


TRANSLATIONAL ARTICLE

Local Data Spaces: Leveraging trusted research environments for secure location-based policy research in the age of coronavirus disease-2019

Jacob L. Macdonald¹ , Mark A. Green², Maurizio Gibin³, Simon Leech⁴, Alex Singleton² and Paul Longely³

¹Department of Urban Studies and Planning, University of Sheffield, Sheffield, United Kingdom

²Department of Geography and Planning, University of Liverpool, Liverpool, United Kingdom

³Department of Geography, University College London, London, United Kingdom

⁴Leeds Institute for Data Analytics, University of Leeds, Leeds, United Kingdom

Corresponding author: Mark A. Green; Email: Mark.Green@liverpool.ac.uk

Received: 04 November 2022; **Revised:** 24 March 2023; **Accepted:** 03 May 2023

Keywords: COVID-19; data trusts; secure data; trusted research environment

Abbreviations: ADR UK, Administrative Data Research UK; CDRC, Consumer Data Research Centre; DSH, Data Safe Havens; JBC, Joint Biosecurity Centre; LDS, Local Data Spaces; ONS, Office for National Statistics; TRE, Trusted Research Environment

Abstract

This work explores the use of Trusted Research Environments for the secure analysis of sensitive, record-level data on local coronavirus disease-2019 (COVID-19) inequalities and economic vulnerabilities. The Local Data Spaces (LDS) project was a targeted rapid response and cross-disciplinary collaborative initiative using the Office for National Statistics' Secure Research Service for localized comparison and analysis of health and economic outcomes over the course of the COVID-19 pandemic. Embedded researchers worked on co-producing a range of locally focused insights and reports built on secure secondary data and made appropriately open and available to the public and all local stakeholders for wider use. With secure infrastructure and overall data governance practices in place, accredited researchers were able to access a wealth of detailed data and resources to facilitate more targeted local policy analysis. Working with data within such infrastructure as part of a larger research project involved advanced planning and coordination to be efficient. As new and novel granular data resources become securely available (e.g., record-level administrative digital health records or consumer data), a range of local policy insights can be gained across issues of public health or local economic vitality. Many of these new forms of data however often come with a large degree of sensitivity around issues of personal identifiability and how the data is used for public-facing research and require secure and responsible use. Learning to work appropriately with secure data and research environments can open up many avenues for collaboration and analysis.

Policy Significance Statement

This work presents the Local Data Spaces program—a collaborative pilot project leveraging Trusted Research Environments and national data from the Office for National Statistics. The program was a blueprint towards using a mix of secure and open data for localized policy analysis around coronavirus disease-2019 inequalities and economic vulnerabilities. The secure infrastructure and record-level data enabled a detailed profiling and comparison of localities across a series of Local Authority reports ultimately released as open resources. As new and novel forms of data—collected both actively (e.g., surveys) or passively (e.g., mobility measures), become

increasingly available for secondary research purposes, security and data disclosure risks must be mitigated at all stages. Accredited researcher training and overall infrastructure and data governance ensure the security of conducted research—and resulting outputs. Learning to work with these secure infrastructures for rapid response policy analysis goes towards making use of our wealth of national data resources.

1. Introduction

The widening of access to novel data sources from the public and commercial sectors has fundamentally impacted the research landscape for practitioners and policy analysts. The increasing availability of data, formats, and software, along with those skills needed to work with them, require new infrastructure and data systems to accommodate their access and ease of use. As the availability of such data has permeated into academic and government departments, researchers are exploiting them to produce novel insights for supporting evidence-based policy discussions. Electronic health records, economic surveys, or granular passive monitoring of mobility or mobile phone data (among many others) can have significant potential for socially conscious, public-facing research (Ricciato et al., 2020). The value of these new forms of data in the UK has been evident throughout the coronavirus disease-2019 (COVID-19) pandemic, as they were increasingly relied upon for local and national policy discussions where traditional datasets could not inform decisions or were slow to be made available.

The Local Data Spaces (LDS) program was a rapid response and research-intensive project in the context of the initial UK 2020 COVID-19 outbreak and national lockdowns. This was a collaborative initiative from the Joint Biosecurity Centre (JBC), the Office for National Statistics (ONS), ESRC's Administrative Data Research (ADR) UK, and academic researchers from the Economic and Social Research Council (ESRC) funded Consumer Data Research Centre (CDRC), piloted from November 2020 to April 2021. The program created a framework for using open and secure data resources for locally focused research, leveraging access to secure data via the ONS trusted research environment (TRE)—the Secure Research Service (SRS). Anonymized record-level national surveys and administrative registries were made available for research to be done at the small area scale to support the varied responses to the health and economic impacts of COVID-19 across the country. Using the SRS platform and infrastructure as a *Local Data Space* enabled LDS accredited researchers with secure access to analyze a range of core national-level surveys and registries related to COVID-19 (e.g., COVID-19 Infection Survey, Test and Trace Data, Excess Mortalities) and economic and labor pressures (e.g., Labour Force Survey, Annual Population Survey, Business Registry Database, Business Impacts of COVID-19 Survey)—at granular (pseudo-anonymized) record level under strict access and monitored use. This research note takes the perspective of the embedded LDS academic researchers (the authors on this paper) as they navigated the data access, preparation, and cleaning of data, analysis and presentation of the outputs.

The LDS project took a geographically localized approach to generate a series of openly available Local Authority District (LAD) reports using the securely held COVID-19 and economic datasets in the SRS and openly available mobility and high-resolution indicators from the ONS, CDRC, and Google Mobility Data. Reports were co-designed with a series of Local Authority teams in response to ongoing and real-world challenges in managing the COVID-19 pandemic. Initial engagements with 25 Local Authority health and economic response teams helped to co-produce timely research questions to match their evidence and data needs. Insights were instrumental for LDS researchers to tailor their work for the rapid co-production of data analyses to ensure outputs remained focused and valuable. Feedback was continually sought from stakeholders on all outputs, helping to refine analyses based on the needs of Local Authorities. This iterative co-production process was time-intensive, but invaluable for generating policy-relevant research. We also assumed that the research questions being asked by our stakeholders were similar to those likely being asked by other Local Authorities not involved in LDS. We therefore designed reports and analyses so that they could easily be replicated for any Local Authority (typically using coded

for-loops and automatically updateable reports) so that other places beyond the original consultees could benefit from the data insights produced during our co-production process.

