Historical Infrastructure Dataset Pilot: Transport in Britain,1870-I937.

Prepared by Dr Oliver Dunn (Cambridge University Technical Services) for the National Infrastructure Commission

See data file with curated data set: NIC_Data_GB.xls

March 2020

Contents

1 SUMMARY

This research report describes the methods and archival sources used to create a new curated historical data set giving variables for transport infrastructure and services in Britain between 1870-1937, such as UK totals for rail track length and rail passenger numbers. This pilot project was undertaken by Dr Oliver Dunn (Post-Doctoral Research Associate at the Faculty of History, University of Cambridge) working as consultant with Cambridge Technical Services Ltd (Cambridge University's consultancy services arm) under the supervision of Elizabeth Horsman (NIC) and David Menzies (NIC).

2 INTRODUCTION

The following pilot brief was as advertised:

> *'The National Infrastructure Commission wishes to develop its economic infrastructure dataset, by complementing the data it currently uses or has developed, with a curated historic dataset. The Commission believes that there is a range of relevant data on the evolution of the UK's infrastructure system available, likely housed in many different historical texts and archives.*
>
> *The Commission believes this will be of benefit to the interested public and assist in our understanding of how the performance of, and investment in, infrastructure has evolved over time, providing a perspective on current performance and levels of investment.'*

Responding to this brief, the aim was to provide historical data covering key aspects of major transport systems in Britain. Sources were identified and data was collected to create annual time series for each transport category with multiple variables covering different transport systems. It is expected that at least some data will be comparable with the NIC's existing historical transport data sets and performance measures, for instance passenger kilometres travelled, and ton kilometres travelled (freight).

Copies of historical sources were digitized, and the data then extracted from the images into the data file NIC_Data_GB.xls. The resulting data set covers England, Scotland, Wales and, in some early cases, Ireland, and was collected in spreadsheets for rail, trams, light rail, roads, road vehicles, and air travel (NIC_Data_GB.xls).

3 METHODS

3.1 SOURCES CONSULTED

The sources consulted for this project were:

1. Mitchell, B. R. *British Historical Statistics.* Cambridge: Cambridge University Press, 2011. Reprint
2. Several printed editions of *United Kingdom Statistical Abstracts* (hereafter SAUK) published annually under Parliamentary authority since the nineteenth century.

These sources provided all the data in NIC_Data_GB.xls.

Transport data for the UK has been published in print by Brian Mitchell (who published tables covering other national-level economic statistics in the same book). This curated data

in print form formed the basis for further data collection from the SAUK. These original documents provided data to extend the time series given by Mitchell and to create entirely new data covering trams and light rail. Professor Mitchell is an acknowledged authority on historical statistics for the UK and so his work presented a convenient place to start this data collection. To give the reader a sense of Mitchell's printed data format, Figure 1 shows a snippet from his table for railways. This table was used to create the series in NIC_Data_GB.xls:

Figure 1 Railway mileage table from Mitchell (Mitchell, B. R. British Historical Statistics. Cambridge: Cambridge University Press, 2011. Reprint.)

Transport and Communications 5. Railway Length in Miles – British Isles 1825–1980

NOTES
[1] SOURCES: 1825–43 – H. G. Lewin, *Early British Railways* (London, 1925); 1844–1980 – *Railway Returns, Abstract, and Statistical Abstract of Ireland.*

[2] Statistics relate to route length, irrespective of the number of tracks, at 31 December (except as indicated).
[3] Narrow gauge lines are not included.

	Great Britain	Ireland		Great Britain	Ireland
			1870	[13,562](b)	[1,975](b)
1825	26·75		1871	13,388	1,988
1826	38		1872	13,723	2,091
1827	40·75		1873	13,981	2,101
1828	44·75		1874	14,322	2,127
1829	51				
			1875	14,510	2,148
1830	97·5		1876	14,715	2,157
1831	140		1877	14,874	2,203
1832	166		1878	15,074	2,259
1833	208·25		1879	15,411	2,285
1834	298				
			1880	15,563	2,370
1835	337·75		1881	15,734	2,441
1836	403·25		1882	15,992	2,465
1837	540·25		1883	16,179	2,502
1838	742·5		1884	16,339	2,525
1839	969·75				
			1885	16,594	2,575
1840	1,497·75		1886	16,700	2,632
1841	1,775		1887	16,904	2,674
1842	1,938·75		1888	17,079	2,733
1843	2,043·75		1889	17,152	2,791
1844	2,148 (a)				
			1890	17,281	2,792
1845	2,441 (a)		1891	17,328	2,863
1846	3,036 (a)		1892	17,430	2,895

Mitchell's tables, such as Table 5 shown in Figure 1, were compiled using original company 'returns' made by transport operators that managed transport services across the UK. They reported the information to the government on an annual basis over the historical period covered by this data.

