Introduction

The histpop website (http://www.histpop.org) provides free online access to over 185,000 pages of published reports relating to Britain’s population. The main contents of the website are reproductions of almost all the published census returns and the Registrar-General’s reports for England, Wales, Scotland and Ireland up until the Second World War, but there is a wealth of other material accompanying these reproductions. This includes over 6,000 pages of archival material from The National Archives (TNA), the complete texts of all relevant parliamentary legislation, a variety of contextual essays relating to the collection of population information across the period, a number of downloadable tables of statistical information extracted from the census and Registrar-General’s reports, plus all the other content and functionality one would expect to find on a website of this nature.

The website is called histpop and is the main output of the Online Historical Population Reports Project. The project is run as part of AHDS History within the UK Data Archive at the University of Essex, and has been directed by the author of this article. The project was funded as part of the first Joint Information Systems Committee (JISC) Digitisation Programme. The project started work in early 2004 and completed at the end of March 2007, though the site continues to be maintained, and hopefully more content will be added over time, enhancing this vital resource for historical demographers and local population historians. The project is predicated on the fact that the various population reports published by the Registrar-Generals and their predecessors are not as widely available as they should be, given their short print runs and their increasing fragility. We agree with David Caradog Jones’ 1949 statement that ‘the student is recommended to examine the various census volumes for himself’, but we can no longer agree that his follow up comment, ‘They should be available in the central Reference Library of any large town’, remains valid, especially for some of the older volumes.
It is also worth noting that many students of local population studies have only used the reports in a cursory fashion, generally to extract information about a place of interest. While the statistical information within both the census reports and the various reports of the Registrar-Generals is important, there is also a wealth of mainly untapped qualitative information within these documents. The reports, to a large extent, reflect the thinking and ideologies of their time. Occasionally, some more quirky passages have been discovered in the reports: William Farr, for example, seems almost obsessed about ensuring that scalding water should be kept out of reach of children, returning to the subject on a number of occasions. Prevailing ‘official’ views on alcohol consumption are also aired frequently. Farr quotes the coroner of Newcastle-upon-Tyne on the subject of sudden deaths of children who were ‘suffocated in bed by the parents, very frequently either one or both of them, going to bed in a state of intoxication’, and appended the comment that, ‘Here is another evil to be referred to the evil effect of excessive drinking’. This ‘trivia’ is to be compared with the considerable written evidence on a large number of other subjects which may not have received the attention they deserve in the scholarly literature: subjects such as disabilities, housing, administrative geography and working conditions, among others. The Irish and Scottish census reports also deserve greater notice, and not just by scholars interested in those areas. The necessity of the different Register Offices working in tandem throughout the late nineteenth and early twentieth centuries means that much interesting discussion of the practices of the English Registrar-General is found in the Irish and Scottish census reports.

Contents of the site in more detail

Census reports

The website contains all printed material submitted to the British parliament relating to the census up to and including the 1931 census. This means that not only are census reports included, but other ancillary documents relating to the census-taking process, such as the evidence and report of the 1890 Treasury Committee inquiring on the census, and the Report of the Committee for the Bill for the 1831 census, both of which are seminal texts for the development of census administration. All of these reports are digitised in full. This means that not only are the pages of each report captured in a digital form, but all the texts have been OCRd. Unlike some other ‘digital library’ projects, considerable effort has been made to ensure that each volume is easily navigable through the use of electronic ‘tables of contents’ (see Fig 1).

In order to create a basic list of the various materials which were available, and to ensure precision in referencing, we have had recourse to a fair quantity of reference materials which have been created over the years. Most of these are listed in the following footnote, but it is worth noting that small errors—sometimes of omission, sometimes of commission—have slipped into these reference works, and without a comprehensive library of these works it was a struggle at the beginning of the project to get to grips with the limitations of

Table of Contents

Display: Sections Tables Page Titles

Title & contents pages (Pages 1-4)

England and Wales (continued) (488 pages)

- Division vii. North-Midland (1 foldout, Pages 1-9, Pages 2-99)
- Division viii. North-Western (1 foldout, Pages 1-7, Pages 2-73)
- Division ix. York (1 foldout, Pages 1-11, Pages 2-107)
- Division x. Northern (1 foldout, Pages 1-7, Pages 2-74)
- Division xi. Welsh (1 foldout, Pages 1-11, Pages 2-91)
Figure 2  Screenshot showing a page
some of these resources. Another problem faced at the beginning of the project was to unravel the different reference numbers for parliamentary papers and other government publications: the relationships between physical volumes, logical volumes, parliamentary paper references and command numbers can sometimes stretch the limits of credulity. For example, each ‘county’ volume for the 1911 Census of Scotland was published separately but with sequential page numbering across all 37 volumes and had the same command number (Cd.6097) but was shared across two volumes of parliamentary papers (BPP 1912–13 CIX and BPP 1912–13 CX). The production of a complete list of materials took a considerable amount of time, and we hope to publish separately a bibliography of parliamentary papers relating to population.