This collaborative workflow made use of the secure data within the SRS environment to generate and export a series of 10 openly available and individualized reports for 323 LADs across England. Researchers worked with SRS analysts to facilitate validation and disclosure checks undertaken for all outputs. Each report follows a domain exploring local patterns in health inequalities and economic vulnerabilities - further compared against national and regional trends. These reports were openly shared with LADs and local stakeholders across England through the CDRC data repository (<https://data.cdrc.ac.uk/datasets/local-data-spaces>), enabling them to acquire data insights from the secure data they would otherwise be unable to access. The focus on repositioning a TRE environment for rapid responses to local COVID-19 health, mortality, and related issues, combined with measures of economic and labor force pressures, enabled researchers to convert securely held data otherwise not used into actionable local research assets.

When secure infrastructure is able to provide access to a variety of data across different domains, local policy research or small-area comparisons can more easily be explored across different angles and contexts (e.g., demographics, locations, typologies) in a centralized working environment with overarching data governance. The issues of data security are particularly important to consider when exploring novel forms of data or small-area research, where personal or business disclosure risks may be higher (Affleck et al., 2022; Kavianpour et al., 2022). TRE infrastructures generally, and a collaborative workflow as demonstrated through the LDS project, play an important role in managing these data security risks while generating non-disclosive and openly available data products and insights available for, and driven by, local stakeholders.

2. Trusted Research Environments and Secure Data Access

In practice, a TRE (data trust; data safe haven) is a secure data store that hosts any number of potentially sensitive data for research and analysis use. Access is restricted, monitored, and often based on clear data use agreements and pre-approval of a research project. The guiding idea is to make record-level (pseudo-anonymized) data available for research use and generating broader insights for the public good (Hardinges et al., 2019). Infrastructures can range from physically secure and monitored lab spaces (e.g., traditional in-house secure lab spaces or ESRC's "SafePod" network¹) to virtual desktop environments where the data is made available and worked on remotely. The most crucial component is that no information, data, or outputs are able to leave the environment without a series of strict checks and vetting for data disclosure issues. Since source data must often be accessed, analyzed, and prepared by the researchers in a controlled environment, projects working with record level and potentially disclosive data cannot implement *open data and research* principles to the same degree as with aggregated or purely open data (Arribas-Bel et al., 2021).

When researchers or policy makers require access to secure datasets, the use of a TRE can centralize the application and analysis process while ensuring proper data stewardship among all parties. Requesting access to a number of datasets individually in a responsible and secure manner can often take up significant time and resources with the need to arrange bespoke data licensing and sharing arrangements. This can be particularly challenging for individual researchers or small project teams working with short time frames and especially if responding to urgent policy issues (Vindrola-Padros, 2019). TREs mitigate these administrative barriers by centralizing data sharing through secure platforms, researcher training and accreditation, project approvals, and strict disclosure checks on any outputs. Accessing these data securely and working with them alongside other potentially open data provides researchers with increased potential for deeper insights from multiple, granular resources, and complementing analyses.

¹ <https://safepodnetwork.ac.uk/>.

These strengths were valued by all Local Authorities we engaged with, who appreciated their centralized infrastructure at reducing the resources they would otherwise require to deliver themselves.

The TRE landscape is complex with varying consensus over terminology, scope, legal frameworks, and guiding principles. In the broadest sense, the Open Data Institute (ODI) employs the definition that: *a data trust provides independent, fiduciary stewardship of data* (Hardinges, 2020). This is meant to reflect the different purposes that infrastructure may have depending on the different contexts and needs for data sharing across stakeholders (e.g., public, private, academic). Many different structures of data trusts can be developed under varying legal or non-legal arrangements. Specific agreements can be arranged between relevant parties as long as the broad legal issues of data protection and disclosure are addressed in guiding terms of references, contracts or organizational policies (Delacroix and Lawrence, 2019; Stalla-Bourdillon et al., 2021; UK AI Council, 2021). This leaves a range of potential public-private data-sharing frameworks where different data resources can be worked on together in a secure environment for socially-conscious research benefits.

The LDS project was conducted in the UK context, making use of national data infrastructure and resources, primarily through the ONS. Secure and responsible access to data and the use of data trusts is a national priority for research, development, and skills advancement (Office for Artificial Intelligence, 2021). The framework of using secure environments for research that is based upon sensitive data is, however, increasingly applicable in many contexts. The use of some form of TRE is recognized as good practice and adopted by many different institutions, internally within public or government agencies, in private sector workplaces, or through research or data services. Public agencies at the national or local level often make use of these infrastructures to manage their secure data used for in-house analysis. Outside of this, TRE-type infrastructures are frequently used in more commercial or private ventures with data stores holding specific data collected or managed by an organization, potentially available for external research use (e.g., private mobile network operator data platforms) (Delacroix and Lawrence, 2019; Hubbard et al., 2020; Kavianpour et al., 2022).

There are many TRE environments dedicated to providing secure data for research purposes—whether for academic, public policy, or socially-facing development work. For example, in the UK there are a range of dedicated secure environments with different surveys, registries, or secondary data resources available, such as the ESRC's UK Data Service or across various different official government departments or ministries, related to health data or otherwise. The Wales-based Secured Anonymised Information Linkage (SAIL) Databank is a model example of a TRE hosting secure and linked health and population data and providing invaluable research opportunities. More topic-specific data research services can also be hosted through research centers or academic institutions hosting multiple data licensing agreements for research purposes (e.g., ESRC Consumer Data Research Centre). The LDS program fits in with this broad context to exploit the secure-open data framework for local policy analysis within one of these TREs—the ONS SRS.