Mitchell states he is confident as to the overall reliability of the data contained in the returns because central government was able to oversee a limited number of companies across the UK effectively at the time (Mitchell 2011 p. 532).

Mitchell did not, however, include all data that the government originally collected for the period covered by the pilot. For example, Trams and Light Rail data were reported from 1870, and Civil Aviation from 1919 to 1937, but Mitchell did not include these data. We collected new additional data for this system, and extended coverage of data that Mitchell did collect (e.g. rail). We did this using the SAUK

Combining these two sources of data was possible because the SAUK were originally also compiled from company returns, the same source Mitchell used in his book. The data was

compiled and published on an annual basis by HM Stationary Office as an official journal publication. Originals survive and were consulted in the Cambridge University Library and on Parliamentary Papers Online. Figure 2 is a snippet from a table in the SAUK journal covering roads.

Figure 2 Table with road mileage from SAUK 1923-1937 (Board of Trade (in conjunction with the Ministry of Labour and the Registrars-General), Statistical Abstract for the United Kingdom…Vol. 83. London: HM Stationary Office, 1940. Print.)

ROADS—GREAT BRITAIN
No. 247—CLASSIFICATION of ROADS and SCHEDULE of MILEAGE

	Class I.—Roads		Class II.—Roads		Other Roads		Total Mileage
—	Mileage	Per cent. of Total	Mileage	Per cent. of Total	Mileage	Per cent. of Total	
TOTAL : GREAT BRITAIN :							
1923–4 … … … …	23,230	13·1	14,739	8·3	139,352	78·6	177,321
1924–5 … … … …	24,048	13·5	14,638	8·2	138,996	78·3	177,682
1925–6 … … … …	24,329	13·7	14,930	8·4	138,947	77·9	178,206
1926–7 … … … …	24,552	13·8	15,625	8·8	138,185	77·4	178,362
1927–8 … … … …	25,112	14·0	15,683	8·8	137,942	77·2	178,737
1928–9 … … … …	25,528	14·3	15,747	8·8	137,820	76·9	179,095
1929–30… … … …	25,996	14·5	15,805	8·8	137,485	76·7	179,286
1930–1 … … … …	26,417	14·9	15,924	9·0	134,915	76·1	177,256
1931–2 … … … …	26,513	15·0	16,482	9·3	133,796	75·7	176,791
1932–3 … … … …	26,585	15·0	16,644	9·4	134,118	75·6	177,347
1933–4 … … … …	26,663	15·0	16,774	9·4	134,385	75·6	177,822
1934–5 … … … …	26,779	15·0	16,837	9·4	134,891	75·6	178,507
1935–6 … … …	27,015	15·2	16,855	9·4	134,233	75·4	178,103
1936–7 … … …	27,142	15·1	16,930	9·5	134,832	75·4	178,904
1937–8 :							
ENGLAND AND WALES :							
Administrative Counties { (a)	3,271	2·3	} 12,185	8·8	106,723	77·5	137,656
{ (b)	15,477	11·2					
County Boroughs … …	1,587	11·5	760	5·5	11,404	82·9	13,751
Metropolitan Boroughs, City of London and L.C.C.	292	12·4	125	5·3	1,933	82·3	2,350

Because we used two separate sources, it was important to ensure data from both Mitchell and the SAUK collection are comparable. This was especially the case where individual time series used data drawn from both sources.

Data from both sources matched up well when arranged together as a time series, i.e. without any apparent obvious discrepancies in the data taken from both sources. The original variable titles in printed tables indicate where data was collected on a similar, if not identical, basis originally by officials. It appears the data contained within Mitchell and the SAUK are comparable.

Because the government seems to have collected data on a standard basis, it would be possible to extend the series using more SAUK in the future.

3.2 DATA COVERAGE
For this project the aim was to provide both an accurate and large dataset with good temporal and geographical coverage.

