Interoperating civil registration of death and census data: old age and marriage as categories of analysis

Interoperating Civil Registration of Death and Census Data: Old Age and Marriage as Categories of Analysis

Ciara Breathnach[1(✉)], Rachel Murphy[2], Alexander Schieweck[3], and Tiziana Margaria[3]

[1] School of History, University College Cork, Cork, Ireland
Ciara.breathnach@ucc.ie
[2] School of History and Geography, University of Limerick, Limerick, Ireland
[3] HRI, Lero and CRT-AI, University of Limerick, Limerick, Ireland

Abstract. As part of the project Death and Burial Data: Ireland 1864–1922 (DBDIrl), a web application was created in DIME, a low-code web application development environment. DIME is based on the popular IDE Eclipse and utilizes three distinct graphical model types (data model, process model and GUI model) as Domain Specific Language (DSL). Web applications are defined by means of these models in the DIME integrated modelling environment in a simple way, and these models are used to generate the web application code that is then deployed and made available to the users. This paper provides an overview of how we use a web application created in DIME, a low-code application development tool to establish interoperability between Irish historical civil registration of death data and Irish cross-sectional decennial census data from 1901 and 1911. The former data collections are complex and require manual interventions like transcription and considerable cleaning to render them machine readable, the latter are fully transcribed. Here we present a case study concerning the analysis of deaths in old age in Dublin City and how they correlate with census returns. We then turn our attention to a use case of marital status and the algorithmic problems caused by irregular replies in census data. Our objective was to test how DIME could replicate the decisions taken by domain experts using old age and marital status as the primary linkage criteria.

Keywords: Data matching · civil registration · census · low-code/no-code application development · model driven development

1 Introduction

Old age offers a very fruitful case study from the perspectives of life course studies and historical demography. We examine this in the context of Ireland in the early twentieth century, which was subject to British administration from 1801 until 1922. At the close of the nineteenth century increased life expectancy caused a shift in demographic trends, which necessitated the introduction of new social welfare schemes like the old age

B. Steffen (Ed.): AISoLA 2023, LNCS 14129, pp. 65–83, 2025.
https://doi.org/10.1007/978-3-031-73741-1_5

pension (OAP). The target cohort were lower-paid working-class people who did not earn enough during their working lives to pay into contributory pension schemes or to save money to support themselves in old age. Bismarck's Germany was the first to introduce a non-contributory pension in 1891 as part of a wave of social welfare reforms, and after extensive investigation and debate, the United Kingdom followed suit [1]. The parameters of old age and its positioning in historiography have been tentative as Johnson argues, and broadly distributed across the themes of labour market participation, welfare and social status [2]. Pat Thane contends that women fared better than men in reaching older age in western societies, but she cautions that the boundaries of older age must be rooted in various cultural constructs and historical contexts [3]. Lines were clearly drawn in the British Old Age Pensions Act of 1908, which stipulated that applicants had to prove that they were 70 years of age [4]. Apart from the chronological markers, applicants had to fulfil two further main criteria: 1) being resident in the United Kingdom of Great Britain and Ireland (as it then was) and a British subject for the previous 20 years, and 2) they had to pass a means test proving that their annual income did not exceed £31 10 shillings. Ó Gráda argues that, apart from its significant impact on national income, the effects of the OAP on household economics placed it as 'the most radical and far-reaching piece of welfare legislation enacted in Ireland in the twentieth century' [5]. From his assessment of post-Famine census reports Gilleard surmises that Ireland had the oldest population in Europe, which he attributes to 'cohort depletion from mass emigration' [1]. Dublin city, like Belfast city, had a more youthful population for two reasons: firstly, mortality rates in younger cohorts (especially infants and children under 5) were high and, secondly, it was where internal migrants gravitated. In rural areas of the West of Ireland traditions of seasonal and permanent migrations were firmly established and cohort depletion led to a larger older populace. Growing old in the city was a tough station and while all sorts of allegations were made about age heaping (rounding up to the nearest 5 or 0) in the 1911 census, it seems that in Dublin it was more difficult to get away with what Ó Gráda terms 'welfare fraud' as civil and religious records were more carefully maintained [5]. The impact of the OAP on Irish fiscal and social life cannot be understated. Family incomes were so low that a steady and reliable weekly flow of 5 shillings for one and 7*s* 6*d* for married couples would have been an inestimable boon to both sole-occupancy and multi-generational households at a time when a week's rental in Dublin could have cost up to 3 shillings a week [6].

Until the 1908 Act, 65 was used as the marker for old age for the purposes of the annual reports of the Registrar-General. In 1909, the year in which the old age pension was introduced, civil registration data shows that deaths from old age amounted to 8,870, which included those aged 65–69. Within the pre-established parameters of ages 65 to 74, 839 people died, with a further 5,184 aged over 75, 2,424 aged over 85 and 423 aged 95 and upwards [7]. Historical demographers are interested in understanding the impact of the OAP on these cohorts, and often use a process called data matching to track trends over time. Matching can occur between different record types, but census matching, in which individuals are located in the previous or subsequent census, is one of the most common forms of data matching in the U.S. through IPUMS [8] and in North-West Europe [9].

Apart from the algorithms used by IPUMS and Mandemaker et al. for census returns, there are recognized basic standards for data-matching across all data types. Christensen describes five steps involved in data matching: (i) data pre-processing, (ii) indexing, (iii) comparisons, (iv) record pair classification, and (v) evaluation. Historians and genealogists working with partially indexed data routinely must conduct this record-matching process manually [10].

