful to not be Immediately biased towards the linker that has left a lengthy comment.

Saving and uploading

Save it as the conflictfile, but delete the linkers name and instead write _Solved.

Upload it to teams solved conflict files.

Update trello: put right tag on (**conflict stats updated**) and move to the right.

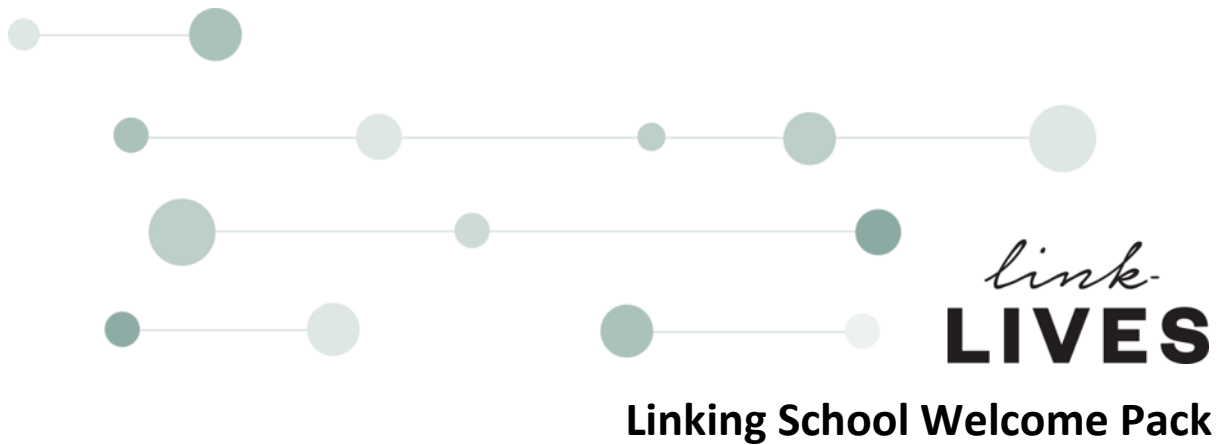
Update **masterlist_burials** up to the **date solved** part. (can be found in Teams under important documents → training data control)

Consistency check

It is vital that the CTD files are identical and have no inconsistencies in them. They might include mixed Upper and lower case text, misspellings (e.g. of linker1, linker 1, Linker1, L1 etc) etc.

Make sure you check anything you add to the file yourself is consistent.

Make sure you check that every decision line has one decision attached. There should be no blanks, not_filled etc.



Welcome to Link-Lives and Linking School! We hope this pack is useful.

Your name
Your email
Date sent 2 Feb 2021
ALA version 1.8
Contact person [Olivia Robinson](#)
User Guide v.8
Best Practices v.8
Linking units We will be working with the following units and date ranges:

Week 1 Parish Record Burials	Sources	Week 2 Parish Record Baptisms	Sources
<i>Viborg Amt, Selde Sogn</i>	1880->1860	<i>Viborg Amt, Junget Sogn, 1842-57</i>	1845->50
		<i>Maribo Amt, Krønge Sogn, 1835-91</i>	1850->60

Linking School plan

The purpose of Linking School is:

1. To experience linking using our bespoke software
2. To practise linking according to the project's Best Practice guidelines
3. To prepare for graduation into the production team.

Part 1 (Wed 2nd – Mon 7th Feb)

- Wed morning: You will be introduced to linking and ALA
- Wed afternoon: You will begin to link the first link unit
- Thur morning: You will join our Link-In with other linkers at KUA
- Thurs afternoon: You will continue/complete linking, then send Olivia your links file
- Mon afternoon: you will receive your conflict file to check and return that day; we will then meet briefly online at 15:30 to answer any burning questions.

Part 2 (Tues 8th – Mon 14th Feb)

- You will move on to linking the next two units and send your links by Thursday evening
- On Monday morning you will receive your conflict files to check and return that day
- On Tuesday afternoon we will meet to discuss your progress and answer questions.

