Institute for Global Prosperity

Using administrative data and artificial intelligence to understand community well-being

Project Team: Lasana Harris[1], Nonso Nnamoko[2], Saffron Woodcraft[3], Saite Lu[4], Jose Gana[3], Jack Procter[1,2], Mrinal Chaudhary[1]

July 2025

Acknowledgements

This project has been funded by the British Academy and Nuffield Foundation collaboration on Understanding Communities. The project team would like to thank the many contributors to this work who attended ethics workshops, served on the advisory board, or provided feedback and advice on the project, including Shayda Kashef, Mary Cowan, Rowan Conway, Yamini Cinamon Nair, Paul Calcraft, Ladan Dirie, Julliet Whitworth, Professor Diane Coyle, Professor Yannis Korkontzelos, Nick Turner, Omid Shiraji, Michal Shinwell, Lara Groves, Professor Lucia Reisch and Afroditi Tsourgianni. We also wish to thank Olivia Stevenson and the Public Policy Team in RIGE at UCL for feedback on earlier versions of this report.

The views expressed are those of the authors and not necessarily of the British Academy or the Nuffield Foundation.

The Nuffield Foundation

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory.

www.nuffieldfoundation.org

The British Academy

The British Academy (BA) is the UK's national academy for the humanities and social sciences. It mobilizes these disciplines to understand the world and shape a brighter future. From AI to climate change, from building prosperity to improving well-being – today's complex challenges can only be resolved by deepening our insight into people, cultures and societies. The BA invests in researchers and projects across the UK and overseas, engages the public with fresh thinking and debates, and brings together scholars, government, business and civil society to influence policy for the benefit of everyone.

https://www.thebritishacademy.ac.uk/

Design by TC & Friends

Cover Footnotes:
1 Department of Psychology, University College London
2 Department of Computer Science, Edgehill University
3 Institute for Global Prosperity, University College London
4 Emmanuel College, University of Cambridge

Contents

Glossary of terms

Community well-being The collective measure of a community's (of place, interest, or identity) ability to meet present and future social, material, and individual needs and aspirations.

Personal well-being A person's evaluation of their own quality of life, also commonly known as subjective well-being.

Administrative data Data about a population collected and held by the government or other public companies or agencies.

Behavioural data Administrative or other data that records actual behaviours and acts.

Synthetic data A simulated dataset that mimics real-world data. It is created using algorithms and statistical models to replicate the characteristics of a real-world dataset, while protecting sensitive personal data.

Algorithm A computer programme or script that describes how data should be aggregated, processed, and inform goals or outcomes.

Artificial intelligence (AI) Artificial intelligence (AI): Algorithms that aggregate, process, and learn from data in the service of a goal or outcome.

Labour productivity Labour productivity: Labour productivity is a measure of how much economic output is generated per hour worked for the whole economy and for each industry or sector. The Office for National Statistics calculates labour productivity by dividing gross value added (GVA) by hours worked for that industry or sector.

Executive Summary

This report summarises the findings of an interdisciplinary research project exploring the viability and utility to local authorities of using administrative data—data that records residents' behaviours rather than self-reported perception data collected in surveys—and AI to gain insight into place-based community well-being.

Administrative data is currently widely used in the private sector to understand consumer behaviour. Yet it is under-utilized in the public sector, even though data recording library membership, parking charges, noise complaints, and recycling are routinely collected by local authorities. These are rich data sources that can be leveraged to better understand community well-being. Given the financial constraints faced by many local authorities and the high costs of collecting survey data, using administrative data could be a cost-effective way to make the most of an under-used public resource. Similarly, the use of AI by the public sector has been proposed by government to find efficiencies and cost-savings.