This aspect of the project had two aims: (i) to take civil registration of death records and match them to census records using a bespoke web application designed in the DIME low-code application development environment[1] and (ii) to ascertain the extent to which it is possible to replicate the historian's thought processes using an algorithmic approach.

We had previously created a bespoke web application with this technology stack to transcribe scanned historical death data [11]. A logical next step was to extend the app to census data using this project as a pilot. While other data-matching solutions exist, they do not align with the full range of Irish records we are analysing both now and in the future. While aiming to replicate the manual process and supporting future work, the matching is limited to some fuzziness, e.g. in names or dates, but no full probabilistic matching at the moment. Also AI is not yet included, as for the moment we are collecting the kind of questions historians would pose, and evaluating the extent to which the approximate matchings we produce without AI are good solutions.

In the remainder of this paper we discuss how we set out to achieve these aims. In Sect. 2 we discuss the datasets we used; Sect. 3 describes the cohorts selected for the case study; in Sect. 4 we introduce the data models and elements of the application designed to conduct the data matching; Sect. 5 discusses the matching algorithm; Sect. 6 reports an analysis of our results, followed by our findings in Sect. 7. We draw our conclusions in Sect. 8.

2 Datasets

Some countries can match civil registration records and census data over a long period of time, e.g. England from 1837 for civil registration and census from 1841 to 1921. For the case study presented in this paper we match civil death records to the 1901 and 1911 censuses of Ireland, which are the only fully extant Irish censuses (see Sect. 2.2). Both datasets are described below. We have refined the metadata structure successively, as this structure and its granularity are at the core of the information on which linkage, interoperability and thus matches are based. We explained and analysed the original data structure in relation to civil registration of death data in [11]. Subsequently we refined the granularity further to arrive at the current, stable level of disaggregation described in [12]. As the census data have been transcribed and indexed in full by the National Archives of Ireland (NAI), it did not necessitate the same levels of attention.

2.1 Irish Civil Registration Records

Irish historical civil registration data (births, marriages and deaths) are openly available for individual searches on irishgenealogy.ie. When these datasets were first digitised, a

[1] DIME is an open-source platform available at https://gitlab.com/scce/dime.

simple index of a subset of information (place, name and date of death) was created and linked to a TIFF file of the original image. Accordingly, the data are not provided in full machine-readable formats nor is there any national infrastructure to support large scale crowdsourced transcription such as Pedersen et al. describe in their contribution to this volume [13]. A further issue is that ownership of each dataset comes under the remit of various government agencies: for example, Irish civil registration data falls under the auspices of the General Registration Office (GRO), whose primary function is to manage the live population register. The historical census comes under the auspices of the (NAI) and the more modern returns under the Central Statistics Office (CSO).

2.2 Irish Censuses

While Irish civil registration records are intact, when it comes to the census (which was taken decennially from 1821 to 1911) Ireland has the unfortunate disposition that some manuscript census returns were legally destroyed by order of the Deputy Keeper of the Public Records in the late nineteenth century, some were pulped for reasons of paper conservation during the First World War effort, and others were a casualty of the Civil War [14]. What survive for the entire country are the full manuscript returns for 1901 and 1911. Owing to the very disturbed state of the country in 1921, no census was held, as martial law prevailed in several counties during the War of Independence, and it was simply unsafe for the Royal Irish Constabulary or the Dublin Metropolitan Police to conduct the work. The next census, which took place in 1926, will be published in accordance with legislation in January 2027 [14].

3 Case Study: Old Age in Dublin South City No. 3 Ward

3.1 Death Data Used in the Study

We manually transcribed deaths registered in the population over 70 years old for one Dublin ward, South City No. 3 (Fig. 1) in 1911. This year was selected as it was the year of the first census taken following the OAP act. We conducted a data matching exercise with these civil registration of death records and the census returns. Working with the GRO death data provided us with insights into the relatively understudied cohort of older individuals. We took a random sample from this dataset to conduct a manual data matching exercise that in turn assisted our understanding of how best to devise a matching algorithm for this use case. A further aim for the historians was to add to the discourses on the questions of age misstatement in pension applications, which until now have relied on aggregate returns.

Dublin South City No. 3, located to the south of the river Liffey, was a predominantly working-class area, with many individuals working precariously as low-paid labourers or servants. Table 1 shows the recorded deaths of people aged 70+ in 1911 by 5-year cohorts. This data was derived from individual level civil registration of death registers of that year, which were fully transcribed to a machine-analyzable format by the first application we developed with DIME for the history research domain [12]. Analysis of this data shows that this age cohort made up 13% of total deaths in South City No. 3 in 1911.