The ONS is just one of many different types of providers of secure TREs for research use. It is in a particularly strong position since they are the national statistics agency for the UK and can offer varying access to the official UK national censuses, surveys, or other collected and administrative data related to the population, economy, and society. The ONS itself is the production arm of its broader parent agency, the UK Statistics Authority, which has a reporting function to parliament on topics related to national accounting, data protection, and use for public policy. Furthermore, the UK TRE landscape is not unique. There are a range of similar strategies being developed in countries around the world to promote socially responsible research using secure data in a trusted environment (Paprica et al., 2020; Zhang, 2021). The LDS program should be easily deployable across different countries and contexts, offering value at brokering data access in TREs and impact-led research.

Safeguards and limitations for TRE researchers

Particular safeguards and limits are enforced when working within the parameters of secure data. These can impose time or resource constraints on researchers if not adequately prepared for them, and it is

important to plan in advance to minimize their disruption—especially if accessing the data within short research project timespans. Working within a TRE, such as the SRS, often requires working in isolated (physical or virtual) lab spaces without connection to external resources (e.g., Internet), limited capabilities for importing or exporting pre-scripted code for cleaning or analysis, and may require additional time if specific coding libraries or packages are not directly available within the TRE. These safeguards on the movement of objects into and out of the TRE can create additional steps that a researcher may need to take in order to complete a project. The core considerations for data analysis within a TRE include:

- Access to TRE will typically require some form of researcher accreditation to validate skills in data security and management. These may also vary between different TREs.
- Technical analysis skills are required to work with data in an isolated secure environment. Researchers often cannot access external resources, internet, or tools while working with the secure data, and thus should be adequately proficient in reading, querying, managing, and running any statistical models and data work offline.
- The movement of any data into and out of these environments typically requires vetting by data analysts trained in statistical disclosure control. While TREs are able to facilitate a combination of working with secure data and other sources such as locally held or open data resources, these must still typically undergo import checks. Research outputs, lookup tables, data insights, model results, and any and all other items which are to be exported from any TRE must undergo a strict export check for disclosure issues that could compromise data protection.
- Linkage of multiple different secure data sources across different TREs - as opposed to working with multiple datasets within one TRE, is still a challenge. Openly available data (or bespoke local data with proper ownership) can often be imported for analysis within the TRE. Data securely held in a separate and isolated TRE cannot always be shared horizontally across infrastructures; depending upon the constraints embedded into the data governance.

3. The Local Data Spaces Project—A TRE Example Workflow

3.1. Local Data Spaces

The use of secure data infrastructure for local policy impact is highlighted here in a case study of the LDS project which ran for six months during the COVID-19 pandemic.² The aim was to leverage secure health and economic data through the ONS SRS, combined with data from the CDRC, to generate research outputs and insights which emphasized *local* dynamics and challenges in responding to COVID-19. Taking a focus on the Local Authority administrative level (many local government decisions are made by Local Authorities), this work used a variety of datasets in tracking, comparing, and presenting variations in local COVID-19 outcomes and relevant economic indicators. Data were accessed through the secure SRS infrastructure by accredited researchers—the academic partners from the CDRC, and resulting outputs, reports, and visuals underwent strict export disclosure checks and vetting.³

A series of local engagement exercises were conducted with relevant Local Authority health and economic response teams at the outset of the project (e.g., surveys, workshops, seminars, one-on-one meetings). These meetings helped to co-produce timely research questions and challenges facing local areas - many of which required more time and resources than local teams were able to dedicate while managing the day-to-day operations in response to a crisis. This included a range of topics that could only be explored with record-level data to provide more detailed breakdowns in patterns or distributions over space, across occupational levels, or demographic groups. For example, small-area (sub-LAD) economic vulnerabilities were difficult to examine using aggregate, openly available, indicators. Provisioning secure access to

² Project details and openly available LAD reports for download from the project at: <https://data.cdrc.ac.uk/dataset/local-data-spaces>.

³ <https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme>.

granular economic registries or labor market surveys can help local stakeholders leverage the wealth of national, regional, and local secure data resources to support their policy analysis and aims. Repeated engagements with stakeholders allowed LDS researchers to develop a research plan and framework using the available data in the secure environment to generate actionable insights for supporting local decisions.

In the context of this project, the LDS took the form of an accredited research project approved through the SRS platform. The overall research objectives for this work were broad, comparing *local* trends in COVID-19 inequalities and economic vulnerabilities across the country. This was intentional to allow the co-production process to work effectively—allowing the natural design of research questions that emerged from discussions with Local Authorities. The areas of COVID-19 inequalities and economic vulnerabilities were identified by Local Authority stakeholders across our engagements as priority concerns. The LDS researchers worked within the secure environment with the wealth of nationally collected resources from the ONS and other organizations, looking to extract comparable, robust, and tractable data analysis for all local areas across England. Through doing the research on behalf of Local Authorities, the LDS team facilitated a process of indirect access to these secure data held in the SRS that Local Authorities were not otherwise accessing or able to.

Our comparison systematically looked at how local (LAD) patterns in COVID-19 and economic vulnerabilities compared with regional and national averages. This ended with a series of automated, bespoke, and comparable reports for all LADs in England built from national surveys and registries. The LDS team identified the relevant datasets from the available catalog, combined with openly available data relevant to the analysis, and applied for project space on the SRS system where the respective datasets and tools for analysis were made available through secure and monitored remote access in dedicated, accredited lab spaces.

Within the SRS infrastructure, accredited researchers had access to a number of secure datasets (Table 1). Particular interest was related to the COVID-19-specific datasets that LADs did not have access to, especially the COVID-19 Infection Survey. Insights from these data were combined with a series of ONS flagship data products, national accounts, surveys, and registries to explore related economic pressures, occupational and industry distributions, and local challenges. Record-level data on business, employment, and the labor force were obtained from the Business Structure Dataset, Business Registry, and Employment Survey, and Labour Force Survey.

Our project adopted a framework of incorporating relevant auxiliary open data resources to complement and contextualize the analysis of secure datasets where possible. We included a range of openly available data on local characteristics, demographics, and regionally observed patterns (Table 2).