Over the past two decades, understanding and improving individual, collective, and national well-being has come to the fore in policy debates. This has been driven by rising inequalities, the decoupling of economic growth and living standards, and widespread calls for noneconomic measures to supplement GDP, following publication of the Commission on the Measurement of Economic Performance and Social Progress, more commonly known as Stiglitz-Sen-Fitoussi report, in 2009[1]. In the UK, the number of frameworks and indices assessing individual and community well-being has proliferated, reflecting interest in a diversity of measurement geographies and ways of conceptualizing personal and subjective well-being. Most of these indices are based on secondary survey data.

This project takes a different starting point by seeking to explore the feasibility of using administrative data that is routinely collected by local authorities, and AI to generate an algorithm to measure community well-being. The goal is to explore the practical and ethical issues that arise in the process of utilizing local authority administrative data, and to determine whether administrative data can generate added value, either in terms of gaining additional insights to community well-being or effectively addressing the drawbacks of collecting self-report survey data, such as timeliness and cost.

The research team partnered with the London Borough of Camden Council to explore these issues as part of the Council's wider initiative to develop an innovative community wellbeing measurement framework at both neighbourhood and Borough-wide geographies. Researchers worked with Camden Council officers to investigate the practical and ethical challenges involved in using administrative data, including concerns around curating disparate datasets, data privacy, consent, and synthetic data. A proof-of-concept Community Well-being Index Dashboard was developed as a demonstrator case for using administrative data and an AI algorithm to generate community well-being scores for each local authority in England. The reason for testing the Index against productivity is to see whether administrative data and AI can meaningfully explain differences in local economic performance. Evaluation of the Community Well-being Index shows that it correlates more strongly with productivity compared to self-report Subjective Well-being measures. This suggests that if a local area is underperforming, policymakers could potentially look at which aspects of community well-being—such as housing, education, or social connection—are falling behind. Focusing on these weaker areas could help improve both the well-being of local residents and the area's overall productivity, supporting more sustainable and inclusive development.

1 Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. http://www.stiglitz-senfitoussi.fr/documents/rapport_anglais.pdf

The report identifies four key findings about the potentials and shortfalls of using administrative data and AI to measure community well-being:

1 *Local authorities do not currently have the capacity or infrastructure to enable the widespread use of administrative data to measure community well-being. Government is leading dialogues about the upgrading of national data infrastructure across the UK. Conversations with ADR-UK and the Office for National Statistics reveal a great desire to enhance public bodies' ability to compare across data sets held by different agencies, and even within a single agency. However, there is, as yet, no unified approach, which impedes projects like this and limits the potential of AI technologies in government. Harmonizing data structures within and across departments, linking data sets better, and training in data science for central and local government officials will only strengthen the infrastructure unlocking better insights into community well-being.*

2 *Along with data infrastructure reform, we recommend legal and policy reform around data and data governance including debate on public and community data ownership and use. Community members need to view their data as a resource and community groups should play a role in not only shaping the assessment of their community's well-being but in controlling who has access to their data and for what purpose. Legal protection for data rights, including consent, use, privacy, exclusion, and protection of communities and groups less able to exercise such rights is necessary to promote public adoption of AI approaches within government and beyond.*

3 *Experts should work alongside community members to outline potential for misuse of data, and discussions around data privacy infrastructure and institutional trust. However, governments are a special case of data use given their unique position as servants of the public and creators and enforcers of regulation. While communities readily make available their data to private companies in exchange for access to greatly valued services, which is already a problem and has created huge dependencies on such private companies, they are more cautious when considering government use of data. Indeed, the public seems to trust private companies more than government with regards to their personal data. Therefore, making AI technologies and their uses explainable, transparent, and decided in conjunction with the public appears to be critical for the development of this field.*

4 *Community well-being is complex and context specific. Further research is needed to investigate the multiple dimensions of community well-being and to better understand the dynamics between community well-being and other socio-economic variables.*

Administrative data and AI hold enormous potential for local government to generate insights about community well-being and other complex social policy challenges like understanding health inequalities and poverty. Insights from this project show that such benefits require not only further research and more test cases but coordinated responses to the infrastructural and legal challenges identified here.

Why use administrative data and artificial intelligence to measure community well-being?