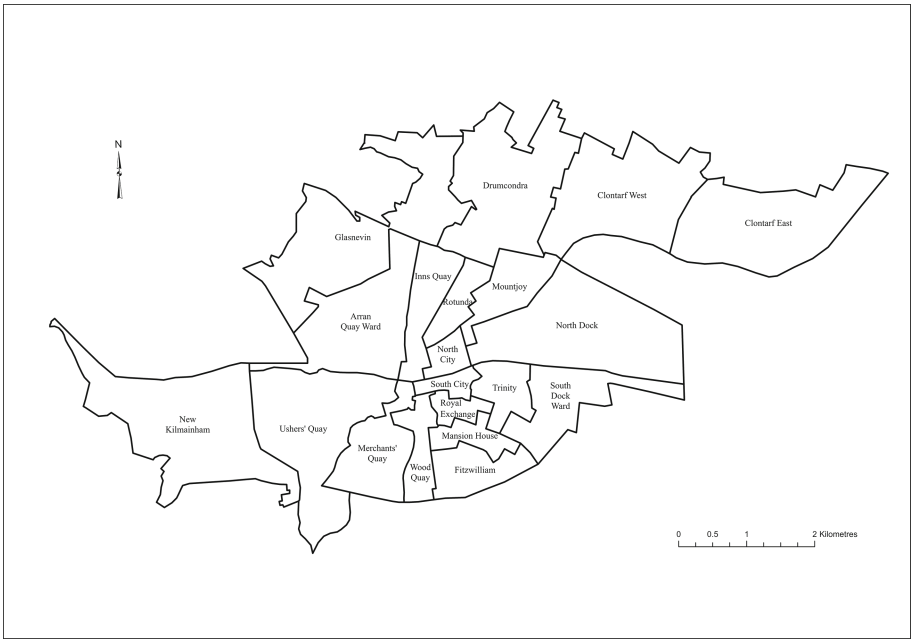

Fig. 1. Map Showing Dublin Wards. (Dublin, North and South Wards, Dublin Metropolitan Police stations shown as points, c. 1900. Map based on information derived from the Dublin Metropolitan Police Committee of Inquiry 1883 [C.3576 C.3576-I], p. 25, and Statistical Tables of the Dublin Metropolitan Police for the Year 1901 [Cd. 1166], p. 33, drawn using OSi historic 25-inch basemap © Tailte Éireann/Government of Ireland Copyright Permit No. MP 003824).

Table 1. Deaths of Individuals aged 70+ in South City No. 3, 1911 (based on analysis of GRO death data, South City No. 3, 1911)

Age Cohort	Deaths	Of which Male	Of which Female
70–74	57	23	34
75–79	40	23	17
80–84	22	8	14
85–89	10	2	8
90–94	2	1	1
Total deaths 70+	131	57	74
All deaths in South City No. 3, 1911	998	512	486

As Table 1 shows, few deaths occurred in older age, which forms the basis of our use case. A total of 131 deaths were recorded in the 70 and over category and, in keeping with Pat Thane's [3] assertions, of these 74 (or 56%) were women. Age, conjugal status and socio-economic status are among the primary features of the profile of these pensioners.

The marital status of this group was as follows: 78 were widowed, 25 were returned as married, 15 were spinsters, 11 bachelors, and 2 unknown.

To perform the matching exercise, we had to first identify all those individuals who were alive on the night of the census, 2 April 1911. This meant removing 46 individuals from the total of 131, leaving a sample size of 84 individuals. The breakdown of these individuals is shown in Table 2.

Table 2. Deaths of Individuals aged 70+ in South City No. 3 registered 3 Apr.-31 Dec.1911

Age Cohort	Deaths	Of which Male	Of which Female
70–74	41	16	25
75–79	24	13	11
80–84	11	4	7
85–89	7	1	6
90–94	1	1	0
Total deaths 70+	84	35	49

4 The Census Application

4.1 Background

We follow here the same design methodology used in the initial positive experience with the DIME platform to design, implement and deploy the project application [12], which allowed users located anywhere in the world to transcribe death data records to a curated database at a fine granular level of 64 searchable and indexable fields, compared to the original 11 of the Victorian registers. The aim of using DIME in DBDIrl was to support the use of data from multiple siloed sources, such as the civil registration data and the census data, and to make them interoperable, this way enabling a greater matching potential. This project focuses on phase one of this: census matching.

The next phase of the project was to create a census application called CensusIRL, which enables the tracing of individuals through time using the census records. We decided to use DIME again. The DIME process, defined as a sketch through blueprints, matches entries between two censuses (see Fig. 2). A preliminary approach successfully tested the feasibility using DIME [15], yielding a fully modelled prototype of the application that served as proof of concept. The current, more advanced, census matching application has been tested so far on the two available Irish censuses, the 1901 and 1911 census, but it is built in such a way that in future it can be applied to the 1926 census as well, which is due to be released to the public in 2027.

4.2 The Extended Data Models

The NAI provided the project with the full census datasets which were ingested into the application. Each census data record contains the 23 data points detailed in Table 3. All