Table 1. *SRS secure data resources*

	Host	Source
<i>Public Health and COVID-19</i>		
COVID-19 Infection Survey (CIS)	ONS	ONS (2023d)
Test and Trace (TT)	DHSC; NHS	DHSC (2022)
Deaths Registrations Dataset	ONS	ONS (2022b)
<i>Economic and Labour Force</i>		
Business Structure Dataset (BSD)	ONS	ONS (2023c)
Business Registry and Employment Survey (BRES)	ONS	ONS Social Survey Division (2023b)
Labour Force Survey (LFS)	ONS	ONS Social Survey Division (2023b)
Annual Population Survey (APS)	ONS	ONS Social Survey Division (2023a)
Annual Survey of Hours and Employment (ASHE)	ONS	ONS (2022a)
Business Insights and Conditions Survey (BICS) ^a	ONS	ONS Social Survey Division (2023a)

^aThis BICS product supersedes the previous *Business Impacts of COVID-19 Survey*.

Table 2. Open data resources

	Host	Source
Access to Healthy Assets & Hazards (AHAH)	CDRC	Green et al. (2018)
Internet User Classification (IUC)	CDRC	Singleton et al. (2020); Alexiou et al. (2018)
Retail Centers	CDRC	Macdonald et al. (2022)
Business Census	CDRC	CDRC (2022)
Index of Multiple Deprivation (IMD)	MHCLG	MHCLG (2019)
COVID-19 Community Mobility Reports	Google	Google LLC. (2022)
Housing Transaction and Prices	Land Registry	HM Land Registry (2021)
Housing Affordability in England	ONS	ONS (2023e)
Mid-Year Population Estimates	ONS	ONS (2022c)
ONS Data Science Campus Faster Indicators	ONS	ONS (2019)

Note. Online Job Adverts; Retail Sales Index; Index of Service; Company Incorporations and Dissolutions; Card Spend; VAT Return Indices.

When working with a number of different secondary data sources from different providers, it is important to understand how (and if) they overlap in terms of study context, geography, and timescale, and to understand the limitations of analyzing any two data sources jointly.

In order to put together a comprehensive local profile of each area, we drew on a variety of open data products with detailed geographic components. These were used to complement and contextualize the secure data which provided the relevant COVID-19, related outcomes, and economic indicators. We used the Mid-Year Population Estimates to get information on local demographics (ONS, 2022c), the Access to Healthy Assets & Hazards (AHAH) provided underlying small area context to health challenges and amenities (Green et al., 2018), and the Index of Multiple Deprivation (IMD) is able to give insights into local relative measures of deprivation (MHCLG, 2019). Openly available geographic data on local business density and retail activity was provided through the CDRC Business Census (CDRC, 2022) and Retail Centre delineations (Macdonald et al., 2022). The geographic elements of these sectors are important to consider as lockdowns particularly impacted retail and hospitality which are often concentrated in city centers and high streets and were notably hard hit during the pandemic.

We further included additional data to profile local areas based on changes in housing prices pre- and post-pandemic, and aggregate measures of local mobility generated and provided openly through Google's Mobility Reports (Google LLC., 2022). These data sources help in providing added local detail and context for each of the areas across England. While not all openly available measures were able to be provided at the most detailed resolution desired, there are still insights to be gained in matching them to data from the SRS to provide additional background on their insights.

Figure 1 provides an overview of the LDS project workflow. When combining a series of openly available and secure data for research work in the SRS environment, it is important to consider this workflow at the beginning of the project to plan and foresee potential challenges or delays. Openly available data are collected and provided to the SRS analysts who run an initial vetting to confirm the files which are then imported to the research project space within the secure environment. Here, accredited LDS analysts have access to their ingested openly available data, along with the secure data needed for the project. The secure environment provides the software and infrastructure needed for the analysts to conduct their statistics and analytical research using the series of record-level data available. Following the analysis, non-disclosive outputs can be requested for export from the SRS. While no data or identifying information can leave the secure environment, aggregate, non-disclosive research or model outputs and information can be checked by SRS analysts who release the request after confirming that no disclosive information is leaving the environment.

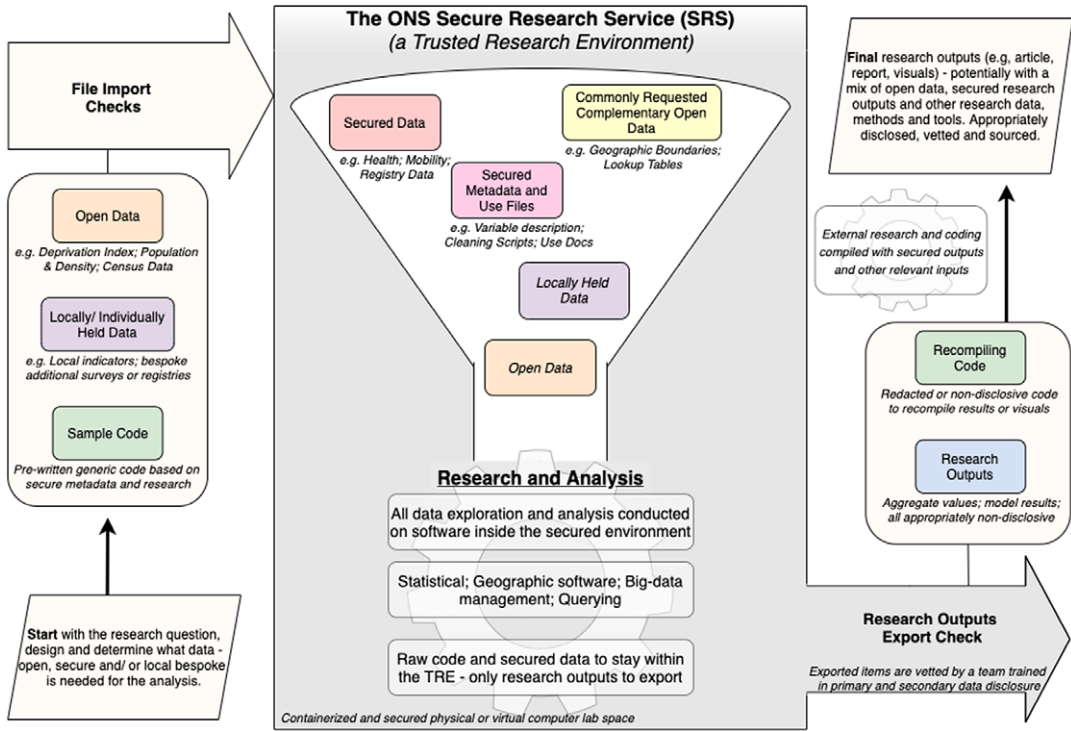


Figure 1. The LDS project workflow—an example of a secure data research project.