Data coverage varies by network according to the nature of the sources but the main bulk of data covers 1870 to 1937. This dataset covers Great Britain, including England, Wales, and Scotland. The modern-day territories of Northern Ireland and the Republic of Ireland were not included in this pilot data set. Ireland was recorded separately in the sources before Irish independence in 1919 and it was possible to omit the Irish data. A small amount of Irish data for Tramways and Light Rail 1876-1912, was however necessarily included here because it was mixed in with the data for the rest of the United Kingdom and was not recorded separately at this time, and therefore we could not easily disaggregate it from data covering England, Scotland and Wales.

With more work, we could remove the Irish trams and light railways data to make it more comparable with the data for the period after 1912 when Ireland was not included.

3.3 DATA COLLECTION METHOD

Most of the data was collected using Optical Character Recognition (OCR) software. OCR allows much more data to be transcribed more accurately compared to manual data entry. Where we could not use OCR, we transcribed data using standard manual data entry. Manual checks of all data were made.

To allow OCR, high definition scans and photographs were made from all printed sources, including Mitchell (see figures 1-3). The images were obtained from Cambridge University Library. Some original copies of surviving SAUK were found to be unavailable and did cover the entire period we aimed to cover. Substitute copies were found instead online at *Parliamentary Papers Online* (ProQuest).

Figure 3 OCR As Fig.2 with OCR overlay indicated by bounding box in 'Mileage' column.

					Mileage
TOTAL : GREAT BRITAIN :					
1923–4 ...	...	...	...	23,230
1924–5 ...	...	...	...	24,048
1925–6 ...	...	...	...	24,329
1926–7 ...	...	...	...	24,552
1927–8 ...	...	...	...	25,112
1928–9 ...	...	...	...	25,528
1929–30...	...	...	...	25,996
1930–1 ...	...	...	...	26,417
1931–2 ...	...	...	...	26,513
1932–3 ...	...	...	...	26,585
1933–4 ...	...	...	...	26,663
1934–5 ...	...	...	...	26,779
1935–6 ...	...	...	...	27,015
1936–7 ...	...	...	...	27,142

Scans were made of the images and these were enhanced using software so OCR could be used to extract the data. This can be done with around 90% accuracy. Manual and global corrections were then made for the remaining 10%. Due to the poor quality of scans available from ProQuest, manual data entry was required to transcribe these. Manual checks were finally made of all data to ensure accuracy. Figure 3 shows the OCR overlay on the road 'Mileage' column from an enhanced image (also shown in Figure 2).

3.4 DATA STRUCTURE

The data was entered into Excel spreadsheets separated by tabs for transport sectors. The original printed table structures found in the sources were replicated in the spreadsheets to maintain a link to the original meaning of the tabulated information. The new column headings and notes in the spreadsheets match those found in the sources for the same reason. The aim was to recreate the original source formats and not introduce too much reinterpretation of the original data.

Original table notes were copied over and arranged to cover rows for individual years or as general notes covering the whole table. The sources used for each column or data are given below the relevant column at the bottom of the new table. A code (1-7) is used to refer to original sources, which is explained by a key in the final tab. There are eight sources in total. Where Mitchell was the source, we cite him *and* the original sources he used to compile his data to maximize the value of his references in the data.

Where figures are unavailable, e.g. those dating from the First World War years 1914-1918, we have marked this in merged cells. Other gaps in the data result from the fact that certain variables were unavailable in early sources and only emerged as government statistics became more detailed over time.

4 RESULTS

4.1 RAILWAYS

Data for rail was taken from tables in Mitchell 2011. These data originate from the *Railway Returns*, collected from railway companies. Mitchell believes the statistics to be reliable, accepting, as noted in the data, minor changes to the way the statistics were collected in 1913, 1928, 1933 and 1948 (Mitchell 2011, p. 532). These data are useful for understanding the extent of Britain's railways and traffic as well as safety on trains and in stations, working conditions, and railway company finances and aggregate national investment trends.