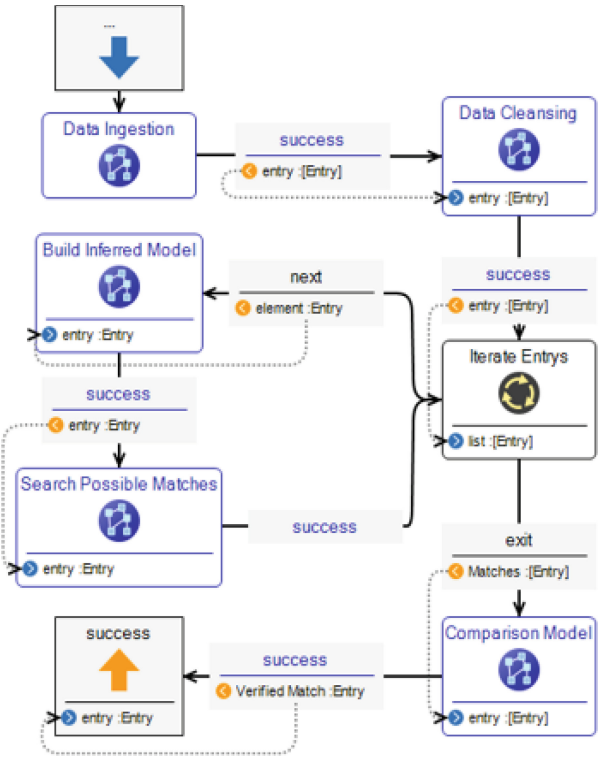

Fig. 2. Search process model sketched using DIME blueprints

census data is included in the scope of the project except crudely anonymised entries of people located in institutions such as workhouses, prisons or asylums. To avoid cases of reputational damage by association with criminality or the 'taint of pauperism' only initials were recorded in the original census returns.

An inferred data model, shown in Fig. 3, was manually defined to enable subsequent data matching based on that extended data model. During pre-processing, some analyses were run on the original data to compute inferred information that is relevant to determine whether a given census record is relevant or not for some matching query. The Entry table (Fig. 3, left) describes the original data entry, while the derived properties are gathered in a separate InferredInfo table (Fig. 3, middle) to enrich the available set of attributes for subsequent searches. For example, this information captures whether one or more middle names are included under the 'Forename' field, or whether an individual's name has been anonymised. This is relevant, as such entries should be excluded, preserving semantic correctness and completeness of the results, but saving computational effort, when trying to match by names. Here we also include Soundex, a phonetic algorithm that relies more on consonants than vowels to index words according to their phonetic pronunciation. Orthography is not always consistent between historical records, particularly where literacy rates varied, and vowel changes are common e.g. between Byrne, Burn and Bourne. In this case these three surnames are all signified by the Soundex code B650.

Table 3. Census Data Records Ingested into the application

1901 Census Household Return (Form A)	1911 Census Household Return (Form A)
1. Surname	
2. Forename	
3. Age	
4. Sex	
5. Relation to head	*As in 1901 but with three additional data points*
6. Religion	
7. Birthplace	
8. Occupation	
9. Literacy	
10. Irish language	13. Years married
11. Marital status	14. Children born
12. Specified illnesses	15. Children living
Enumerator's Abstract (Form N)	
16. County	
17. District Electoral Division (DED)	
18. Townland	
19. House number	
Additional Data	
20. Census year	
21. ID/Image group	
22. Images str	
23. Inferred model	

The *SearchingModel* table (Fig. 3, right) describes the current, fuzzy search and is used by the matching algorithm to find potential candidates in the census records. Some of the fields directly correspond to the original *Entry* elements, like *birthplace*, *firstname*, *surname* and *sex*, but many refer to derived information in the *InferredInfo*, like *canonicalFirstname*, *married* (that connects with the inferred hasMarriageInfo attribute), or exist only in the *InferredInfo* table like Soundex. Furthermore, only the *approximateBirthYear* is known due way in which the original data is structured, so we limit the window of years used to approximate matches. The *censusTarget* can be used to selected a specific census, currently 1901 or 1911.

In the model we also see that the *Entry* and the *InferredInfo* tables have mutual pointers, so are connected in 1:1 fashion, but they are kept separate. There are two main reasons for this choice:

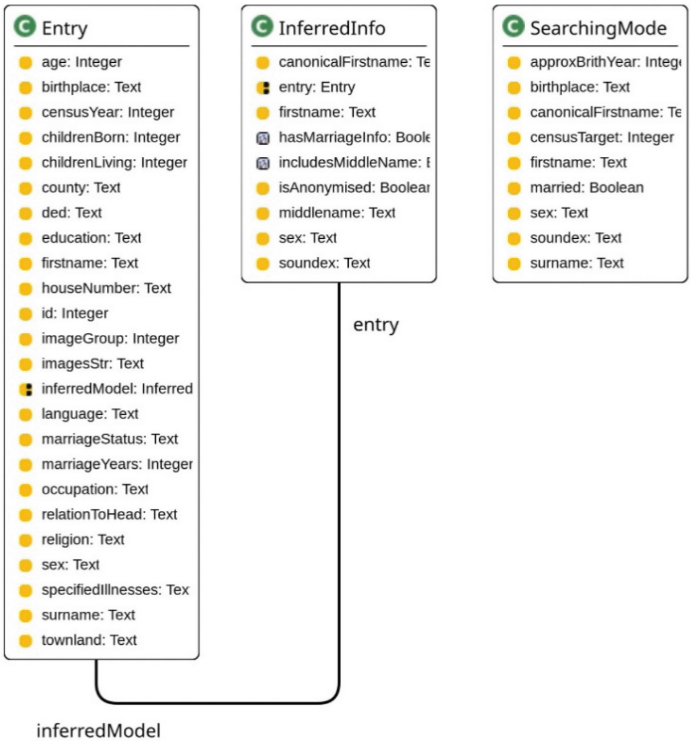

Fig. 3. The Data model used for the search: original Census Information (left), Inferred Information (middle) and the Searching Model (right) covering the fields that are currently used in the matching process.

- While the original census *Entry* structure is fixed, as the project progresses, we may wish to update and extend the nature and amount of *InferredInfo* made available for the searches.
- Additionally, it is curatorial best practice to keep the original data separate from any additional fields that may be useful for specific purposes but are not in the original format and dataset.

