Written evidence from the Greater London Authority (GLA) and the London Office of Technology & Innovation (LOTI) at London Councils (DTA 24)

Public Administration and Constitutional Affairs Committee
Data Transparency and Accountability: Covid 19

The following response to the Public Administration and Constitutional Affairs Committee’s Call for Evidence on Data Transparency and Accountability during Covid-19 represents the views of the Greater London Authority (GLA) and the London Office of Technology & Innovation (LOTI) at London Councils.

The response combines the direct and relevant experience of the GLA and LOTI in responding to the Covid-19 pandemic.

For further information on any of the points outlined below, please contact: Theo Blackwell, Chief Digital Officer for London, GLA; and Eddie Copeland, Director, LOTI.

Introduction
As signatories and co-founders of the MHCLG Local Digital Declaration we are committed to designing safe, secure and useful ways of sharing information to build trust among our partners and citizens, to better support our communities, especially the most vulnerable, and to target our resources more effectively.

Response

1. Did Government have good enough data to make decisions in response to Coronavirus, and how quickly were Government able to gather new data?

The pandemic has revealed a number of problems with Government data which have led to disproportionate death and suffering by some of the most vulnerable residents in our society. Covid-19 has proved not to be the ‘great leveller’ that was first thought by government. These issues need to be addressed to improve future resilience both in response to further waves of the pandemic that we are seeing now, and to address future major shocks. Outlined below are two key challenges faced by the GLA and LOTI during our response to the pandemic.

Firstly, disease surveillance reporting was inadequate to identify and prevent the spread of Covid-19 until it was too late -- the disease had already spread across the entire London region and deaths were increasing exponentially before a lockdown was introduced. Even where data was being reported this was partial - for instance when early test data started to be released it
became apparent that key test centres were missing from the data, providing a misleading picture of the spread across London.

A second challenge was the lack of detail in reported data to enable the identification of disproportionate impact among population groups. We focus on ethnicity below, where the first evidence of impact was identified in April by ICNARC who were studying the impact of Covid-19 on intensive care service. However, the GLA’s Rapid Evidence Review [https://data.london.gov.uk/dataset/rapid-evidence-review-inequalities-in-relation-to-covid-19-and-their-effects-on-london](https://data.london.gov.uk/dataset/rapid-evidence-review-inequalities-in-relation-to-covid-19-and-their-effects-on-london) has identified a disproportionate impact on a much wider range of communities, many of which are not measured in official government reporting.

The lack of recording of ethnicity data for Covid-19 cases and Covid-19 deaths limited the Government’s ability to identify the disproportionate impact of Covid-19 by ethnicity. We welcome the announcement by the Minister for Equalities in the first quarterly report to the Prime Minister and Health and Social Care Secretary on progress to understand and tackle Covid-19 disparities experienced by individuals from an ethnic minority background, that recommendations included mandating the recording of ethnicity data as part of the death certification process, as this is the only way to establish a complete picture of the impact of the virus on ethnic minority groups.

In 2003 the GLA, London Health Observatory and DH commissioned a report to respond to the government consultation on ethnic data collection at birth and death registration. The report is here: [https://kar.kent.ac.uk/7772/1/Aspinall_MissingRecord_July_2003.pdf](https://kar.kent.ac.uk/7772/1/Aspinall_MissingRecord_July_2003.pdf)

The report argues for recording ethnicity data at birth and death registration:

‘When stratified by ethnic group the burdens (incidence, prevalence, and mortality) of many diseases are known to vary. There are well documented inequities in access to preventative, treatment, and palliative health and social care services based on ethnic group. There are, too, reported differences in the quality of services received across the different ethnic groups and of outcomes of treatment and care. Many of these inequities are amenable to change. However, in order to address them they must, first of all, be comprehensively defined and documented. Mainstreaming ethnic monitoring/data collection is a vital step in the process. The history of such data collection in the NHS is poor, whichever of the key datasets is examined: hospital episode statistics, general practitioner data, cancer registrations, and disease registers. While steps are now being taken to remedy some of these deficiencies, the continued non-availability of ethnic monitoring data and in some cases of compatible ethnically-coded denominator data remains a problem. In particular the lack of ethnic group in births and deaths data has been the subject of widespread comment by specialists in demography and public health
and is probably the single action that could most improve the evidence based for addressing ethnic/racial inequalities in health and health care’.