The coverage of census reports in the histpop website is comprehensive in relationship to its terms of reference. All English, Welsh and Scottish census reports published before 1951 are included. In essence this means that the last complete censuses included is the 1931 round, even though the General Report for the 1931 census was not published until 1950. For Ireland, all published reports relating to the censuses are included, covering censuses from 1813 to 1911. There was no census in Ireland in 1921, and the first census of the Republic of Ireland was taken in 1926. For various reasons we could not include the relevant reports in histpop, but we did include the reports from the separate censuses taken in Northern Ireland in 1926 and 1937. For good measure, we have also included a number of other volumes presented to parliament including returns of expenses, the British Empire census report for 1901, the returns of the census of London of 1896, and the rather remarkable letter from the Registrar-General apologising to the people of Wales for suggesting that they had ‘fraudulently’ filled in their schedules in order to inflate the number of Welsh speakers.

Registrar-General’s reports

The selection criteria for the reports of the Registrar-Generals were slightly different. All the Annual and Decennial reports of the Registrar-General (of England) published until 1920 have been included, as have all ordinary and detailed reports of the Registrar-General of Scotland for the same period. So what is missing? All reports of the Irish Registrar-General, and all weekly and quarterly reports of the British Registrars-General have been excluded. The additional number of pages would not have been great but the issue of size, along with the question of finding copies which were suitable for scanning, caused us to ignore these volumes for the time being. The cut-off date of 1920 was selected to bring coverage up to the beginning of the change of name and format of the reports in 1921. As this was a suitable cut-off point for England and Wales, it was assumed that this might be similarly so for Scotland.

As with the census reports, a number of additional related publications are also included on the site. Most importantly, both editions of the Annual Reports of the Registrar-General (ARRG) for the period 1837–1854 have been included. During this period two versions of these reports were published, the ‘Sessional (or
Parliamentary) Paper’ series and ‘Registrar-General’s’, which had considerable differences in terms of content. Furthermore, not only have the decennial supplements been included, but also other ad hoc reports, including the reports on cholera and influenza published in 1868 and 1920 respectively.

In all, the number of pages of reports reproduced in this section of the website totals 186,005. Each page image was scanned at a high resolution, and ‘cleaned’. The cleaning involved despeckling and straightening. After this process was undertaken for each volume, all the pages were OCRd to enable searching on the site. Clean OCR, with a very high accuracy rate, was produced for each page which only contained text, whereas a lower rate of accuracy was allowed for tabular information. This means that there will be times when a word appears on the printed page, but cannot be found by searching on the web site.

**Material from the National Archives**

This third section of digitised images comprises 6,244 pages of material from The National Archives. One member of the project team spent the best part of six months at TNA selecting and scanning material from the collections of the Registrar-General. These 6,244 pages are a very small proportion of the total archival material available at TNA relating to the census, but we are confident that they are a reasonably representative selection of the complete archive. However, examples of some key documents, like Census Enumerators’ Books (CEBs)—for all available years—have been scanned, as have as many instructional guides as possible. The CEBs are scanned in full colour, and it might come as a surprise to those who are used to viewing microfilmed images that the 1851 enumerators’ books were printed on blue paper, and that the tickers who abstracted data from the CEBs used different coloured pencils for their work. Examples of original householder’s schedules, where they survive, have also been included, highlighting the differences between the schedules and the CEBs.

Possibly of more interest to LPS readers are the various occupational dictionar-ies known as Instructions to the clerks employed in classifying the occupations and ages of the people: the three unpublished lists for the 1871, 1881 and 1901 census have been added to the site. Other instructions, forms and maps used in taking the census in England and Wales, in 1841 and from 1861 to 1921, have also been incorporated. For example, there are instructions issued to local officers regarding their duties in taking the census; specimens of census sched-ules distributed to households, institutions and vessels in England, Wales and the Channel Islands; examples of enumerator’s memorandum books, in which enumerators recorded their progress in delivering and collecting census schedules in their enumeration district; and forms for the division of registrars’ sub-districts into enumeration districts. There are also many examples of General Register Office correspondence with various governmental and non-governmental bodies in the preparation of the census, for example regarding census legislation; the preparation of, and questions to be included in, the census schedules; the co-ordination of census activities between England, Wales, Scotland and Ireland; the adjustment of enumeration district bounda ries, and the appointment of staff; and the tabulation and publication of results.
Finally there are also around 1,500 maps used at the time of various censuses, showing registrars’ districts and sub-districts, and containing manuscript additions recording changes to the administrative geography.