The data is stored in the extended format in a PostgreSQL database.

4.3 Embedded Processes for the Inferred Fields

Process models (see Fig. 4) describe the algorithms used to filter the census records in order to "infer" the additional information in the remaining two fields of the InferredInfo table: the hasMarriageInfo process (left) and the includesMiddleName process (right). While originally in the prototype these were fully modelled in DIME, the scalability of the queries has improved by just using database queries. As the data is stored in a standard PostgreSQL database, it is easy to formulate queries and embed them in rather simple DIME processes.

We are thus confident that this technique can be used again to provide in the future additional inferred fields, if needed. Given the similarity of these two processes, we may be able to provide a generic template process with the elements common to all the queries, that can be extended on demand just with a few Service Independent Building-blocks (SIBs) and data fields specific to the query in question.

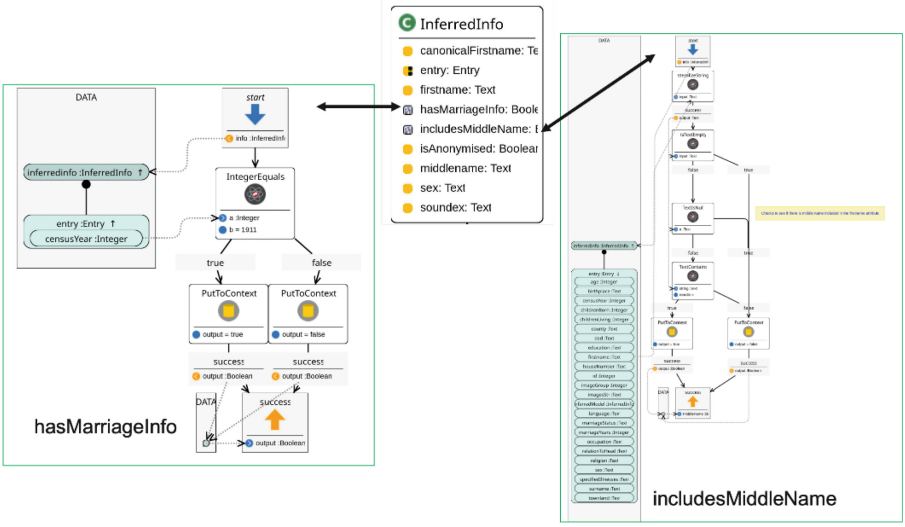

Fig. 4. The embedded process models used to compute inferred information: the hasMarriageInfo process (left) and the includesMiddleName process (right).

5 The Matching Algorithm

5.1 Matching Properties

The 84 deaths that occurred after 2 April 1911 were matched in the application against the entire census dataset. The matching occurs on the following six properties:

- **First name.** For some individuals the first name may be consistent between censuses, but in many cases the first name may vary because individuals have nicknames. For instance, a woman named Margaret could also be referred to as Madge, Mag, Mags, Maggie, Margarita, Marguerita, Meg, Peg or Peggy, and her name might also be abbreviated to Mgt or Margt. The genealogy website Roots Ireland includes a list of common abbreviations for first names, and we have used the examples in the list to generalise names [16].
- **Last name.** While spelling was becoming more standardised during this period, there can be slight variations, therefore Soundex has been used to account for similar sounding names. Soundex is a phonetic algorithm that indexes names according to their sound. It encodes the first letter of a surname, and subsequent consonants, with

the aim of matching surnames that might have small inconsistencies in spellings. It is not a perfect system, but it has long been used by genealogy companies to underpin surname searches, to a reasonable degree of success. It is based on the English pronunciation of names, so it may have a lower degree of effectiveness for Irish surnames. As will be discussed later, in the case of women, surnames changed upon marriage and so this record field does not help to capture women who have married in the intervening period between the censuses.

- **Sex.** This information should be consistent across the censuses.
- **Approximate birth year.** The birth year should be consistent, but in the nineteenth and early twentieth centuries remembering birthdays was not prioritised as highly as it is now. While some families noted down birthdays in a family bible (probably more a Protestant tradition, or one used by Catholic emigrants) for many this was not information of great significance. Neither was an individual's age, until the coming of the Old Age pension. For these reasons, people may not have known their precise age. So, when matching birth years, we initially broaden the parameters to ± 1 year, and possibly extend the interval up to ± 5 years.
- **Birthplace.** This piece of information should be consistent across all records about an individual. Potential issues include inconsistency in spelling of placenames, particularly if the individual concerned was illiterate and relying on another to write the name on their behalf. It is also possible that in rural areas the name of a farm might be given in one census, and a townland in another. Issues such as changing/multiple placenames occur as well; many Irish placenames were Anglicised by the Ordnance Survey for instance.
- **Marital status.** In most cases, following marriage an individual's marital status moves from single to one of the following: married/widow/widower. It would have been highly unusual for a once-married person to describe themselves as single, as divorce was the preserve of the wealthier classes at that time, and even couples who were estranged or those who were deserted would generally maintain their status as married. For instance, returned migrant Mary O'Connor described herself as 'married in America' [17]. However, there are a few examples of individuals who described their status as 'divorced' such as Alfred Manning a 59-year-old retired merchant [18].