3.2. Local profiling of all LAD areas in England

Over the course of the project, the LDS researchers developed a series of local profiles built on the range of datasets and information available and ingested into the secure environment. LAD profiles were generated algorithmically to report on a series of consistent trends and patterns observed in COVID-19 and related economic outcomes—comparing local dynamics to regional and national trends where possible. This resulted in a set of comparable LAD profiles built along a variety of domains for 323 areas across England.

One of this project’s broad objectives was to emphasize and highlight the *local* aspects of the national data products and facilitate analyses across areas. This ultimately resulted in a valuable set of resources available for researchers and local stakeholders allowing them to examine and compare their area to others in terms of COVID-19 infections and impacts, sector-based economic vulnerability, or local mobility. A package of reports were prepared for each LAD area, broken into a series on public health and COVID-19 measures and another series on economic or sector-based labor market and business vulnerability. Table 3

Table 3. LDS openly available LAD data reports

Public health (COVID-19) and inequalities	Economic/sector-based vulnerabilities
Occupational inequalities in COVID-19	Industry and Sector Workforce Densities
Ethnic inequalities in COVID-19	Economic Vulnerabilities
Demographic inequalities in COVID-19	Retail and Accommodation Sectors
Geospatial inequalities in COVID-19	Population and Housing
Excess mortality	Human Mobility Patterns

Note. Reports are available for each LAD in England via the CDRC Geodata Packs platform: <https://data.cdrc.ac.uk/geodata-packs>.

highlights the 10 reports, five across each series. With the use of algorithmically coded *for-loops*, we efficiently adapted these reports for a variety of spatial extents.

Each report contains the relevant (non-disclosive) breakdown of indicators and measures to track and benchmark each LAD. This could be in terms of providing infection rates or mobility pattern changes over time, or static distribution of small area occupation and industry sector densities of those most impacted by ongoing lockdowns. LAD indicators are compared with respective regional and national level equivalents, where feasible, to better understand how the local area is fairing in the broader context. In the end, these 10 reports for each LAD area provide a detailed picture of local COVID-19 outcomes and related economic pressures over the course of the pandemic.

This work leveraged the record-level aspect of the data in the secure research environment to overcome common problems in using aggregated statistics. A significant amount of detail can be lost in the naive aggregation of data into higher spatial units (e.g., regions). This can mask important sub-region patterns and potentially introduce spatial biases into the outcomes. When working with high-resolution data, a significantly more detailed analysis can be undertaken and further, the data can be explored in tandem with other data representative of the same location or demographic. Even more powerful is when datasets can be linked between themselves for additional layers of richness and robustness. Through the local-focused work inside the secure environment, small area and granular concepts of spatial densities and distributions or demographic stratifications can be explored—highlighting additional dimensions to consider in local policy planning and response.

The LDS project and the generation of a package of openly available reports for each LAD area is at its core an exercise in working through a hybrid open and secure research project. We use the TRE to generate an open data research product for Local Authority and related stakeholders to use - applying the principles of open data to generate non-disclosive aggregate research outputs openly available for sharing amongst stakeholders (Arribas-Bel et al., 2021). TREs thus allow sensitive data to be analyzed in a secure environment and incorporated into the broader research workflow.

3.2. Project Impact and outcomes

A key impact of the project has been co-producing research with 25 Local Authorities, resulting in the production of 10 individualized reports for 323 LAD areas across England—all built on secure data, yet providing openly available content.⁴ All reports were bundled into an LDS packet which includes for all LADs each of their 10 reports in a zipped file. In the initial seven months since the reports were launched in May of 2021 (until December 1, 2021) there were a total of 813 packets downloaded for their online repository covering most all LAD areas. Figure 2 highlights the key stakeholders of interest downloading respective LDS packets, as obtained via the CDRC self-reported access statistics. Primarily, we see a bulk of academic and local government interest.

The LDS initiative was designed to be flexible and responsive, allowing us to reposition the needs of the project based on research questions and timely local policy issues identified during the co-production process. In some instances, the creation of short reports was used as “conversation starters” to help this process and often led to bespoke analyses for Local Authorities. For example, work conducted towards building a report on occupational inequalities highlighted that furloughed populations (i.e., temporarily not working) were more likely to have tested positive for COVID-19 in Norfolk. Norfolk County Council were interested in these local insights and requested further evidence on who was more likely to have been furloughed since they did not have any local data. Through these channels, we were able to produce bespoke analyses to support their needs and generate tailored local data insights to support policy discussions. The secure aspect of TREs meant that Local Authorities could

⁴ While, as of 2021, there were 309 Local Authority Districts in England, different datasets from varying time periods used a mix of current and previous (pre-merger or split) areas. We produce results for the lowest common denominator in cases where secure data cannot be translated to other geographic codes.