It’s vital to have good quality data on both ethnicity and other groups impacted by health inequalities, in order to monitor the disproportionate impact of Covid-19 and assess the effectiveness of measures to reduce disproportionality. This links to the points made in question 3 and 4 regarding improved transparency in regional reporting and dissemination of data.

By reporting figures nationally, this can downplay the significance of factors such as ethnicity given the different demography of these groups and their concentration in certain cities and regions.

We would like to note that we welcomed the excellent work of the Office for National Statistics (ONS) in making a much wider range of data and reporting available at speed. Good examples include the matching of death registration and Census records to quantify disproportionality by ethnicity and other factors, and the work to publish google data on mobility. Also the production of new surveys, datasets and bulletins to help track the socio-economic impacts of the pandemic.

2. Was data for decision making sufficiently joined up across Departments?

There was and still continue to be issues in consistent reporting of cases and deaths data from different sources - NHS/PHE and ONS. For instance this prevented an understanding of the extent to which Covid-19 had affected care homes and other institutions such as prisons, at an early stage of the pandemic.

3. Was relevant data disseminated to key decision-makers in: Central and Local Government; other public services (like schools); businesses; and interested members of the public?

Throughout the crisis, there has been a strong sense that local authorities and other local public services have consistently been omitted from central Government’s initial thinking on designs for data sharing.
This has manifested itself in challenges related to shielding lists, volunteering, testing data and tracing of complex cases, plus difficulties in accessing relevant data about people who are furloughed or economically vulnerable. And also the need for bodies such as the GLA to publish a wide range of regional reporting to provide greater transparency to stakeholders such as the media, civil society and the public (covered in question 6).

For example, in the case of shielding lists, different lists were being sent to councils and Clinical Commissioning Group (CCG) leads with explicit instructions for them not to be shared, even though many CCGs share chief officers with councils. Similarly, councils with responsibilities for splitting out data for other councils were not told they had this responsibility for some days after they had started to receive the data.

Data on testing and tracing capacity, activity and results has been improving, but published data at the national England level is not all available for regions, for example testing turnaround times, distance to test, ethnicity data is not available.

Data about levels of activity (for instance, footfall, trips and spend) was made available to central Government (or was offered) by a number of private sector organisations. It was not known or easily discoverable what part of government had secured this data or on what terms. The data was also not made available to local government and had to be purchased independently by London government, including from O2 and Mastercard. This type of data, if made available earlier, could have better informed local messaging and planning for social distancing measures in public areas.

Currently some organisations are charging for these data and others are continuing to make data available free of charge (for instance the Purple WiFi company whose data we use in the GLA’s mobility report is provided free of charge). There is a need for a more consistent approach to private data that can be used for the public good.

4. Were key decisions (such as the “lock downs”) underpinned by good data and was data-led decision-making timely, clear and transparently presented to the public?

The spread of Covid-19 was and continues to be uneven. The focus on national reporting for England has made it harder to communicate key regional messages in good time and transparently present the rationale for key decisions.

For London, this has led to the need for much greater reporting by the GLA to fill information gaps and provide a comprehensive picture. To support demand from local media we produced our own analysis of daily cases and deaths bringing together the different data sources to provide a transparent and consistent analysis for London. These include a daily summary and more
detailed daily analysis on the London Datastore. These more detailed daily pages have been the most accessed data on the London Datastore for the last 6 months. 

After the introduction of lockdown, media coverage of London parks led to concern that there was widespread non-compliance with these measures. The data we brought together for our mobility report was used by both government and media to understand this issue in more detail, https://data.london.gov.uk/dataset/coronavirus-covid-19-mobility-report

At the start of the pandemic there was a lag in publication of statistics on job support schemes such as the Coronavirus Job Retention Scheme and the Self-Employment Income Support Scheme. In order to understand impacts on London’s economy, we had to rely on ONS survey data until administrative data from HMRC started being released from June onwards. Although release of these data is very welcome and is now regular, it took a few iterations for releases to include the level of information required to properly understand regional impacts, and there has been a degree of inconsistency in content and presentation of successive releases.