The material included is a small selection of that available. Much of it will be new to users of the census, and much of it helps shed light on the manner in which the statistics were created. Most researchers will find evidence in this collection to assist them in interpreting the quantitative data, understanding
why certain practices took place, and generally learning about the construction of statistics from the middle of the nineteenth century through to the middle of the twentieth century.

Essays

Under this heading can be found a collection of specially commissioned essays which provide contextual detail for the other resources within the site. These essays, ranging in length from 1,500 to 2,000 words, have been written by Dr Edward Higgs and myself. They provide background information and are designed to lead students and researchers through the different types of material available on the site. Some are reasonably didactic, like the essay entitled Constructing a parish population history; some are descriptive, like the Introduction to administrative units of England and Wales; while others again are analytic, like Higgs' The development of population statistics. At the time of writing there are 108 of these essays on the site, and the project has been especially commended by researchers for providing this contextual information, which helps both experienced and less experienced population researchers to get a firm grip on the material published in the reports.

Legislation

The final section, Legislation, contains the complete texts of all parliamentary legislation relating to the collection of population statistics between 1801 and 1937. For copyright reasons we did not create reproductions of the Acts of Parliament, but the full texts of 59 pieces of legislation have been provided on the site. Some may consider the Acts of Parliament to be rather dry, but the complete collection of legislation shows clearly how the registration and census processes evolved. The detail of these acts often provides insights into the administrative procedures for the censuses and registration which cannot be found at the National Archives. This legislation is essential reading for anyone who wishes to fully understand the collection of population statistics in the British Isles. Further research could be carried out to examine the debates relating to the legislation as it went through parliament. Hopefully, some of this legislation will spur people to examine the development of British demographic statistics into the twentieth century.

How it all works

In writing this article, I have assumed that most LPS readers are not really interested in the technical details of the website. However, it is important to note that since the project was funded by the JISC, which promotes standards in information technology, some mention of the technical side of the project must be made. What really needs to be explained is that the website is only the visible front-end of an online repository which contains over one million images, a vast quantity of associated metadata and a highly complex database. Metadata is needed for users of the site to find what they want, but it is also vital for the long-term preservation of the material. Best practice in
digitisation, good metadata and plans to sustain the site will all help to ensure that the material is available for use well into the future.

The website makes use of the latest open source software technologies and uses open standards for the presentation of the available resources. No client-side Java scripting is used and the only restriction on use is that users’ web browsers must be XHTML 1.x compliant, which means that most potential users will find that the site looks and functions identically whatever browser they choose to use. The whole of the back-end of the system is designed to make the user experience as simple as possible, but the highly complex nature of some of the material has meant that some functionality is perhaps a little more opaque than it could be. With a website of this complexity users may have to persevere a little in order to find what they want, but the time saved and the facility with which most material can be found will prove worthwhile to most users.

**Navigation through the site**

Every effort has been made to ensure that the site is as usable and intuitive as possible. This has meant that we have paid special attention to the navigation bar which is on the left hand side of the screen when viewed. In layman’s terms, we expect most users to want to browse the volumes rather than to search them, though the option of searching is also available.

To take a simple example, let us assume I wanted to find the population of the parish of Northolt in 1871. From the navigation bar on the left hand side of the browse page, I would click on the small arrow to the left of the word census, then expand 1871, click on England, and choose the relevant volume (Registration Counties). After this I would see a complete table of contents for that volume and scroll down to Division 4: Middlesex. The table of contents shows that Table 4 contains the population of parishes. Then I would click on the relevant link on the far right hand side. The population of each of the parishes in the Registration District of Uxbridge are found on the first page.

From entering the site to finding that statistic takes six clicks of the mouse! This, of course, is with knowledge of the site and an understanding of the general arrangement of the different volumes, but from entering the site to any page of original report should take no more than ten clicks. I could also have found this page in slightly fewer clicks by searching for Northolt in the simple search. Searching will prove more valuable for searching texts rather than tables. Both searching and browsing can be refined to access the data by date and/or by geography.