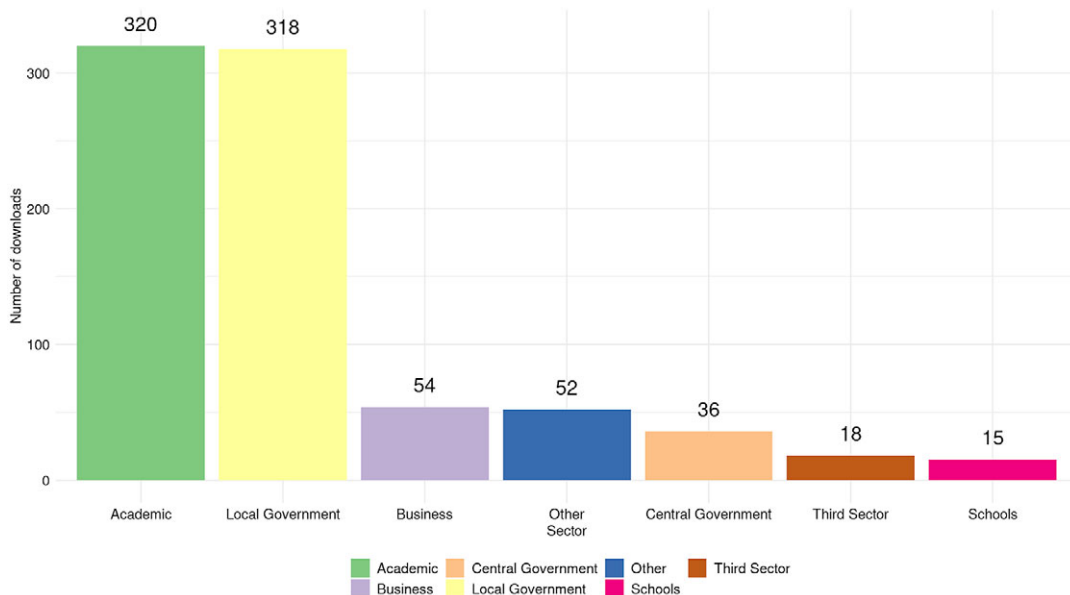


Figure 2. Downloads of LDS packets (i.e., all 10 LAD reports collated in a zipped file): May–Dec., 2021.

not “see” what data was held within, making it difficult to understand the opportunities available. These conversation starter short reports helped make the data feel “real,” allowing Local Authorities to actively engage and refine research questions further.

This collaborative, flexible, and iterative process to evidence generation helped us to support key policy issues. One example was the piloting of lateral flow testing in Liverpool (often titled the “mass testing” pilot). Embedded LDS researchers were able to provide additional analytical capacity and data insights to Liverpool Local Authority during the pilot. Test and Trace data were analyzed to investigate inequalities in uptake, identifying that communities less confident in using internet technologies had low uptake and leading the Local Authority to avoid social media routes to advertise testing (Green et al., 2021). Geospatial analyses were supplied to optimize the coverage of test sites, helping address their lack of skills in this area (Green, 2021). An additional benefit was the ability to investigate mortality trends before, during, and after the pilot given that the council did not have access to the same timely data as in the SRS, showcasing the benefits of making data resources available to Local Authorities faster than normal.

LDS researchers working within the TRE were also able to respond to national requests for evidence on timely policy issues and provide additional analytical capacity. The UK Government’s Scientific Advisory Group for Emergencies (SAGE) were concerned about COVID-19 disproportionately impacting younger women during the second wave and approached the team as a potential avenue to help gather evidence. From the secure data in the TRE, the researchers were able to supply data insights on gender inequalities in COVID-19, demonstrating that occupational differences were not evident by gender (EMG, 2021; ONS, 2021). Through having established data governance and access setup, the LDS team were able to respond quickly to urgent requests and supply data insights faster than many groups. We approached discussions with the UK Government similar to working with Local Authorities, helping to embed co-production while working “at pace.”

Finally, the success of LDS was demonstrated by winning the ONS (2021) Project Award for Research Excellence. The award recognizes innovative research that has delivered public good or informed policy decisions, and LDS was praised for its collaborative approach to working with Local Authorities to co-produce timely evidence at pace for responding to the pandemic on various fronts.

4. Lessons Learnt

A formal project evaluation report of the LDS was undertaken by the JBC and ADR UK (Henggeler et al., 2021). Key strengths of the project included:

- (i) The scheme was popular with Local Authorities who appreciated additional analytical support at a time when their resources were stretched in responding to COVID-19;
- (ii) The SRS provides centralized technical infrastructure for hosting high-resolution data and analytical resources, meaning that users did not have to individually invest in providing resources themselves and therefore save costs;
- (iii) Local Authorities did not always have the skills, resources, or time to apply and make the most of the SRS, meaning that academic partnerships to help them indirectly access data;
- (iv) Co-produced reports were relevant and help Local Authorities gather evidence differently, complementing some of the more on-the-ground methods of data collection used locally. Many of the datasets LDS made available to Local Authorities were not previously being used by them.

This model of engagement ultimately relied on the academic partnership and having the LDS-accredited researchers with dedicated time to provide data analysis and generate reports. It was recognized that the short timeframe for the partnerships and limited dedicated research resources (six months) was a barrier to uptake from Local Authorities when dealing with longer-term issues or larger projects. Some Local Authorities who were approached did not have sufficient time and resources to take part. These were likely to be the Local Authorities who could have benefitted the most from the additional analytical capacity. The time commitment towards becoming an accredited researcher and finding or setting up a nearby secure lab environment was also a drawback for the non-academic partners. If no secure lab environment was nearby, then the alternative was a relatively intensive accreditation process to set up a dedicated lab space—often not possible for Local Authority workplaces. As such, no Local Authorities were able to apply for direct access themselves to the TRE in the 6-month pilot.

A significant amount of pre-emptive work was needed to be done by the academic LDS researchers before accessing the data. When working on these limited timeframes, it is key to develop a clear outline of a research proposal and specifically, how this can be answered using the secure data. The general research community interested in using these secure data and infrastructure would benefit from knowing specific details and formats of the data they require (e.g., metadata, variable lists, synthetic data). This is particularly important when using multiple data resources in the same TRE or importing external open data to complement the analysis. Sometimes the opaque barriers of TRE security hindered how external users could see their potential, ultimately putting off some Local Authorities.

Having a well-structured research proposal with flexibility built in not only provides a clear outline for the researchers to follow, but supports the TRE staff who need to vet and approve project proposals, data requests, and outputs. It is necessary to understand how certain surveys or registries can be used effectively. Particularly when it comes to the geographic component, many surveys have limited statistical power in small sample sizes. Managing these expectations prior to the project beforehand, and coming up with solutions to what is feasible, was key for having a flexible project.