Over the past two decades, understanding and improving individual, collective, and national well-being has come to the fore in UK policy debates. Since the UK Measures of National Wellbeing launched in 2010 - reporting on personal well-being, social connections, mental health, civic participation, and satisfaction with place - the number of indices assessing individual and community well-being at a local and national level has proliferated (Moore and Woodcraft 2023).[2] Co-op's Community Well-Being Index[3], Greater London Authority's London Wellbeing and Sustainability Measure[4], Thriving Places Index[5], and Community Needs Index[6] are among the many frameworks and tools that have emerged, all experimenting with different ways to measure the well-being, strength, and resilience of local communities. These tools, and others like them, aim to generate more holistic evidence and insights for policymakers about the social and economic life of communities to supplement conventional economic measures like Gross Domestic Product (GDP).[7] These approaches to measuring community well-being use a broad range of indicators from secondary datasets, often combining objective measures such as employment rates, public health, and air quality, with subjective measures such as self-reported perceptions of life satisfaction and happiness.

Social science research has long shown that attitudes, a large focus of self-report data, do not effectively predict behaviour. Self-report attitudinal surveys are susceptible to personal biases and demand effects, where survey conditions or survey questions themselves influence responses, and cannot take account of unconscious influences on behaviour. Furthermore, primary survey data is costly to collect and analyse.

Administrative data that records the actual behaviours of citizens, communities, service users, has characteristics that could make it well-suited to provide complementary insights to survey data about community well-being. Local authorities routinely collect administrative data about the behaviours and activities of community members in the process of delivering services. Parking charges, noise complaints, library memberships, community centre bookings, council tax payments, and closed-circuit television feeds are typical examples. This type of administrative data is widely used in the private sector to understand consumer behaviour. However, ethical concerns, capacity constraints, and obstacles to data sharing are among the reasons why administrative data is under-used by local government and other public agencies. Among the issues identified by this project are: a lack of data infrastructure that inhibits data sharing between services and directorates within local authorities; capacity constraints that make processing and analyzing large datasets challenging; ethical concerns around the level of government access to certain personal behaviours (such as those derived from video, physiology, or other sensitive behaviours); and lack of consent for future uses at the point of data collection.

Despite these challenges, we argue administrative data is an untapped public resource with significant potential to generate locally specific, policy-relevant insights and evidence, in this case, to better understand community well-being. Given the extreme pressures on local government finances, using existing administrative data could be a cost-effective way to generate new insights about community well-being.

2 Moore, H. L. and S. Woodcraft (2023). Local meanings and 'sticky' measures of the good life: redefining prosperity with and for communities in east London. Prosperity in the Twenty-First century: Concepts, models and metrics. H. L. Moore, D. Matthew, N. Mintchev and S. Woodcraft. London, UCL Press.

3 Co-op/Young Foundation.The Community Wellbeing Index https://communitywellbeing.coop.co.uk/

4 Mayor of London. The London Wellbeing and Sustainability Measure https://apps.london.gov.uk/wellbeing

5 Centre for Thriving Places. Thriving Places Dashboard https://www.centreforthrivingplaces.org/thriving-places-index/

6 Local Trust/OCSI. Community Needs Index https://ocsi.uk/2023/05/24/community-needs-index-2023/

7 Stiglitz, J. E. (2010). Mismeasuring our lives: Why GDP doesn't add up. New York, The New Press.

Stage 1

What is community well-being and how is it currently measured?

Key points

1 *There is no single widely accepted definition of community well-being in theory, policy, or practice.*

2 *The community well-being measurement landscape in the UK includes a growing number of indices and frameworks with a common methodology – self-report subjective measures from secondary survey data, combined with objective measures of economic and environmental conditions.*

3 *Most measurement tools use a definition of community well-being from academic research or developed by government policymakers - opportunities for communities of place or interest to have a voice in defining community well-being and how it should be measured are limited.*

In the first stage of the project, we carried out a review of well-being literature and measurement frameworks to examine how community well-being is defined in research, policy and practice. Next, we describe the process and findings from this work.