Welcome to Link-Lives

Link-Lives is an academic research project collaboration between CPH University, Rigsarkivet and KBH Stadsarkivet. It is funded by Innovationsfonden and Carlsberg and runs until 2024. We are developing a database that will connect individual-level sources to eventually create life-courses for all people in Denmark since 1787. These are being built by linking data from transcribed censuses, parish records and burial registers. The sources have been transcribed by volunteers at Rigsarkivet, Københavns Stadsarkiv and by the genealogy firm Ancestry. More about the project, its aims and the people involved can be found on our website: link-lives.dk.

Linking overview

Linking is the backbone of our work. Our approach is that each linking unit is linked by at least two different linkers before being sent for comparison to a linking coordinator. The coordinator produces a conflict file containing all cases where there is disagreement between the two linkers (usually only 10-15% of links are contested in this way). After the linkers have conflict-checked their files and returned them, the linking coordinator takes on the role of 'conflict solver', examining the two sets of results and deciding which option is more likely according to the project's Best Practice guidance. The final decisions for each linking unit are saved as consolidated training data in a CTD file.

We create manual links connecting an individual between two sources (census, burials, marriages, baptisms etc) for three reasons:

1. To develop expertise in handling and linking different sources
2. To test the accuracy of our rule-based algorithms
3. To train the machine algorithms to automate future linking

In our Best Practices document (see below) you will get a fuller picture of how we link and what guidelines to follow. Broadly speaking, we are guided by these overarching commandments:

Linking Commandments

1. Only link one person to one person
2. Always challenge a potential link
3. Always search for a competing candidate
4. Always link using primary variables first, before assessing secondary variables
5. Never use unauthorized sources outside of ALA
6. When in doubt, do not link

Getting started with linking

We have developed a standalone software programme that enables us to link, called ALA. To get started with linking, follow these instructions:

1) Download the latest version of ALA

- a. Follow the link you have been given to the zip files containing ALA, to download and install on your computer (PC only).
- b. Make sure you follow the instructions in the User Guide and read the Best Practice files before you start linking.
- c. A set of [instructional videos](#) can be found online. They refer to an earlier version of ALA but many of the principles are helpful introductions.

2) Data linking

- a. All new starters on the project go through 'Linking School', where you go through a process of linking three linking units (usually parishes or parts of parishes). You will find your linking unit allocation at the top of this document.
- b. Once you've finished each of the parishes in Linking School, send your links.csv file [to Olivia via email](#).
- c. We will process your links and compare them to our existing information.
- d. You will be sent an Excel feedback file so that you can help us investigate the differences. This process will take around 1-2 hours more. We call this conflict checking.

Conflict checking

Once we have processed your links we invite you to review what we call 'conflicts' in an Excel 'conflict file'. A conflict is when your linking decision differs from that of another linker (e.g. one of you finds one link while the other finds multiple links). In Linking School, that 'linker' is actually a consolidation of all those who have linked these parishes, so we can be reasonably confident the decisions are in line with our Best Practice guidelines. (Once you graduate from Linking School, for each parish you link, you will be paired with another linker from the team. However, the conflict-checking process will be the same).

- a) For this task, we would like you to go over the cases in the Excel spreadsheet and see whether you stand by your own link decision (linker2) or whether you agree more with the amalgamated data (linker1).
- b) You will see linker1's decisions on the left side of the sheet (shaded blue) and yours on the right (shaded green).
- c) Use ALA to look up each case as necessary.
- d) Write 'linker1' or 'linker2' in the column headed 'mere_sandsynlig' according to your assessment. You can also write 'none' if you now believe neither is the correct decision.
- e) In the column headed 'kommentar', we ask you to write a short description as to why you made that choice (see below).
- f) Return your file [to Olivia via email](#).

1) Commenting

Our conflict solver takes the final decision on contested links – those where both linkers disagree. All information provided in the comments section helps, but the following are most useful:

1. If the place of birth is a place name in the parish we are looking for, please write that. It saves a lot of time.
2. How did you search for your result? If you did something special to find it, which you might think the other linker and the solver wouldn't think of, share it!
3. If you have a link and the other person says it's a multiple, and you decided that you still think your link is the best, then you have to say in the comments why you still think it's a link.
4. If you cannot decide, please write that in the comments and why.

