Mapping the Great Irish Famine

Paul S. Ell
The Centre for Data Digitisation and Analysis
School of Sociology
The Queen’s University of Belfast
Northern Ireland.

This paper describes the creation of a new paper atlas – *Mapping the Great Irish Famine: A Survey of the Famine Decades* – which has recently been published by Four Courts Press in Dublin in the Republic of Ireland.¹ The paper breaks down into three distinct sections. In the first section, the Database of Irish Historical Statistics, a large body of statistical material from which data were drawn to create maps in the atlas is described. Then, in the second brief section comment will be made on how this statistical resource was created – essentially using state of the art Optical Character Recognition Technology. Even more briefly, mention will be made of how the Centre for Data Digitisation and Analysis have applied the same technology to other printed sources. Finally, in the third section some of the geographies of the Famine that the Atlas reveals are examined.

To the first section, the collection of statistical data. The majority of work done mapping the famine is drawn from the vast body of data collected as part of the Database of Irish Historical Statistics research project.²

The Irish Database project had several aims. The main one was to gather recurrent census statistics for Ireland from the published census returns from the first comprehensive Irish census in 1821 up to 1971 thus including all censuses that were not available in a machine-readable form. A second aim was to facilitate regional and national research on Ireland. It was considered that by making a large body of statistical material readily available on Ireland, this would encourage researchers to

² The Database of Irish Historical Statistics may be obtained, in machine readable form, through application to the History Data Service, ESRC Data Archive, University of Essex, Colchester, England – dawww.essex.ac.uk.
use the material. Furthermore, the availability of the material would reduce the research costs of any project since it would not be necessary for them to gather data.

Several historiographical imperatives were identified as the driving force behind the project. First, the published Irish census returns are not widely available. In fact, even within the British Isles, few Universities have comprehensive holdings of the published returns, in part because, unlike most of the early British returns, the Irish publications have not been re-printed, although the text of the Registrar General’s Reports for the period between 1821 and 1911 recently have been.\(^3\) Second, advances in computing began to make possible the gathering, storing and manipulation of very large quantities of numeric data. Third, there was an increasing interest in Irish history, not just in Ireland but in the United States, in Great Britain and in Australia and New Zealand. The database exploited this interest and the lack of source material outside Ireland.

So, what is the Database of Irish Historical Statistics? Essentially, it is a very large body of temporal data at a range of different spatial levels. In total the Database holds over 32 million individual data values. When it is remembered that these are raw data values, after the creation of what might be considered to be a limited range of derived variables, the total dataset exceeds 150 million individual values, with the potential, through the creation of new derived variables to increase it still further. In other words, the Database is one of the largest historical datasets for the British Isles in existence.

The Database includes most recurrent census statistics for Ireland. It contains comprehensive demographic data, occupational data, statistics on literacy, language and religion, information relating to poor law union relief, trade statistics, figures concerning agricultural output, data on industry and industrial structure, unemployment data, wages data, crime figures and statistics concerning housing quality. Statistics from some non-census sources are included, for example annual vital birth and death statistics. Generally, data are available every tenth year when a population census took place. For certain periods however, particularly in the 1830s, 40s, 50s and 60s, for the period before, during, and after, the Great Irish Potato Famine, annual statistics, where available, have been collected.

---

The statistics in the Database are found at a variety of spatial levels. The spatial levels themselves also vary over time. Generally, data are available for Ireland as a whole, for the 32 counties, 160 Poor Law Unions and 320 baronies in the nineteenth century. For the work on the Famine we have produced maps at county, poor law union and barony level depending on which statistics were available for each spatial unit. The same statistics, with the exception of demographic indicators, are rarely available at all spatial levels. It is from this large repository of quantitative data that material was selected for mapping as part of the Famine Atlas project.

The second part of this paper concentrates on how the statistical data that encompass the Database of Irish Historical Statistics was computerised. Initially, this vast body of data was manually keyed, normally using a double entry system – one operator would key a page of numbers and another would do the same page again. The two files were compared, differences highlighted and investigated. Clearly, this was a slow and cumbersome process.

During the project the use of Optical Character Recognition or OCR technology was identified as a possible option to circumvent this tortuous keying process. However, this was in 1993 when the approach was not well developed and was really only advised for modern clear documents, and even then significant error rates could be expected. The census volumes were far from this ideal. They were old, they were brown, they were torn, and the information we were interested in was numbers set into tables.