6 Results

6.1 Matches

Table 4 explains how many individuals received a given number of matches for each of the censuses. One individual received 139 matches in the 1901 census, while another received 125 in the 1911 census. Three quarters of individuals had a match in at least one census, but 1 in 4 individuals could not be matched in either census.

Table 4. No. Matches per individual in each census

No Matches	1901 Census	1911 Census
No matches	22	25
1–2	17	17
3–10	22	18
11–25	13	12
26–50	7	6
51–100	3	7
101+	3	1

6.2 Individuals with One or Two Matches

First, we considered a random sample of all individuals with one or two matches. Using traditional research techniques by the historians we looked at whether we could identify the correct individual in both censuses, and if so, whether this match was correctly identified by the application. The results are shown in Table 5: in all these cases it was possible to find the individual in the other census. In most cases the difficulty arose due to differences in the spelling of names, though in two cases the ages were very different.

Table 5. Individuals with one or two matches

Name and age	Located in Census	Correct Match?	Historian's comments
Louisa Malcomson (71)	1901	Yes	Yes, same address in death record and census [19]
William Claffey (72)	1901	No	Incorrectly matched to William Ashley. He was traced to the 1901 census in which he stated he was 59. This data was used to trace him forward to 1911 where he provides an age of 66 – in this case a margin greater than 5 years needed to be applied [20]
Henry Giblin (74)	1911	Yes	1 Bride Road (address on census) was part of Iveagh Buildings (death certificate). There is a potential match in 1901 but this would need to be verified [21]

(*continued*)

Table 5. (*continued*)

Name and age	Located in Census	Correct Match?	Historian's comments
Charlotte Holmes (76)	1911 only	No	On the death record Charlotte Holmes is living at 29 William Place. A search locates a 'Charlote Homes' aged 91 living at 29 William's Place. The informant on death certificate is Eileen Beadfield, likely a relative of Martha Bradfield, one of the people Charlotte was living with in the 1911 census. There is a possible match, Charlotte Rachel Holmes in Brooke Street, Enniskillen in 1901 [22]
Hans Murphy (78)	1901 only	Yes	Murphy is also found in the 1911 census; the 1901 census gives his address as Guinness Trust and his death record states Iveagh Buildings, the same location. He was married in both 1901 and at the time of his death so we searched on his wife's name 'Harriet Murphy' and found Hans living there at this address – but his name was written as 'Hands' [23]
James Riordan (82)	1901 only	Yes	In 1901 he was an army pensioner living in Cork, aged 73. In 1911 he cannot be located at his Upper Beechwood Avenue address but was found in the private hospital in which he died under the name James 'Reardon' [24]

6.3 Individuals Who Could not Be Matched in the App

Next, we took a random sample of five individuals who could not be matched to see if the historian colleagues were able to match them manually. Historians and genealogists start with what is known and work backwards, so we started with the death record, and then traced individuals to the 1911 census first, followed by the 1901. To emulate the application, we did not look at any other records to assist us in the matching process. In reality, if we could not locate an individual, we would cast the net more widely among available sources to build up a more detailed profile. Our findings will inform future iterations of the app (Table 6).

Table 6. Manual matching of individuals who could not be matched in the app

Name, age, occupation marital status and address	Located in Census	Historian's comments
Isaac Hearney (71), Coach Builder, widower St Albans Road	Yes, 1911 and 1901	We were able to match this individual to the census return of Isaac P. Kearney, a 73-year-old coachmaker born in Limerick and living at 19 St Alban's Road. In 1901 he was living at Greenville Terrace, Dublin [25]
Jane Elizabeth Roche (75), widow of a gentleman, 15 Earlsfort Terrace	Yes 1911 and 1901	Jane Elizabeth Redington Roche aged 75 was located at 15 Earlsfort Terrace. The issue in identification is down to the fact that her surname was indexed under Redington Roche. Using this surname, she was also easily located at Adelaide Road in the 1901 census [26]
David John Fitzgerald (76), Grocer, 14 Emor Street	Yes, 1911 and 1901	David Fitzgerald was easily located in the 1911 census, but his surname was written FitzGerald. A retired grocer's assistant aged 76, he was born in Wexford and lived at Emor Street, so a positive identification. The David Fitzgerald located in Royal Canal Bank in the 1901 census is likely to be the same person. He was born in Wexford, aged 60 and a 'packer' (mistranscribed as 'paller') [27]
Leibe Schein (83), widowed schoolteacher, 11 Walworth Road	Yes, 1911 and 1901	In the 1911 census Louis Shein aged 82 is a boarder in the Jaffey household (10 Walworth Road). While no occupation is given this is likely to be the correct person. The first name does not correspond, but members of the Jewish community often Anglicised their names so Russian-born Louis is likely to be Leibe. This is confirmed in the 1901 census when Russian-born Lewis Shine of 10.1 Walworth Road, is described as a teacher in Hebrew born in Russia [28]
Mary Townsend Saunders (85), Lady, Spinster Rest for the Dying	Yes, 1911 but not 1901	In the 1911 census Mary T Saunders aged 85 was located living in 8 Ulverton Road, Dalkey, Dublin. She was an annuitant (living on a sum of money she received each year) and single, born in Co. Cork. All this information matches her death record perfectly [29]. We searched the 1901 census for a Mary Saunders but were not able to locate her

6.4 Individuals with a High Number of Matches in the App

Finally, we reviewed a random sample of 5 individuals with a high number of matches to show how we might be able to narrow down the search results in the application. Typically, when doing this kind of research historians start narrow and broaden their search when no positive findings are forthcoming (Table 7).