Generally speaking the results of browsing and searching facilities are presented in two stages. First a table of contents is displayed which lists the volume and section titles from which users can select the page they wish to view. Having made that selection, the selected page is displayed as an image together with information about where the page is positioned in the volume from which it has been extracted. By means of this information users can navigate to adjacent pages or to other sections in the same original volume.
Other general facilities include image zooming, rotation and download (in both high resolution TIF and PNG formats) and the ability to download a selection of tables in spreadsheet (.xls) format, and—most pertinently for academic use—the correct citation for each original report is given. Many of the footnotes in this short article have simply been cut and pasted from the website.

Considerable effort has been put into ensuring that the browse function is as flexible as possible: new ‘contents’ pages have been designed to reflect the overall structure of each volume of material and, through the use of metadata, every single document in the system is linked by subject material to other documents. This ‘Associated Content’ feature maximises flexibility in browsing and allows users to move easily from one document to another, while also providing a link of interest to those who are less familiar with the contents of the reports.

The website was officially launched in mid-January 2007. Duncan Macniven, the Registrar-General for Scotland, formally launched the site, and presentations were given by a number of specially invited researchers, including Eddy Higgs, Colin Pooley, Anne Hardy, Andrew Hinde, Paul Ell, Tony Franklin and myself. For the OHPR team, the day was the successful culmination of three years hard work, and it was very satisfying to receive such a good response to the site.

The real impact of a resource can partially be found in the web statistics, and I can report these for the first 51 days of 2007. In this period we had just over 9,500 unique visitors, looking at a little under half a million pages. In the 21 days of February 2007 which have elapsed at the time of writing, over 20 per cent of ‘visits’ to the site have been for longer than five minutes. We are hugely encouraged by these numbers, and hope that through careful marketing we will encourage greater use. To this end, I would like to offer a histpop mug to the first five readers of LPS who email info@histpop.org the occupations of the women who died through being struck by lightning on 18 August 1876.

Acknowledgements

I would like to acknowledge the work of the OHPR team, Tony Franklin, Hervé L’Hours, Victoria Holmes, Anja Nieth, Juergen Neuhoff, Steve Warin and Ole Wiedenmann, for their work on the project and for their assistance in the writing of this article.

NOTES

1. For further details of the JISC Digitisation Programme see: http://www.jisc.ac.uk/index.cfm?name=digitisation_projects (available 13 February 2007).
3. See, for example, Nineteenth annual report of the registrar-general (1856) BPP 1857–58, XXIII (2431) 200–1 and Twenty-fourth annual report of the registrar-general (1861) BPP 1863, XIV (3124), 227–9.
4. Thirty-seventh Annual report of the registrar-general (1874) BPP 1876, XVIII [C.1581], 224.
5. Treasury Committee to inquire into questions connected with taking of census report. Minutes of evidence, appendices BPP 1890, LVIII [C.6071]; Committee on Bill for taking Account of the population of Great
6. OCR or Optical Character Recognition is the application of a software tool to convert images containing text into machine-readable texts.


8. Return of expense for census of GB, 1841 and 1851, BPP 1854, XXXIX (442); Return of expense for census of England, Wales, Scotland and Ireland, 1871, BPP 1875, XLII (377).

9. Census of the British Empire, 1901. Report with summary and detailed tables for the several colonies, &c., area, houses and population; also population classified by ages, condition as to marriage, occupations, birthplaces, religions, degrees of education, and infirmities, BPP 1905, CH [Cd.2660].


11. Letter of Registrar General relative to complaint against certain remarks in census report of 1891 as regards inhabitants of Wales, speaking Welsh only, BPP 1894, LXIX (331), 1.

12. See: E. Higgs, Life, death and statistics: civil registration, censuses and the work of the General Register Office, 1837–1952 (Hatfield, 2004), 221–8. These ‘Registrar-General’ versions are not acknowledged by a number of writers: for example, R. Woods and N. Shelton, An atlas of Victorian mortality (Liverpool, 1997), state on page 155 that the Tenth ARRG was ‘very short, only uses counties, nothing on cause or age of death...’: The web site contains a report of 375 pages which we chose to cite as: General Register Office, Tenth annual report of the Registrar-General for 1847 (London, 1852) and General Register Office, Tenth Annual report of the registrar-general (1847), BPP 1849, XI.563 [1113].


14. The curious may have noted a mismatch with the number of images mentioned earlier. For each digital reproduction of a paper original there are six digital versions on the site. Five are small, compressed PNG files which can easily be displayed across the internet; the other is a copy of the master image which is usually much larger. These TIFF images are on average about 2 MB each, though the largest file which can be downloaded by a user of the site is in the region of 103 MB.