However, after extensive research, an OCR package, which was far more powerful than the industry standard at the time, was identified – the ProLector system. Of course, in the adoption of cutting-edge technology there are often difficulties – the software might not actually work and it is frequently expensive. ProLector was certainly expensive. At the time, with associated hardware – a then powerful personal computer and a large flatbed scanner – the system cost in excess of £20,000, but it did work. The system allowed the rapid, and very importantly, accurate semi-automated capture of numeric material. Equally, the system could accurately capture alpha data as well.

The advantages of the system are manifold.

- The lack of pre-installed font bases which are standard in most OCR systems. ProLectoris completely operator trained. A new font base can be developed for each census table, for example, taking into account not only the changing font but
also the quality of the print. Very large font bases can be created so there might be 30 different variants of a number five entered into the font base.

- The linkage of the software to the scanner is well developed. Whereas other systems require the user to scan a full document saving the images in computer graphics formats such as TIF or JPG files, under our system each page is scanned and then immediately OCR’d. This allows the operator to adjust scanner settings according to the condition of the document. Dependent on OCR accuracy levels the operator can adjust scanning resolution and contrast.

- The system can be trained to ignore lines separating columns within tables. It was found that other systems insisted on processing these lines normally ‘recognising’ them as either and I or a 1 for example.

- It is possible to give weightings to individual characters. A higher weighting forces the software to be more critical when processing it – useful where I and 1 are confused.

- The system can preserve document formatting. This goes beyond simply recording the position of text with tabs. ProLector can store bold, italics, superscript and subscript and can mark changes in font type.

- Of greatest significance is the systems ability to recognise almost any typeface from all but the poorest document and quickly and accurately process the material.

The ProLector system performs extremely well. One operator can process over 150,000 data values or individual words or whole numbers in a day. In practice accuracy levels are extremely high with in excess of 99.995% of characters being correctly recognised. We expect no more than five errors per ten thousand characters digitised.

The system does have a few disadvantages. The main one is that it does not handle well large amounts of numeric data mixed with small amounts of alpha data in column format. In practice numeric columns have to be scanned separately from alpha columns. In addition, the size of the font base can be large but does have a physical limitation of 256kb. Further, it is occasionally necessary to photocopy poorer quality material using the copier to maximise contrast and increase page size. In addition, very small documents might not be worth digitising. The system has to be trained to recognise each new font. Whilst this does not take long, it may be quicker to key the document. Second, where characters run into one another the OCR system will not recognise them – unless of course the same characters consistently run into one another in which case the software can be trained to recognise this recurrent error. Should the characters not be recognised, the operator has, at this point, to intervene,
and type the correct information. If this happens frequently, as it does with early nineteenth-century newspapers for example, again it might be best to type the whole text in.

The Centre has not just utilised OCR technology in the creation of the Database of Irish Historical Statistics and the Famine Atlas, but on a host of other projects. In April 1998 we received a grant of £260,000 from the Economic and Social Research Council to construct an historical Geographical Information System for the British Isles with colleagues at the Universities of Portsmouth and Leeds. We are responsible for populating the time variant GIS with statistical data – almost all of it collected using OCR techniques. The work will result in a GIS with associated non-locational data available as a tool for interested scholars and, in addition, an Atlas celebrating the bi-centenary of the first modern British census in 1801 should be published in late 2001.

Staff from the Centre have also digitised Hearth Tax Data from nineteenth-century antiquarian transcriptions, created a full transcript of the printed 1871 Census for Scotland, digitised John Williams Digest of Welsh Historical Statistics, computerised nineteenth century hospital statistics from hospital yearbooks, and are currently computerising the 10 volume Scottish National Dictionary which records, in very considerable detail, lowlands Scots.

The third section of this paper examines some of the geographies of the Great Irish Potato Famine. It is important to realise that whilst, in the past, much of the work by scholars concerned with the Famine employed often complex quantitative or qualitative techniques, and not infrequently references were made to the spatially differing impact of the disaster, very few maps have actually been created, and no attempt has been made to chart systematically the regional impact of the Famine. Those maps that do exist often are at a poorly detailed spatial level such as the county or are merely redrafting of maps already created by earlier scholars.

Against this backdrop Mapping the Famine project was launched which has resulted in Mapping the Great Irish Famine: A survey of the Famine decades. The project covers the years from 1831 to 1871 and maps data taken from the decennial population censuses, one-off sources such as information from the British Parliamentary Papers Famine series and annual data from a variety of sources such as the Irish Agricultural Statistics – all extracted from the Database of Irish Historical Statistics. The project has two principal aims - the first, to establish at a detailed
spatial resolution the geography of the famine. The second, to seek to explain through quantitative and cartographic analysis that geography.