1.1 What is the difference between personal well-being and community wellbeing?

In the past decade, the number of well-being indices and frameworks has proliferated, reflecting the growing importance of well-being as a policy agenda and the challenges of defining and measuring a concept that is acknowledged to be ambiguous and contextual. Well-being is acknowledged to be an ambiguous concept that is often used interchangeably with overlapping concepts like happiness, quality of life, life satisfaction (Oishi et al., 2013).[8]

Broadly speaking, definitions of well-being fall into two main categories: subjective (or personal) well-being and community well-being. Definitions of subjective well-being have historically inclined themselves either to a 'hedonic' or 'eudaimonic' conceptualization. The hedonic view emphasizes emotions, feelings, and a cognitive or evaluative dimension, while the eudaimonic view foregrounds the realization of a person's potential (Ryan and Deci, 2001, Das et al., 2020).[9, 10] A well-known definition of subjective well-being is that of Diener et al., (2002, 2009)[11, 12], who focus on an individual's cognitive and affective evaluations of his or her life, regardless of how others see it. According to Dolan and Metcalfe (2012)[13], there are three main concepts of subjective well-being in the literature – evaluation (life satisfaction), experience (momentary mood) and eudaimonia (purpose) – and they argue that policymakers should seek to measure all three.

In the UK, the Office for National Statistics Personal Well-being Measure adopts a hedonic view inviting people to evaluate their feelings of well-being and life satisfaction using four standardized questions (ONS4) in national and regional surveys.[14]

The ONS4 are:
Life Satisfaction: "Overall, how satisfied are you with your life nowadays?"

Feeling life is worthwhile: "Overall, to what extent do you feel the things you do in your life are worthwhile?"

Happiness: "Overall, how happy did you feel yesterday?"

Anxiety: "Overall, how anxious did you feel yesterday?"

Globally, the annual World Happiness Report uses self-report survey data from the Gallup World Poll, which also invites people to evaluate life satisfaction, positive and negative emotions.[15] The use of standardized measures enables international comparisons of subjective well-being, although this approach has received criticism for its lack of attention to different cultural interpretations and meanings of well-being.[16]

The Organization for Economic Co-operation and Development (OECD) adopts a broader definition of subjective well-being in the context of its "How's Life?" initiative[17], which examines quality of life in combination with material living conditions and sustainability. The World Health Organization (WHO) conceptualizes subjective well-being largely based on the OECD framework. These frameworks move closer to the concept of community wellbeing, which considers collective resources and responsibilities alongside individual evaluations of quality of life.

Definitions of community well-being also vary. Wiseman and Brasher (2008) define community well-being as "the combination of social, economic, environmental, cultural, and political conditions identified by individuals and their communities as essential for them to flourish and fulfil their potential."[18] This definition recognizes that 'community' involves more than a simple aggregation of individual values at a larger scale – conveying a larger sense of shared group values, places, and experiences (Atkinson et al., 2017).[19]

Diener et al., (2009) propose that collective well-being is "about the civic virtues and the institutions that move individuals toward better citizenship: responsibility, nurturance, altruism, civility, moderation, tolerance, and work ethic."[20] This is heavily inclined towards the eudaimonic conceptualization of well-being at the group level.

8 Oishi, S., Graham, J., Kesebir, S. & Galinha, I. C. 2013. Concepts of happiness across time and cultures. Pers Soc Psychol Bull, 39, 559-77.

9 Ryan, R. M. & Deci, E. L. 2001. On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being. Annual Review of Psychology, 52, 141-166.

10 Das, K. V., Jones-Harrell, C., Fan, Y., Ramaswami, A., Orlove, B. & Botchwey, N. 2020. Understanding subjective well-being: perspectives from psychology and public health. Public Health Reviews, 41, 25.

11 Diener, E., Lucas, R. E. & Oishi, S. 2002. Subjective well-being: The science of happiness and life satisfaction. Handbook of positive psychology. New York, NY, US: Oxford University Press.