Key outcomes of the conflict-checking stage:

- a) An opportunity for you to see how closely you are aligned to our Best Practice approach
- b) To practise the process of conflict-checking for when you graduate from Linking School and join the production team.

Two important reminders during conflict-checking

1. You should not change your links in ALA. They must stay as you first linked them.
2. In many cases, there is no right or wrong answer. The conflict-checking stage is a key part of the Link-Lives methodology that not only reduces the time it takes a conflict-solver to resolve persistent conflicts (the next step), but offers all of us a chance to look a little more carefully at the decisions we've made and to improve with each parish.

Graduating from Linking School

When you have completed your allocated linking units, you will meet to discuss the process with Olivia. If everything has gone smoothly, you will then join the production team as a junior linker. You will have an opportunity to raise any issues or concerns about the process before continuing.

User Guide

This gives instructions on how to download, install and operate ALA. It can be found attached to your welcome email.

Best Practices document

This gives instructions on how to assess the sources and link in a way that is consistent with the linking team. It can be found attached to your welcome email.

FAQs

- **I can't find my initials in the users screen.**

Navigate to ALA/data folder and find users.xlsx. Open in Excel, add your initials and numbers to the list, then save and close.

How long should I link at a time?

It is important not to sit for too long while linking, both for your back and for you brain! We suggest not more than 1-2 hours at a time.

Am I taking too long over each link?

An important part of Linking School is to train your brain to make consistent decisions in line with the Best Practice guidelines for the moment, so the time each one takes is less important at this stage. Once you 'graduate' from Linking School, we tend to find people link approx. 20-25 links per half hour. This varies, however, depending on the time distance between sources 1 and 2.

Add your own questions here to be answered by either Olivia or others in the linking team:

-

-

-

-

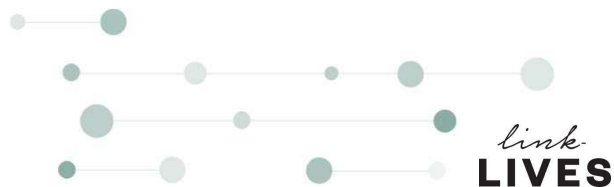
-

-

Mini Linking Workshop

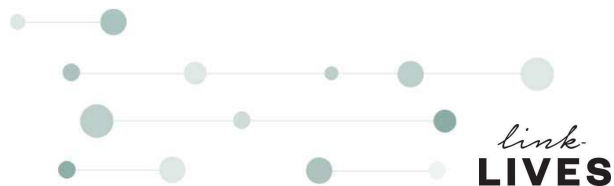


Zoom, 20th Nov 2020



Purposes

1. Align linkers & linking within the LL method
2. Improve and co-create LL's best practices



Structure of the day

- Background context (morning)
 - How ALA generates potential links
 - How linking decisions are defined in LL + their purpose
- Conflict discussions (morning + afternoon)
 - Gothersgade 1901->1880 link sample
 - Introduction to the conflict
 - About a minute to refresh our memories
 - When you're ready, press the decision on the poll
 - Unrecorded – illustrative purposes only
 - Reach a consensus on each conflict & record it
- Addressing questions / themes (afternoon)

link
LIVES

Outputs

1. FAQs document
2. Conflict-solved set of Gothersgade
(using consensus method)
3. Review of linking definitions

link
LIVES

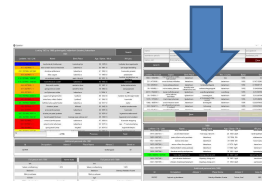
Housekeeping

- No question is a bad question – no suggestion is a bad one
- Keep the chat window just for questions - no parallel chats during discussions
- To keep discussion flowing we won't crowd manage – we'll review if it gets to chaotic
- Screenshots of our workshop for later publicizing – do you agree?
- Taking a break? – turn off your video

link
LIVES

Potential links

- Should always be challenged!
- Why?
- Construction of potential links:
 - Blocking: Age (+/-2) and Gender (and maybe some name and birth place junk)
 - Score: combined *distance score* of Birth Parish & County, First Name, Patronym & Family Name (max 0.5, max 10 potentials)