The Atlas looks in detail at the period surrounding the Famine largely based on the 1831 and 1841 censuses – affording a view of the period leading up to the Famine, and the 1851 and 1861 censuses showing post Famine development. Specifically, sections in the atlas are concerned with the Unpeopling of Ireland – with maps series on population, emigration and mortality. In the section sex, marriage and families – changing gender ratios in Ireland pre and post Famine, marriage patterns, and household structure are discussed. The Atlas then goes on to examine the condition of the people with map series on diet, housing quality, religion, literacy, the Irish language, disease, and the relief of the poor. The final section charts changes in the Irish economy with maps examining wages, occupations, and changes in agriculture.

Detailed visualisations reveal fascinating patterns. No paper on the famine would be complete without a description on its demographic impact and here we concentrate on population loss. In the century before the Great Famine the population of Ireland grew faster than anywhere else in western Europe. Population growth was particularly marked up to the 1810s after which growth rates declined due to later marriage in the east and increased emigration, particularly to England. We cannot say whether this trend would have continued for the potato crop failed and the Famine intervened. What we can say is that between the 1841 and 1851 censuses of Ireland the Islands population fell by 20 per cent. We can also say that this trend continued. In the 1850s population fell by a further 11 per cent and almost 7 per cent in the 1860s. The maps show the dynamics of this change. They plot population change between 1841 and 1851, 1841 and 1861 and, finally, between 1841 and 1871. Between 1841 and 1871 308 of the 320 baronies in Ireland lost population. Of these 60 lost more than half of their population. Population loss was, therefore, widespread, but was at its most severe in inland baronies in Connacht, Munster and Leinster. Ulster, with its more industrial economy, suffered less as did many coastal baronies, perhaps their populations relying on the resources of the sea. Some cities actually increased their population over the period such as Dublin, Cork and Belfast as work was sought in towns as an escape from the countryside.

Moving on to another map series, this on housing. Unusually, early Irish censuses classified houses into one of four quality categories. Fourth classes houses were defined as ‘all mud cabins having only one room’, third class as ‘a better description of cottage, still built of mud, but varying from 2 to 4 rooms and windows’.
A second-class house was defined as ‘a good farm house, or in a town, a house on a small street, having 5 to 9 rooms’. Finally, first class housing was all housing of better quality than the previous three classes. Of the entire map series that can be produced from this rich seam of data perhaps the most interesting is that showing the distribution of fourth-class housing. In 1841 fourth-class housing was widespread, probably a consequence of the rapid rise in Ireland’s population over previous decades – there was great demand for shelter. Whilst widespread, there is a strong regional dimension to the distribution of these poor mud huts. In some baronies in the far west of Ireland over 80 per cent of housing fell into this poorest of categories. At the other end of the scale, in the urban areas of Belfast, Dublin and Cork, fourth-class housing was rare accounting for less than three per cent of the housing stock. Overall, fourth-class housing was most common in 1841 in the west of Ireland – in Donegal, in Mayo, in Galway, in Clair, Kerry and Cork. By 1851 the pattern had changed in a process that continued into 1861. The preponderance of fourth-class housing declined. It was still most frequently found in the west but, by 1861, few baronies had more than half their housing stock in this class. Conversely, by 1851 almost all of Ulster was completely free of fourth-class housing. It is clear that, as the population was decimated by death and emigration, it was these houses that were abandoned first. It might also be suggested, if we use housing quality as a proxy measure for wealth, that it was the occupants of these houses who were most likely to be affected by the famine and forced from their homes either through death, disease or the search for a better life in Ireland’s towns and cities or abroad. At county level the maps perhaps show even more clearly this process underway.

Other housing classes largely mirror, in terms of spatial distribution, the fourth class pattern. As the numbers of fourth-class houses declined, increasingly Ireland’s population lived in second or third class houses. To a fair degree they always had in the east, but this accommodation became increasingly common first in the midlands and, by 1861, in the west as well. A county map of the distribution of second-class housing makes this point well.

To conclude, it is quite clear that the famine had a spatial dynamic. One only needs to look at the changes in the distribution of fourth class housing to demonstrate this. It is clear that spatial approaches to the study of the famine should take their place alongside the other, more traditional, quantitative and qualitative approaches. This new spatial approach – new to the study of Irish history at least – has only been possible through the construction of the Database of Irish Historical Statistics. Its creation, in turn, rests on the use of Optical Character Recognition technology.
Percentage population decline at barony level
Percentage fourth class houses at county level
Percentage second class houses at county level
Percentage second class houses at barony level