12 Diener, E. 2009. The science of well-being: The collected works of Ed Diener, New York, NY, US, Springer Science + Business Media.

13 Dolan, P. & Metcalfe, R. 2012. Measuring Subjective Wellbeing: Recommendations on Measures for use by National Governments. Journal of Social Policy, 41, 409-427.

14 Office for National Statistics Personal Well-being User Guidance (2024). https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide#:~:text=The%20ONS4%20measures%20ask%20people,period%20(both%20positive%20and%20negative)

15 Helliwell, J. F., Richard Layard, Jeffrey Sachs, and Jan-Emmanuel De Neve 2021. World Happiness Report 2021. In: John Helliwell, R. L., Jeffrey D. Sachs, Jan-Emmanuel De Neve, Lara Aknin, Shun Wang (ed.) New York: Sustainable Development Solutions Network.

16 Fadijia, A. W., Meiring, L., & Wissing, M. P. (2019). Understanding well-being in the Ghanaian context: Linkages between lay conceptions of well-being and measures of hedonic and eudaimonic well-being. Applied Research in Quality of Life. https://doi.org/10.1007/s11482-019-09777-2

17 OECD 2011. How's Life? Measuring well-being. OECD publishing.

18 Wiseman, J. & Brasher, K. 2008. Community Wellbeing in an Unwell World: Trends, Challenges, and Possibilities. Journal of public health policy, 29, 353-66.

19 Atkinson, Sarah, Anne-Marie Bagnall, Rhiannon Corcoran, Jane South, with Sarah Curtis, Salvatore di Martino, Gerlinde Pilkington (2017). What is Community Wellbeing? UK: What Works Wellbeing.

20 Diener, E. 2009. The science of well-being: The collected works of Ed Diener, New York, NY, US, Springer Science + Business Media.

A 2017 systematic review of community well-being indicators identified 43 measures or indices in use in the UK by governmental agencies, academic researchers, and civil society organisations with new measurement tools being added to this list regularly.[21] The review identified community well-being as less clearly defined than personal well-being. This lack of clarity can be seen in the wide range of synonyms and concepts evident in the diversity of measurement frameworks, which have developed to foreground aspects of well-being. These include the Thriving Places Index, the Social Progress Index, and city-focused measures like the Young Foundation's Civic Strength Index, and the Greater London Authority's new London Wellbeing and Sustainability Measure.

