

Since the Census Helps Plan Infrastructure and Housing, Could a National Framework for Data Help Overcome the Shortcomings of the COVID-19 Census?

*Mark Coates, International Director of Public Policy and Advocacy, Bentley Systems
Dave Philp, Chief Value Office, Cohesive*

What comes to mind when you think of the census? Or rather, when comes to mind?

Many people think of it as a historical document. For them, it might conjure images of rapidly growing industrial cities. There is some justification for that—the census has been part of our lives for a very long time. The first census of England and Wales was taken in 1801, partly to accurately record the number of men who might fight in the Napoleonic Wars.

Since that time—and particularly in the digital era—we have seen huge improvements in our ability to collect, store, analyse, and present data. Yet, the census remains important not just for historical reasons, but also for modern-day planning. It provides a large and consistent dataset on a range of important topics, which can be built up to national level or broken down into tiny constituent parts. Information from the census helps the government and local authorities plan and fund new transport infrastructure, as well as services, such as education and health.

For example, the figures that came out on December 8 showed how people in different parts of England and Wales [get to work](#).

Previous iterations have provided enormously valuable insight into patterns of public transport use, gaps in provision, places that are most dependent on cars, and the relative importance of different modes in different areas.

These key insights have been vital in correcting an occasionally London-centric narrative, which fails to understand, for example, how crucial buses are outside the capital, or just how much other areas are reliant on cars as their only realistic option for getting to work.



Because this is census data, sample sizes remain perfectly usable, even when we drill down to tiny areas for hyperlocal insight, making it useful for local and national planning.

We also know that the census is measuring the same thing in the same way—whether in Cardiff, Carlisle, or Canterbury. There are no gaps. North and south, rural and urban, city and village—everywhere is measured.

When we then combine this data with other consistent national datasets—on deprivation levels or demographic characteristics like race and age—we can understand what is happening on multiple levels.

There are other datasets on public transport use, but they often have limitations that make them far less usable. They might only cover a specific geography, such as Greater London or Greater Manchester, or they might not measure all modes. They might use different methodologies so that it is impossible to put together to form an accurate picture: a jigsaw where the pieces don't fit.

So, the census data remains vital. And yet, sadly, the transport data in the 2021 census was compromised by COVID-19. Census Day was March 21, when more than 4 million workers were on the government's furlough scheme and with millions more working from home. Non-essential retail remained closed and social gatherings were still restricted—as was public transport.

The effect on the data was that the data was reduced to the level of historical document—a snapshot of life under COVID-19, but little more. It tells us little to nothing about the underlying shifts in public transport and car use since the previous census. It does not help us form future-focused policy.

For example, 31.2% of working adults told the census takers they worked "mainly from home." While it may have been true in March 2021, it wasn't true before COVID-19 and it isn't true now. In fact, separate data collected by the Office for National Statistics (ONS) shows the number of people working from home had fallen to below 15% by April this year.

Of those who said that they did commute to work, 65.6% said that they drove a car or van. At the time of the last census, in 2011, that figure was 60.8%. While we might conclude that we have failed

on public transport, and that more people indeed been choosing to use the car, we simply cannot make that assumption given the circumstances. When the census was taken, many people either could not use public transport—because it wasn't available—or chose not to for reasons that no longer apply—because they were worried about their own health and the health of others.

Another census taken now would likely give very different results. However, we do not take a second census. Instead, we have to try to piece together a far fuzzier picture based on different datasets [on car use](#), [bus use](#), [train use](#), and [walking and cycling](#).

It isn't just commuting data that was skewed by the pandemic. COVID-19 affected all parts of our lives and, consequently, all parts of the census. Even the most basic data—on where people live—was tainted.

For example, the census figures for 2021 show that several London boroughs had significantly lower populations than in 2011. Camden was down from 220,338 to 210,100, Westminster from 219,396 to 204,300, and Kensington and Chelsea from 158,649 to 143,400. Those figures were very different from what statisticians believed the situation to be before COVID-19 struck. Population estimates provided by the ONS suggested Camden, for instance, had a population of 279,516 in 2020.

Were those estimates wrong? Almost certainly not. Rather, the census was taken at a time when tens of thousands of people were working from home or left London and other cities to live more cheaply somewhere else. It is reasonable to assume that, as people have returned to offices, they have also return to live in cities.

A second census isn't a realistic option. Indeed, there has been talk that the 2021 census might be the last. One reason for this idea is cost. While the 2011 census cost GBP 482 million, the 2021 edition was said to have cost in the region of GBO 900 million. That isn't the whole story, as the sum includes a major modernisation process at the ONS, which will make its data more widely available and relevant. But the fact remains that conducting a census is not cheap.

It is also not necessarily the best way of achieving its very important ends. As our data literacy improves, as more data becomes easily accessible, the census already looks like a cumbersome

anachronism. COVID-19 taught us the importance of timeliness in data—and information from the census, which is conducted every 10 years, can hardly be said to be timely. How useful is it to know commuter patterns from 2011 when setting transport policy in 2015 or 2020? And should we have to wait until 2031 to know conclusively what impact our decisions have had?

The key data challenges we face now have relatively little to do with collection, or storage, or analysis. Rather, they have to do with our national data architecture—the tedious but vital work of planning what we want to collect and why, as well as of making sure individual acts of data collection add meaning to a deeper pool.

A local authority collecting live data about public transport use is valuable to that local authority. All local authorities collecting the same data in the same way is infinitely more valuable to national and local decision-makers alike. It does not have to be costly or labour intensive, as much of the most valuable data is already being collected by the private sector, if not the public.

However, this data is not always shared, and it is not always recorded in a way that is commensurate with other datasets. Of course, private companies—even those that are paid for or subsidised by the public purse—should not be expected to publish genuinely commercially compromising data. Still, there is a great deal that we could and perhaps should expect them to publish.

A national framework for data—with consistent definitions, categories, quality standards, and protocols for sharing and storage across central and local government—might be a good starting point.

A country could leave it up to individual towns and cities to connect themselves, build their own train tracks, and create their own designs. However, if you want a true rail network, you must make sure that all the tracks are of the same size and type.

We have made great strides in harnessing the power of individual datasets. Now, the real power of data lies in bringing them together. If 2021 is indeed to be the final census, perhaps its legacy will be to spur us to a better approach.