While there was an appetite for this project, the outputs and partnership between local stakeholders and more nationally focused datasets, there is limited feasibility of scaling up the SRS and academic support to all Local Authorities directly without significant dedicated resources specifically targeting the local aspect of these national data resources (both in academic time, but also to support increased processing of TRE requests/outputs). At the same time, there were significant economies of scale that were realized in the methods applied (e.g., generating reports for all LADs in England based on discussions with a smaller set of organizations). The systematic creation of reports for all LADs using a pre-packaged template built in consultation with local stakeholders is an effective way at leveraging the secure data and resources to their fullest extent and benefit.

5. Conclusion

Our work serves as a blueprint for incorporating the use of sensitive, granular location-based data from a secure research environment into broader co-produced research projects with external stakeholders. Leveraging a secure trusted research environment is an accessible way in which researchers and local stakeholders are able to access granular and secure data which may otherwise be unavailable or only provided at aggregate levels above what local policy may find useful. Further, such an environment provides a secure platform from which local stakeholders, whether economic or health policy-makers, academics, or others, can complement their own data with the use of external granular data resources.

High-resolution, record-level, and granular data from surveys and administrative registries are the foundations upon which a wide variety of local and regional policy decisions are made. No clearer has this been seen recently than over the course of the COVID-19 pandemic within the UK with the constant need for timely, robust, and safe data generated from sensitive sources of information. Key aspects of everyday life across all domains were impacted—not least among them including local patterns and dynamics in health and infections, economic vulnerability, city center or high street mobility, and social sector impacts, among others. Local policy and research must necessarily consider these angles and more to develop and design resilient functioning areas—and having access, to and availability of reliable data at the appropriate geographic scale is key for responding, moving forward, and leveling up these localities.

Acknowledgments. The authors are grateful for the support provided throughout the LDS project. Overall LDS project governance and organization was provided by the Joint Biosecurity Centre (Department of Health and Social Care). The Office for National Statistics (ONS), and specifically the Secure Research Service (SRS) Support Team, were instrumental in supporting and managing secure access to the SRS and export of outputs. Administrative Data Research (ADR) UK provided mentoring of the team, as well as general feedback on research ideas and directions. We thank all the Local Authorities who engaged with LDS. Finally, the authors would like to thank the attendees of the Data for Policy 2021 Conference session for discussion and feedback.

Funding statement. This research was supported by grants from the ESRC Consumer Data Research Centre, ES/L011840/1; ES/L011891/1. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interest. The authors declare no competing interests.

Data availability statement. Where openly available data were used, cleaned or generated, replicable code is available through the online LDS repository: <https://github.com/ESRC-CDRC/LocalDataSpaces>. Internal code within the SRS environment using secure data is available upon request.

Author contribution. Conceptualization: J.L.M., M.A.G., M.G., S.L., A.S., P.L.; Investigation: J.L.M., M.A.G., M.G., S.L.; Supervision: A.S., P.L.; Writing original draft: J.L.; Writing review & editing: M.A.G., M.G., A.S. All authors approved the final submitted draft.

References

- Affleck P, Westway J, Smith M and Schrecker G (2022) Trusted research environments are definitely about trust. *Journal of Medical Ethics*. <https://doi.org/10.1136/jme-2022-108678>
- Alexiou A, Riddlesden D and Singleton A (2018) The geography of online retail behaviour. In *Consumer Data Research*. London: UCL Press, pp. 96–109. <https://doi.org/10.2307/j.ctvqhsn6.10>
- Arribas-Bel D, Green M, Rowe F and Singleton A (2021) Open data products- A framework for creating valuable analysis ready data. *Journal of Geographical Systems* 23, 497–514. <https://doi.org/10.1007/s10109-021-00363-5>
- CDRC (2022) Business Census: Open Access [data collection]. Available at <https://data.cdrc.ac.uk/dataset/business-census> (Accessed June 7, 2023).
- Delacroix S and Lawrence N (2019) Bottom-up data trusts: Disturbing the ‘one size fits all’ approach to data governance. *International Data Privacy Law* 9(4), 236–252. <https://doi.org/10.1093/idpl/izp014>
- Department of Health and Social Care (DHSC) (2022) NHS Test and Trace (England): Secure Access. [data collection]. Available at <https://www.gov.uk/government/publications/nhs-test-and-trace-statistics-england-methodology/nhs-test-and-trace-statistics-england-methodology>.