5. Was data shared across the devolved administrations and local authorities to enable mutually beneficial decision making?

There have been a number of issues with the data shared with London local authorities for the purposes of infection and outbreak control, and for Test and Trace. Issues with each individual dataset are outlined in detail below. Continually having to work around these issues has significantly added to the time local authorities have had to invest in accessing, analysing and drawing insights from data. The result has been that interventions to support vulnerable residents have been delayed.

Simple good practice such as publishing the required data descriptions and metadata and facilitating access to the data via API rather than a spreadsheet export could have sped up response times and reduced the risk of error. These issues should be addressed now as a matter of urgency, as the risk associated with inaction increases with a growing infection rate.

The following issues were flagged with LOTI and the GLA in their work with local authorities between March and September 2020:

**Public Health England Line List - Infections Data**

- **Metadata missing:** Missing metadata and insufficient data descriptions have made it harder for analysts to understand this data in the context of other government datasets such as the public infections data published to https://coronavirus.data.gov.uk/
Not possible to automate data extraction: Boroughs report that automating the data pull from PowerBI is not possible. The current system, which forces the use of spreadsheets, increases the risk of error and data loss during the download and transformation process. A particular risk is that PHE’s PowerBI dashboard has a row export limit.

Shielding Patient List Data - Published by NHS Digital via NHS SEFT portal

Not possible to automate data extraction: Accessing this data still requires a time-consuming manual process that increases the chances of data errors during transformation. If the list moves back to daily publication this could cause a bottleneck to accessing the data, slowing down the time to intervention.

Lack of alignment with legacy shielding data (GDS SPL): Councils have built up a more comprehensive picture of shielded individuals through the original shielding period. Since the transfer of publishing responsibility shifted from the Government Digital Service (GDS) to NHS Digital, it has become difficult to match current records up with the legacy GDS Shielding Persons List (SPL). Following internal analysis in multiple boroughs, it is evident that the NHS SPL data misses out individuals from the original lists. As there is low trust in the data quality, boroughs are worried that individuals may be being removed in error.

Poor data quality: As with the GDS Shielding Persons List, boroughs have reported data quality issues, including very old contact details/phone numbers, people included on the basis of out of date medical information and deceased individuals.
Test and Trace - Data passed to councils via NHS Test and Trace (formerly CTAS)

- **Poor data quality**: Councils report that the data has quality issues (e.g. incomplete addresses, missing or incorrect phone numbers and emails), though some of this is to be expected as the contact data is for those people that the central trace system could not find.

- **Lack of clarity on information governance and permissions**: Councils lack clarity on how they can use Test and Trace data. For example, guidance is unclear on whether councils are allowed to link it with existing council datasets - such as individuals known to adult social care - to provide targeted support for those asked to quarantine.

- **Not possible to automate data extraction**: Accessing this data still requires a time-consuming manual process that increases the chances of data errors during transformation. Delays to accessing this data inhibit the successful slow down of virus transmission as the time to contact is increased.

- **Restrictive access control and security**: The security requirements for accessing this dataset make it difficult for council officers to use the data while working from home (as many have to do to follow government policy). This includes the requirement to have a whitelisted IP address to access the data. Additionally, only a limited number of council officers can download the data, creating a bottleneck in the process.

6. Is the public able to comprehend the data published during the pandemic. Is there sufficient understanding among journalists and parliamentarians to enable them to present and interpret data accurately, and ask informed questions of Government? What could be done to improve understanding and who could take responsibility for this?

Local authorities have reported that journalists and members of the public have queried the local infections data they have published, since it does not always align with other government data. This has caused confusion and additional work for council data analyst teams who have to try to reconcile the differences between restricted datasets and publicly published data. Ensuring that adequate data descriptions and meta data are published with datasets can mitigate this issue. Keeping methodologies consistent where possible over time also enables all parties to better understand published data.