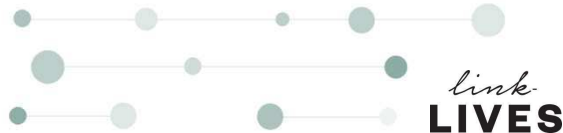
The image shows a screenshot of a data table with multiple columns and rows. A large blue arrow points downwards towards the table. The table appears to contain various data points, possibly related to the 'Potential links' section.

link
LIVES

Linking decisions

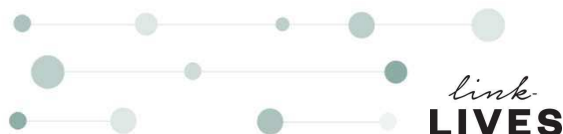
User Guide:

1	Link	Links	YES
2	Maybe	Links, but marks this link as somewhat insecure: Used when the linker is not completely sure – but thinks the link is really probable. In some instances, "maybe" would be considered as candidates for links.	
3	Occupation	Registers a link which –without considering the occupational info – would have been a "Maybe", but the information about the person's occupation makes the link sure	
4	Not found	Registers that the person has not been found in Source 2 (no plausible match found) - even though nothing excludes the possibility of him/her actually being there.	NO
5	Multiple	Registers that the person has not been linked to a person in Source 2 because there are more than one plausible candidate	
6	Unborn	Registers that the person is not possible to link to Source 2 (because he/she was not born yet at the time of Source 2).	



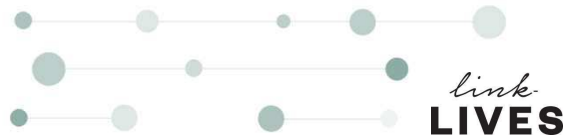
Afternoon session

- Demo of linking
- Which external sources can we use?
 - All those which give general context not specific detail. Places, for example:
 - <http://www.krabsen.dk/stednavnebase/>
 - <https://www.danishfamilysearch.dk/>



Linking decisions

- Could we consider to skip the "maybe" option?
- Attempt to define the difference between 'multiple' and 'not found' – or does it not matter?
- Does the 'multiple links' option change how we use 'not found'?
- Should 'multiple links' be changed to something like 'multiple possibilities'?
- How can we best use the reverse search function?



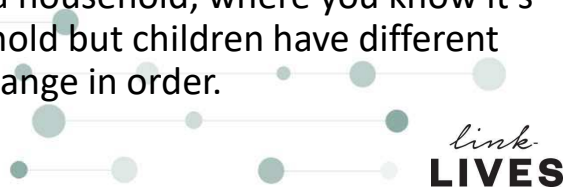
Multiples/not founds

- If there are candidates in neighbouring hamlets, but age/name is exact, how do we decide?
E.g. a lot of hits on the same name in a county but where the parish does not come up. Which to choose?
- -"multiple links" because there are several matches with the name in the county and no-one in the parish
- -"not found" because the parish is not the same but the names are right
- -"maybe" in the few cases I have a hunch that one of the choices is better than the others



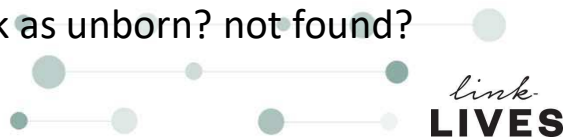
Names

- When women's surnames change, if name and place of birth are distinctive enough, can we disregard the surnames?
- What is "normal" name practice for widows' surnames? do they revert to maiden names generally?
- Experience and best practice in linking situation with children in a household, where you know it's the same household but children have different ages and they change in order.



Places

- A lot of the residents in Gothersgade don't write the specific parish they come from, but only jutland or sjælland (or slesvig). How do we search?
- it would be nice to have place names combined with the parish put into the computer so that more links are matched.
- Ubeboet = mark as unborn? not found?



If there's time

- Mapping of linker numbers increasing link rates (to be explained)