Railways variable types:

1. Railway length in miles
2. Paid up share and loan capital (on £ million)
3. Receipts passenger trains (£)
4. Receipts goods trains (£)
5. Total working expenses (£)
6. Train mileage (passengers)
7. Train mileage (freight)
8. Passenger journeys
9. Freight loaded (tons)
10. Passengers killed
11. Passengers injured
12. Servants of companies and contractors killed
13. Servants of companies and contractors injured
14. Number of locomotives

4.2 TRAMWAYS AND LIGHT RAIL

The tramways and light rail data included in this pilot data collection project were not covered by Mitchell at all, perhaps because of their smaller scale compared to railways. Trams were a popular form of urban transport and provided a very important service for passengers. Light railways were used to move people and goods short distances. The data for these were collected here entirely from original sources (listed below). No data was ever given for freight carried on either trams or light rail.

Tramways and light rail statistics appear alongside those for railways in the SAUK. Because they were collected in a similar way to railway statistics, an acceptable level of accuracy and reliability of tramway and light rail data can be safely assumed.

Unfortunately, London's tramways network was removed from the published statistics after 1929 and could not be included in the data after that date. With more research this network could be put back into the data using different sources. A small number of Irish data for the limited number of trams in that country is mixed in with the UK statistics to 1912. In addition, SAUK statistics to 1898 did not include data for light railways. This is probably because few existed.

Tramways and Light Rail variable types:

1. Route miles open for traffic
2. Paid-up capital of tramway undertakings (£)
3. Number of passengers conveyed

4. Revenue receipts of all tramway undertakings (£)
5. Expenditure of all tramway undertakings (£)

4.3 ROADS

UK road statistics were collected by the government from the 1920s, when road management was centralised (Mitchell 2011, pp. 532-533). For statistics dating before this, information is scattered across the archives of local authorities and more research would be needed to locate these data. Road data were taken from Mitchell's tables. Mitchell gives very little information about the road types and in fact historians know surprisingly little about the quality of Britain's roads at this time, given the importance of roads to the wider economy.

Road Variable types:

1. Trunk roads (miles)
2. Class I roads (miles)
3. Class II roads (miles)
4. Class III (miles)
5. Unclassified roads (miles)
6. Total roads (miles) (original totals)

4.4 ROAD VEHICLES

Data was collected from 1904 on motor vehicles and was derived from Mitchell 2011. A very large number of vehicle types are covered. Over the period covered by this pilot data set, automotive technology advanced dramatically. The number of trucks shown in the data increases exponentially over this time period. This indicates how trucks became important as a way of transporting goods around Britain. This trend depended on improvement in the efficiency of internal combustion engines, improvement of road surface quality, and geographical coverage and connectivity with towns and cities.

Vehicle variable types:

1. Private cars
2. Motorcycles
3. Buses, coaches and taxis
4. Taxis
5. Goods vehicles
6. Other (except trams)
7. (Original) Total (except trams)
8. Tramcars
9. Road accidents fatal
10. Road accidents non-fatal (so far as reported)
11. Total road licence receipts (motor vehicles only) (£)

4.5 CIVIL AVIATION

One novel aspect of this data set is its inclusion of air traffic data from 1919, when civil aviation began in the UK. British Civil Aviation refers to the forerunners of companies like British Airways, i.e. private companies operating domestic and foreign services between London and cities including Paris and Leeds. In 1937, the statistics for flights are described as connecting the UK with countries including the USA, and the continents of Africa and East Asia. Unfortunately, the original tables are not very clear, and more research will be needed to clarify which routes were covered up to what time.

Like trams and light rail, the data were copied from the SAUK. Air traffic was closely regulated and there were few operators, meaning the government was likely furnished with accurate and reliable information by aviation companies via the Air Ministry. These statistics only relate to British operators of aircraft and not to aircraft belonging to other countries. Some of the early aviation statistics collected originally by the Air Ministry are conflicting and ambiguous, and some conflicting data had to be omitted.

These data are useful to demonstrate the exponential growth of early civil aviation but also how safe it was compared to railways, as shown by the accident statistics. More information about global destinations could reveal the expanding network of air routes for the 20th century. More data can be collected from Mitchell after 1937 to continue the Civil Aviation series.

British Civil Aviation variable types:

1. Aircraft miles flown
2. Number of passengers carried
3. Weight of cargo carried (tons)
4. Fatal accidents
5. Non-fatal accidents
6. Crew killed
7. Crew injured
8. Passengers killed
9. Passengers injured

5 RECOMMENDATIONS FOR NEXT STEPS

This pilot data can be extended and improved in two main ways: extending the time series for transport categories forwards and backwards in time and adding further contextualising qualitative data for individual transport modes.