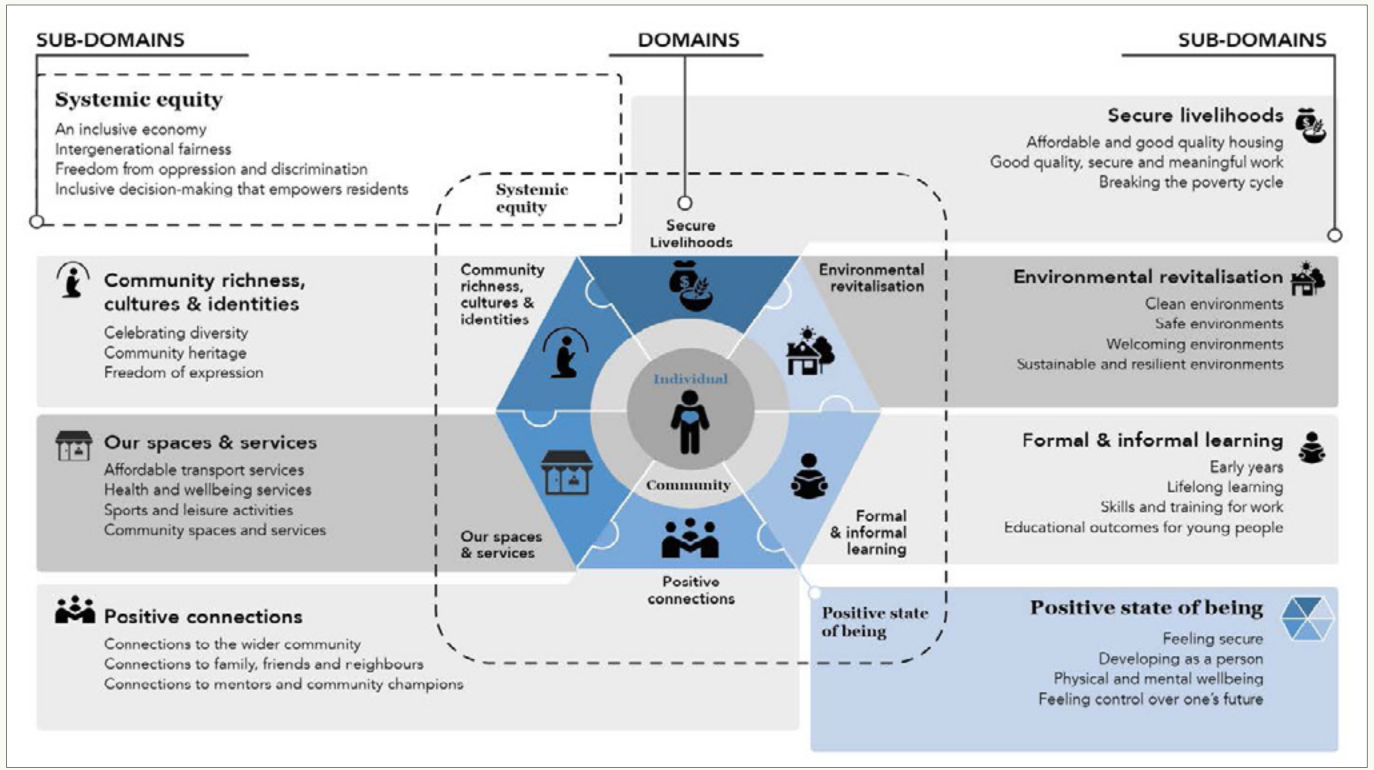
Despite this diversity of work, a widely accepted definition of community well-being does not exist in theory, policy, or practice. For the purposes of this project, we follow the What Works Centre for Wellbeing (an independent body for well-being evidence, policy, and practice in the UK) in adopting Wiseman and Brasher's definition of community well-being.

Community well-being is "the combination of social, economic, environmental, cultural, and political conditions identified by individuals and their communities as essential for them to flourish and fulfil their potential."

Wiseman and Brasher (2008) Community Wellbeing in an Unwell World: Trends, Challenges, and Possibilities. Journal of public health policy, 29, 353-66.

21 Atkinson, Sarah, Anne-Marie Bagnall, Rhiannon Corcoran, Jane South, with Sarah Curtis, Salvatore di Martino, Gerlinde Pilkington (2017). What is Community Wellbeing? UK: What Works Wellbeing.

Figure 1: Good Life Euston Model Domains and Sub-Domains



1.2 Reviewing the UK's community well-being measurement landscape

1.2.1 Good Life Euston: A co-produced wellbeing measure

Camden Council is leading efforts to innovate in community well-being measurement, developing novel approaches that involve citizen scientists, residents, and community networks in defining what a good life is and how it should be measured.

Good Life Euston is a community-based and community-led research project about the drivers of, and obstacles to, a good life for people living in Euston, an area of the Borough affected by long-term regeneration linked to HS2 and the redevelopment of Euston station. The purpose of the Good Life Euston project is to understand how local communities are affected by major physical, economic, and social transformation and for residents to have a voice in determining how the outcomes of change should be measured. The Good Life Euston model is one of the key outputs from the project (see figure 1). Co-produced by citizen social scientists (local residents trained by UCL's Citizen Science Academy to work as social researchers in their neighbourhoods) drawing on their qualitative research, the Good Life Euston model is a multi-dimensional conceptual framework representing the individual and collective resources, conditions, and outcomes that enable people to thrive and live well. It is the model underpinning the Good Life Euston Index - a new local prosperity measure that Camden Council and partners are using to assess the impacts and outcomes of regeneration on local communities and has been used as the basis for developing Good Life Camden – a Borough-wide community well-being measure based on secondary data.