Table 7. Manual matching of individuals who could not be matched in the app

Name, age, occupation marital status and address	Were they located by the algorithm?	Historian's comments
William Brown 70, married, clerk, Meath Hospital/69 Brighton Square	No	Located at 69 Brighton Square under the surname Browne. Age 68 (2 years less than death record so not located in search) and married. Occupation is 'commission agent, retired' rather than clerk [30]
John Obrien 70, married, army pensioner, 5 Mclean's Lane	?	No John O'Brien at 5 McLean's Lane in the census. Death cert shows his wife was Mary. Possible candidate is John O'Brien of 86 Crumlin, Terenure, whose wife was Mary, but further research required [31]
Margaret Byrne, 73, widow of coach driver, 1 Earlsfort Terrace	Yes, 1911	Margaret was located at 1.2 Earlsfort Terrace, aged 73. Spelling of name, age and address are all precise matches to death record. She was not located in the 1901 census [32]
Michael Byrne, 71, widowed labourer, 36 Upper Kevin Street	No	Michael Byrne located at 36.3 Upper Kevin Street in the 1911 census aged 77, widower, labourer. The algorithm did not identify him due to too great an age discrepancy. He did not live at this address in 1901; Michael Byrne of 73.4 Francis Street is a possible match, but more research required [33]
Mary Breen, 75, widow of horse dealer, 12 Golden Lane	Not in 1911 but in 1901	Mary Breen, aged 78, was located at 11.2 Golden Lane. Age is 3 years above that in death certificate, and street name is identical (number different). She was located at 11.1 Golden Lane in 1901 aged 65. This is exactly 10 years younger than the age in the death certificate which is why she was matched in 1901 [34]

7 Discussion

This exercise allowed us to identify several challenges with the data matching:

Dataset Size. While we only used a small sample of data for this trial, the subsequent attempt to run the matching application with a larger data set of ca. 8 million records (the full 1901 and 1911 census) pushed our low-code technologies to their limits, and we needed to use custom database code.

Anonymised Entries. Of the full dataset, over 155,000 records relate to anonymised entries. It will be very challenging to be able to match these records which only utilise initials of first name and last name.

Birth Year. Some individuals were highly accurate in recording their birth year, but others much less reliable. For some individuals this is because knowing their own age or date of birth was not as important as it is now, but for others this might be due to age heaping, where dates are rounded. In other cases, this imprecision or differences may have been intentional (particularly after the old age pension had been introduced). A search of plus or minus one year is too limited a time window – a range of ten years either side is likely to capture most eventualities.

Inconsistent Responses. Even where specific responses were required, some individuals still provided an alternative response. For instance, though there were 4 options for marital status ('married', 'widower', 'widowed' and 'single'), across the entire census dataset we found 227 distinct status descriptions, which included non-standardised responses such as 'no', 'still single', 'unmarried (unfortunately)', 'on the lookout' and 'not yet'.

Name Changes After Marriage. This is a major issue when tracing women. However, it could potentially be solved by also incorporating the marriage certificates into the data pool.

Transcription Inaccuracies. Even though marital status should be relatively easy to transcribe, some issues were identified. For instance, in the case of Edward Flynn, Cookstown his marital status is mis-transcribed as 'less than one year' yet in the original census return he stated 'married' [35]. This additional information should have been transcribed into the 'years married' data field.

Record Sets Containing Different Information. All these records were designed for highly specific purposes, and being able to link them was not a factor in their original design. Thinking of future applications for data matching, one issue will be identifying common information between all these records. For instance, the birthplace is very useful in census matching, but it is missing from the marriage and death records.

We also identified some areas that could be further enhanced. The historians noted that the address can be a very useful property to include in searches, particularly where a death occurs one or two years after a census. In the cases where the historians checked the data manually, using an address to match allowed for a positive identification. If the address does not match, it makes sense to search first in close proximity (e.g. within the same District Electoral Division), and only then broaden the search to a county, or Ireland as a whole. This gradual spatial approach should lead to more accurate results.

Inconsistent spellings of names posed a problem as shown in the tables above: e.g. Hans was modified to Hands, Riordan to Reardon. A system other than Soundex might be used, or we could develop a solution that handles the specifics of Irish names. Finally, in many cases the standard plus or minus 5 years window for age searches was not broad enough. A plus or minus 10 years would work better in some instances. We also noted that the highest number of matches appeared with common surnames, but that using Soundex many similar surnames were returned, e.g. Brien, Breen and Brown. These are distinct surnames in terms of pronunciation, so Soundex may not be the best solution.

8 Conclusions

While the initial findings are promising, more refinement is required for the application to work well. Rather than having a single standardised set of search criteria it may work better to create a decision tree to better emulate the flexibility of the knowledge discovery mode in which historians would progressively work their way through the manual matching process. This is the case because at the start of any search historians look at all the information available to them and decide which starting point (name plus address, birthplace, birthdate or even occupation) seems most likely to give the right results. In the future, we hope to expand this process within the application using some kind of AI or machine learning technique, which could help to identify similarities in the entries, which have not been obvious to the human eye. Such innovation will deepen understandings of the extent of 'welfare fraud' [5] in early twentieth-century Ireland and will add to the growing body of work in age-cohort studies using individual level data [36].