Mitchell includes more rail data going back to 1825 that could be added to the current rail time series. For all other categories, the earliest data available in these sources has already transcribed, but all the time series could be extended forward, for example data for Civil Aviation and railways. Mitchell's data for motorways could also be added to Roads. However, SAUK variable types and those in Mitchell change over the twentieth century, meaning some more work will be needed to make it comparable with the current data.

This pilot dataset covers a very long period. The extensive coverage and project scope meant we could not provide all the historical context that might accompany the data. Technology changed significantly over the time period covered for this pilot, however the dataset doesn't reveal very much about this change. Questions might include, which trains are covered by the category 'Locomotives'? How efficient were they and were diesel trains included? What kind of roads are represented by the various classifications of roads used in the sources (class I, II, II etc.)? What about the large proportion of recorded 'unclassified' roads: were these simply muddy tracks? Primary research would be required to clarify some aspects of the data.

Mitchell tells us the government statistics that form the basis for this data set are essentially reliable. However, more research using the literature and original sources could be usefully undertaken to investigate remaining ambiguities, for example how data for Civil Aviation was collected.

Another entirely separate transport system is omitted by the current pilot data. Coastal shipping was a very large system for internal communications during the era of the steam ship. From the 1830s to the 1960s passenger and freight services were very large in scale. Coastal transport was a competitor of railways but was also highly integrated with other modes like rail. Ample data exists for coastal services and passenger numbers and could well be added to this data set taken from the SAUK.

Motorways were built from the 1960s, and new aviation routes connected the UK with other parts of the world in previously unimaginable ways. Trams and railways became less important over time and coastal shipping disappeared entirely. One can envisage a larger historical dataset that captures all these ebbs and flows of our major transport networks over the last 200 years by collecting more data from the sources found for this pilot project.

6   REFERENCES

1. HM Stationary Office, *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1871 to 1885. Thirty-Third Number.* London: Printed by Eyre and Spottiswoode, 1886. Print.
2. HM Stationary Office, *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1887 to 1901. Forty-Ninth Number.* London: Printed by Wyman and Sons, 1902. Online, accessed 12 March 2020.
3. HM Stationary Office, *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1898 to 1912. Sixtieth Number*. London: Printed by Eyre and Spottiswoode, 1913. Online, accessed 12 March 2020.
4. Statistical Department, Board of Trade, *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1910 to 1924.* Vol. 69. London: HM Stationary Office, 1926. Print.
5. Board of Trade (In conjunction with the Ministry of Labour and the Registrars-General), *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1913 and 1922 to 1935.* Vol. 83. London: HM Stationary Office, 1937. Print.
6. Board of Trade (In conjunction with the Ministry of Labour and the Registrars-General), *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1924 to 1938.* Vol. 83. London: HM Stationary Office, 1940. Print.
7. HM Stationary Office, *Statistical Abstract for the United Kingdom. In each of the last fifteen years from 1883 to 1897. Forty Fifth Number*. London: Printed by Eyre and Spottiswoode, 1886. Print.
8. Mitchell, B. R. *British Historical Statistics.* Cambridge: Cambridge University Press, 2011. Reprint.
    a. For all railways data, Mitchell cited as his sources: 1844-1980: Parliamentary Railway Returns; To 1913: Railway Returns, Capital formation - B. R. Mitchell, "The Coming of the Railway and United Kingdom Economic Growth, Journal of Economic History, xxiv, 3 (1964); 1843-1938 Railway Returns, *Transport Statistics Great Britain*, and, Abstract, whichever provides the latest revisions.
    b. Roads: 1922-7 -Annual Reports of the Road Fund; 1928-80 – British Road Federation, Basic Road Statistics and Abstract.
    c. Vehicles: 1904-20 (except trams) - Society of Motor Manufacturers and Traders,
        Taxation Motor Industry of Great Britain 1935; 1921-5 (except trams) - Return of Motor Taxation; 1926-38 (including trams from 1933) - Annual Vehicle

Census 1904-32; - Annual Return of Tramway and Light Railway (Street and Road) Undertakings. (trams) - Annual Return of Tramway and Light Railway (Street and Road) undertakings.

d. Air: Compiled from Returns furnished by the Air Ministry.