- EMG** (2021) COVID-19 Risk by Occupation and Workforce, Eightieth SAGE meeting on COVID-19, 11 February 2021, Scientific Advisory Group for Emergencies. Available at <https://www.gov.uk/government/publications/emg-covid-19-risk-by-occupation-and-workplace-11-february-2021>.
- Google LLC** (2022) Google COVID-19 Community Mobility Reports: Open Access. [data collection]. Available at <https://www.google.com/covid19/mobility/>.
- Green M** (2021) Thinking spatially to communicate and evaluate the roll-out of ‘mass’ testing in Liverpool, 2020. *People, Place and Policy* 15(1), 54–56. <https://doi.org/10.3351/ppp.2021.9589727428>
- Green M, Daras K, Davies A, Barr B and Singleton A** (2018) Developing an openly accessible multi-dimensional small area index of ‘access to health assets and hazards’ for Great Britain, 2016. *Health & Place* 54, 11–19. <https://doi.org/10.1016/j.healthplace.2018.08.019>
- Green M, García-Fiñana M, Barr B, Burnside G, Cheyne C, Hughes D, Ashton M, Sheard S and Buchan I** (2021) Evaluating social and spatial inequalities of large scale rapid lateral flow SARS-CoV-2 antigen testing in COVID-19 management: An observational study of Liverpool, UK (November 2020 to January 2021). *The Lancet Regional Health - Europe*, 6, 100107. <https://doi.org/10.1016/j.lanepe.2021.100107>
- Hardinges J** (2020) Data Trusts in 2020. Open Data Institute. Available at <https://theodi.org/article/data-trusts-in-2020/>.
- Hardinges J, Wells P, Blandford A, Tennison J and Scott A** (2019) Data Trusts: Lessons from Three Pilots. Open Data Institute. Available at <https://theodi.org/article/odi-data-trusts-report/>.
- Henggeler A, Stewart B, Bennett I, Ledden, S. and Egglestone S** (2021) Local Data Spaces: Pilot Study Evaluation. Administrative Data Research (ADR) UK. Available at <https://www.adruk.org/news-publications/news-blogs/local-data-spaces-pilot-demonstrates-importance-of-local-level-data-and-analysis-to-inform-local-decision-making-440/>.
- HM Land Registry** (2021) Price Paid Linked Data: Open Access. [data collection]. Available at <https://landregistry.data.gov.uk/app/root/doc/ppd>.
- Hubbard T, Reilly G, Varma S and Seymour D** (2020) *Trusted Research Environments (TRE)*. Green Paper. UK Health Data Research Alliance. <https://doi.org/10.5281/zenodo.4594704>
- Kavianpour S, Sutherland J, Mansouri-Bensassi E, Coull N and Jefferson E** (2022) Next-generation capabilities in trusted research environments: Interview study. *Journal of Medical Internet Research* 24(9), e33720. <https://doi.org/10.2196/33720>
- Macdonald J, Dolega L and Singleton A** (2022) An open source delineation and hierarchical classification of UK retail agglomeration. *Scientific Data* 9, 541. <https://doi.org/10.1038/s41597-022-01556-3>
- MHCLG** (2019) The English Indices of Multiple Deprivation: Open Access. [data collection]. Available at <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>.
- Office for Artificial Intelligence** (2021) National AI Strategy. Department of Digital Culture, Media and Sport. Available at <https://www.gov.uk/government/publications/national-ai-strategy>.
- ONS** (2019) Data Science Campus Faster Indicators. Available at <https://datasciencecampus.ons.gov.uk/faster-indicators-of-uk-economic-activity/>.
- ONS** (2021) Differential impacts of the Coronavirus pandemic on men and women, Eighty-fourth SAGE meeting on COVID-19, 25 March 2021, Scientific Advisory Group for Emergencies. Available at <https://www.gov.uk/government/publications/ons-differential-impacts-of-the-coronavirus-pandemic-on-men-and-women-24-march-2021>.
- ONS** (2022a) Annual Survey of Hours and Earnings, 1997-2022: Secure Access. [data collection]. 21st Edition. UK Data Service. SN: 6689. <https://doi.org/10.5255/UKDA-SN-6689-20>
- ONS** (2022b) Death Registrations in England and Wales, 1993-2021: Secure Access. [data collection]. 8th Edition. UK Data Services. SN: 8200. <https://doi.org/10.5255/UKDA-SN-8200-8>
- ONS** (2022c) Estimates of the population for the UK, England, Wales, Scotland and Northern Ireland: Open Access. [data collection]. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>.
- ONS** (2023a) Business Insights and Conditions Survey: Waves 1-70, 2020-2022: Secure Access. [data collection]. 16th Edition. UK Data Service. SN: 8653. <https://doi.org/10.5255/UKDA-SN-8653-16>
- ONS** (2023b) Business Register and Employment Survey, 2009-2021: Secure Access. [data collection]. 12th Edition. UK Data Service. SN: 7463. <https://doi.org/10.5255/UKDA-SN-7463-12>
- ONS** (2023c) Business Structure Database, 1997-2022: Secure Access. [data collection]. 15th Edition. UK Data Service. SN: 6697. <https://doi.org/10.5255/UKDA-SN-6697-15>
- ONS** (2023d) Coronavirus (COVID-19) Infection Survey QMI: Secure Access. [data collection]. Available at <https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/covid19infectionsurvey>.
- ONS** (2023e) House price to workplace-based earnings ratio: Open Access. [data collection]. Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/ratioofhousepricetoworkplacebasedearningslowerquartileandmedian>.
- ONS Social Survey Division** (2023a) Annual Population Survey, 2004-2021: Secure Access. [data collection]. 21st Edition. UK Data Service. SN: 6721. <https://doi.org/10.5255/UKDA-SN-6721-20>
- ONS Social Survey Division** (2023b) Northern Ireland Statistics and Research Agency, Central Survey Unit. Quarterly Labour Force Survey, 1992-2022: Secure Access. [data collection]. 35th Edition. UK Data Service. SN: 6727. <https://doi.org/10.5255/UKDA-SN-6727-28>
- Paprica P, Sutherland E, Smith A, Brudno M, Cartagena R, Crichlow M, Courtney B, Loken C, McGrail K, Ryan A, Schull M, Thorogood A, Virtanen C and Yang K** (2020) Essential requirements for establishing and operating data trusts: Practical

- guidance co-developed by representatives from fifteen Canadian organizations and initiatives. *International Journal of Population Data Science* 5(1), 31. <https://doi.org/10.23889/ijpds.v5i1.1353>
- Ricciato F, Wirthmann A and Hahn M** (2020) Trusted smart statistics: How new data will change official statistics. *Data & Policy* 2, e7. <https://doi.org/10.1017/dap.2020.7>
- Singleton A, Alexiou A and Savani R** (2020) Mapping the geodemographics of digital inequality in Great Britain: An integration of machine learning into small area estimation. *Computers, Environment and Urban Systems* 82, 101486. <https://doi.org/10.1016/j.compenvurbsys.2020.101486>
- Stalla-Bourdillon S, Carmichael L and Wintour A** (2021) Fostering trustworthy data sharing: Establishing data foundations in practice. *Data & Policy* 3, e4. <https://doi.org/10.1017/dap.2020.24>.
- UK AI Council** (2021) Exploring legal mechanisms for data stewardship. The Ada Lovelace Institute. Available at <https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/>.
- Vindrola-Padros C** (2019) What is rapid research and why is it relevant for health care? Nuffield Trust comment. Available at <https://www.nuffieldtrust.org.uk/news-item/what-is-rapid-research-and-why-is-it-relevant-for-health-care#what-are-we-working-on-now>.
- Zhang X** (2021) A commentary of data trusts in MIT technology review 2021. *Fundamental Research* 1(6), 834–835. <https://doi.org/10.1016/j.fmre.2021.11.016>

Cite this article: Macdonald J. L., Green M. A., Gibin M., Leech S., Singleton A and Longely P (2023). Local Data Spaces: Leveraging trusted research environments for secure location-based policy research in the age of coronavirus disease-2019. *Data & Policy*, 5: e22. doi:10.1017/dap.2023.14