Acknowledgements. "Death and Burial Data: Ireland 1864–1922" is a project funded by an Irish Research Council Laureate Award IRCLA/2017/32. This publication has emanated from research supported in part by a grant from SFI R@ISE SPP under grant number 21/SPP/9979.

References

1. Gilleard, C.: Old Age in Nineteenth-Century Ireland. Palgrave Pivot, London (2017)
2. Johnson, P.: Historical readings of old age and ageing. In: Johnson, P., Thane, P. (eds.) Old Age from Antiquity to Post-Modernity: From Antiquity to Postmodernity, pp. 2–3. Routledge, London (1988)
3. Thane, P.: Social histories of old age and aging. J. Soc. Hist. **37**(1), 93–111 (2003)
4. Old Age Pensions Act 1908, 8 Edw. 7. c. 40
5. Gráda, C.Ó.: 'The greatest blessing of all': the old age pension in Ireland. Past & Present **175**, 124–161 (2002)
6. Breathnach, C., Murphy, R.: Fine lines: locating commercial sex work in official data, Dublin 1901 and 1911. Urban History **51**, 435–456 (2023). https://doi.org/10.1017/S096392682300010X
7. Annual Report of the Registrar-General, p. 137 (1909)
8. Helgertz, J., Price, J., Wellington, J., Thompson, K.J., Ruggles, S., Fitch, C.A.: A new strategy for linking U.S. historical censuses: a case study for the IPUMS multigenerational longitudinal panel. Hist. Methods **55**(1), 12–29 (2022)

9. Mandemakers, K., Alter, G., Vézina, H., Puschmann, P.: Sowing: The Construction of Historical Longitudinal Population Databases. Radboud University Press, Nijmegen (2023)
10. Christen, P.: The data matching process. In: Data Matching. Data-Centric Systems and Applications. Springer Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31164-2_2
11. Breathnach, C., Ibrahim, N.M., Clancy, S., Margaria, T.: Towards model checking product lines in the digital humanities: an application to historical data. In: ter Beek, M.H., Fantechi, A.. Semini, L. (eds.) From Software Engineering to Formal Methods and Tools, and Back, pp. 338–364. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30985-5_20
12. Schieweck, A., Murphy, R., Khan, R., Breathnach, C., Margaria, T.: Evolution of the historian data entry application: supporting transcribathons in the digital humanities through MDD. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), Los Alamitos, CA, USA, pp. 177–186 (2022). https://doi.org/10.1109/COMPSAC54236.2022.00033
13. Pedersen et al (in this volume)
14. History of Irish Census Records. https://www.census.nationalarchives.ie/help/history.html
15. Doherty, A.J., Murphy, R.A., Schieweck, A., Clancy, S., Breathnach, C., Margaria, T.: CensusIRL: historical census data preparation with MDD support. In: 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, pp. 2507–2514 (2022). https://doi.org/10.1109/BigData55660.2022.10021106
16. Roots Ireland First Names. https://www.rootsireland.ie/help/first-names
17. Mary O'Connor. http://tinyurl.com/Mary-O-Connor-1911
18. Alfred Manning. http://tinyurl.com/Alfred-Manning-1911
19. Louisa Malcomson. http://tinyurl.com/Louisa-Malcomson-1911 and http://tinyurl.com/Louisa-Malcomson-1901
20. William Claffey. http://tinyurl.com/William-Claffey-1911 and http://tinyurl.com/William-Claffey-1901
21. Henry Giblin. http://tinyurl.com/Henry-Giblin-1911 and http://tinyurl.com/Henry-Giblin-1901
22. Charlotte Holmes. http://tinyurl.com/Charlotte-Holmes-1911 and http://tinyurl.com/Charlotte-Rachel-Holmes-1901
23. Hans Murphy. http://tinyurl.com/Hans-Murphy-1911 and http://tinyurl.com/Hans-Murphy-1901
24. James Riordan. http://tinyurl.com/James-Riordan-1911 and http://tinyurl.com/James-Riordan-1901
25. Isaac P Kearney. http://tinyurl.com/Isaac-P-Kearney-1911 and http://tinyurl.com/Isaac-P-Kearney-1901
26. Jane Elizabeth Redington Roche. http://tinyurl.com/Jane-Redington-Roche-1911 and http://tinyurl.com/Jane-Redington-Roche-1901
27. David John Fitzgerald. http://tinyurl.com/David-John-Fitzgerald-1911 and http://tinyurl.com/David-John-Fitzgerald-1901
28. Leibe Schein. http://tinyurl.com/Leibe-Schein-1911 and http://tinyurl.com/Leibe-Schein-1901
29. Mary Townsend Saunders. http://tinyurl.com/Mary-Townsend-Saunders-1911
30. William Brown. http://tinyurl.com/William-Brown-1911
31. John O'Brien. http://tinyurl.com/John-O-Brien-1911
32. Margaret Byrne. http://tinyurl.com/Margaret-Byrne-1911
33. Michael Byrne. http://tinyurl.com/Michael-Byrne-1911 and http://tinyurl.com/Michael-Byrne-1901
34. Mary Breen. http://tinyurl.com/Mary-Breen-1911 and http://tinyurl.com/Mary-Breen-1901

35. Edward Flynn. http://tinyurl.com/Edward-Flynn-1911
36. Breathnach, C.: Parish pump politics and the introduction of Old Age Pensions in Ireland. Accepted and forthcoming History Workshop Journal