The Good Life Euston Index reports on eight domains, 28 sub-domains, and over 60 indicators. Like the community well-being indices discussed in section 1.2, the Good Life Index primarily uses self-report survey data collected from 3,000 households in the Euston area to report on whether communities are thriving or struggling.[22] For example, the Secure Livelihoods domain includes composite indicators reporting on real household disposable income, job satisfaction, debt burdens, ability to save and keep accommodation warm, and the Positive State of Being domain includes composite indicators reporting on perceptions of control and agency, and how secure people feel about the future.

Before adopting Camden's Good Life Euston model as the framework for the demonstrator Community Well-being Index, the research team the landscape of community well-being measures in the UK to identify other indices and dashboards currently used by government, local authorities, and public agencies. The purpose of this exercise was to compare the characteristics of these tools to determine:

How community well-being is defined and what role if any, communities play in definition setting

The rationale and logic for each Index or measurement framework - How indices are constructed including data sources and metrics

Identifying common domains and indicators of community well-being across the different frameworks to establish if key domains were missing from the Good Life Euston model.

22 The first Good Life Euston Index will launch in 2025 based on household survey data collected in the last six months of 2024.

1.2.2 Mapping Good Life Euston to other community well-being measures

There are several community well-being measures and indices in the UK that are used by local government and public agencies. These include national, local authority and city-level measures that use different conceptual frameworks and indicators, such as:

Global, regional and national: The Social Progress Index reports on 12 components of progress globally, nationally, regionally, and is beginning to develop local government and city-level tools. Some of the Social Progress Index domains overlap with definitions of community well-being although it does not explicitly use the term well-being.

National: Wellbeing of Wales measures national progress against seven goals intended to improve the social, economic, environmental, and cultural well-being of Wales. The seven goals are contained in law under the Well-being of Future Generations (Wales) Act 2015. The seven goals are: a prosperous Wales; a resilient Wales; a healthier Wales; a more equal Wales; a Wales of cohesive communities; a Wales of vibrant culture and thriving Welsh language; a globally responsible Wales. Over 50 indicators are used to measure these seven goals.

Local authority and city-level: Several indices focus on the local authority and city-level including the Centre for Thriving Places Thriving Places Index and the Greater London Authority's London Wellbeing and Sustainability Measure.

Community-level: The Co-op Community Well-being Index measures three pillars (people, place, relationships) and nine domains (education and learning, health, economy work and employment, culture heritage and leisure, transport mobility and connectivity, housing space and environment, relationships and trust, equality, voice and participation). This Index is a partnership between Co-op, The Young Foundation, and Geolytix, and uses Geolytix 'Seamless Locales' tool as the reporting geography. Seamless Locales have been created to reflect areas people identify as their neighbourhood or community, and on average have 2,230 inhabitants and 973 homes. There are 28,317 locales in the Co-op Community Well-being Index.[23]

As this project is focused on using local government administrative data, we selected four indices that measure community well-being at this scale to review how they are constructed and to map common domains and indicators. The four indices are:

The London Sustainability and Well-being Measure

The London Borough of Barking & Dagenham's Social Progress Index

The Thriving Places Index

The Co-op Community Well-being Index

Figure 2 illustrates the domains and sub-domains in these four indices and identifies where common themes exist. For example, all four indices include health, employment/work/income, belonging and trust, social relationships, subjective experience of the economy, and the environment. Education, skills and learning are included in three of the frameworks. Each framework contained specific dimensions (e.g. transport, culture and heritage, mobility, vibrancy of community) and each measured domains and subdomains differently, including the weighting of the importance of each dimension. All measures relied on a combination of self-report survey, economic, and service provision data that measure proxies of behaviour relevant for community well-being.