Public data (for instance, daily Covid-19 cases) changed formats several times and was supplied in a way that was almost impossible to develop efficient, automated data pipelines for. It has required a substantial effort from several highly skilled data scientists to get from what was supplied nationally to a clean, consistent picture of cases on the London DataStore for Londoners
to use https://data.london.gov.uk/dataset/coronavirus--covid-19--cases. This time could have been better spent on deeper local analysis. Countless other organisations were repeating this exercise, having to piece the picture together from different data sources and make their own judgements about what was sensible to do with it. We have welcomed the recent improvements to the API offered at coronavirus.data.gov.uk. It is well documented and has been provided alongside a Python / R package, making the data much easier to use.

However, we would like to note that there continues to be a lack of transparency regarding testing and contact tracing data in England, including:

1. No regional or sub regional testing capacity and activity and performance data is published.
2. No regional or sub regional data on Care Home testing is published.
3. No regional and sub regional data on contact tracing performance is published.

In order for journalists and parliamentarians and the public to have sufficient understanding of testing and contact tracing capacity, activity and performance it is necessary for NHS Test and Trace to publish detailed regional and sub regional level testing and contact tracing data including:

1. Allocation of testing capacity to London compared to other regions
2. Testing rates by local authority
3. Capacity of London regional test sites (RTSs), Local Test Sites (LTSs), mobile testing units (MTUs) and home test kits per week
4. Test to result turnaround times
5. Weekly regional/sub-regional analysis to determine testing rates by gender, age, ethnicity, socio-economic status
6. Breakdown of testing activity and positivity rates by Local Test Sites to understand footfall and determine whether the right communities are being reached
7. Home testing performance metrics
8. Regional and sub regional data on contact tracing performance

7. Does the Government have a good enough understanding of data security, and do the public have confidence in the Government’s data handling?

Some London boroughs have reported that the security in place for access to restricted government datasets including the Shielding Persons List, Public Health England Line List data and Test and Trace has been prohibitive to effectively accessing the data.
Restrictions include:

- IP / Device whitelists (restricting access to data to specific devices or office connections)
- Limited number of authenticated borough users
- Borough officers unable to access data while working from home

Overly onerous security limits the number of borough officers (and the devices they can use) who can extract data, causing bottlenecks in the data flow to analysts and decision makers.

8. How will the change in responsibility for Government data impact future decision making?

We have no view on changing departmental responsibilities, only their efficacy in relation to local government and our ability to serve our citizens. In our longer experience, data-sharing arrangements and practice between Whitehall, combined authorities/GLA and local authorities vary significantly and are inadequate.

Poor data sharing can have consequences that range from inconvenience for citizens and poor use of public funds to failure to prevent serious harm to vulnerable individuals. There are numerous cases where data sharing is difficult or the experience across government departments is variable. Where this occurs, it often prevents local authorities from providing the best service to the public and is a source of frustration and inefficiency.

We are particularly concerned about (a) lack of access to sufficiently detailed data around in-work and out-of-work poverty (b) inconsistent access to the data we need for the discharge of our statutory functions / devolved powers (for example, skills).

Finally, work at a GLA and local government level to join up data, for example via the London Datastore, takes place independently of central government funding or resourcing. It is not clear from the National Data Strategy whether the Government has considered the full scale and granularity of the ‘delivery’ data held by local authorities and their technology estate supporting hundreds of frontline services, commissioned by local authorities but operated largely by third parties.

Our primary recommendation is for greater interoperability, which we are working towards in London, but would benefit greatly from a national steer and resourcing. We also request that

- Local government receive additional powers for data sharing, current enabling legislation (e.g. the provisions of the Digital Economy Act 2015) seem to have been drawn up without local government in mind.
- Regional expertise developed over the last decade with successful pan-authority data platforms (e.g. London, Manchester, Leeds) is represented in future policy design.
Sufficient resource is identified by the National Data Strategy to enhance local government data expertise sufficiently.