The Co-op Community Wellbeing Index and London Sustainability and Wellbeing Measure involved communities and civil society organisations in participatory processes to define community well-being and consider the importance of different dimensions of wellbeing.

Following the review of community well-being indices described here, we adopted the Good Life Euston model as the framework for developing an Index using administrative data and AI. The Good Life Euston model does not explicitly use the language of community wellbeing, yet it shares multiple domains and subdomains with other community well-being indices. As it was co-produced by Camden residents based on local priorities and lived experiences, we felt it provided a good baseline from which to develop a proof-of-concept administrative data index.

23 Hill-Dixon, A., Suzanne Solley and Radhika Bynon (n.d). Being Well Together: The creation of the Co-op Community Wellbeing Index. The Young Foundation/Coop/Geolytix.

Figure 2: Mapping shared domains and sub-domains (in parenthesis) in four community well-being indices

London Sustainability and Well-being Measure	Barking & Dagenham Social Progress Index	Thriving Places Index	Co-op Community Wellbeing Index
Being healthy	Health and well-being	Equality (Gender, ethnicity, income, social, health)	People (Health, education and learning, economy, work and employment)
Good employment and opportunities to succeed	Employment/Enterprise	Local conditions (Education, economy, place, community, health)	Place (culture, heritage and leisure, transport, mobility and connectivity, housing, space and the environment)
Positive connections and belonging	Community engagement	Sustainability (Green infrastructure, waste, energy use)	Relationships (Relationships and trust, equality, voice and participation)
Having a decent home	Housing		
A clean and sustainable environment	Environment		
Feeling financially secure	Deprivation		
Vibrant neighbourhoods with accessible services	Skills and education		
	Safety		

Stage 2

What are the ethics and challenges of using administrative data and AI to measure community wellbeing?

Key points

1 *Public concerns about personal data processing, sharing and ownership are a critical concern and obstacle to use of local government administrative data.*

2 *Local authorities currently lack the data infrastructure to process and maximise the use of administrative data.*

3 *Synthetic data can guarantee privacy and address concerns about government use of administrate data.*

In this stage of the project, we examined the ethical implications and practical challenges of using administrative data and AI to measure community well-being. Three expert workshops were convened, each exploring a different ethical and data challenge. The workshops involved relevant stakeholders, including representatives from Administrative Data Research UK, The Office of National Statistics, The Office for Statistics Regulation, the Ada Lovelace Foundation, the Nuffield Foundation, Camden Council, The Greater London Authority, Jakarta Smart Cities Initiative, and SHIFT London, who were invited to participate because of their expertise in data in government, community well-being measurement, and data policy. The aim of the workshops was to identify ethical and practical concerns both to inform the development of the proof-of-concept Community Well-being Index and to consider issues that should be incorporated into best practice guidelines for future work in this area.

2.1 Workshop 1 — Ethics and challenges of using administrative data

This workshop explored the ethics and challenges of using local government administrative data for community insight with a group of officers from Camden Council. A summary of key discussion points and implications for this project and wider research in this area are below (see Appendix for full recommendations).

Data privacy and trust in government: Public concerns over personal data privacy, processing, sharing, and use across various platforms is a critical concern. In this context, there is a trade-off between individuals' need to access public services and their right to maintain privacy over their data and authority over the purposes for which their data are used. There is a need for transparency and ethical handling where the public should be able to perform their cost-benefit analysis regarding their data usage. A gap in public understanding of data processes is a major trust barrier that hampers effective data sharing practices. Enhanced public information campaigns and targeted educational initiatives seem essential to overcome these barriers, along with making public institutions and entities more trustworthy so that the public can better trust them with their personal data. Variability in the definition of 'public good' across different sectors is another challenge for data-driven policy decision-making. This discrepancy makes it difficult to formulate a universal framework that satisfies all stakeholders involved in data handling and sharing. For example, the criteria for what constitutes public good in housing differ markedly from those in social care or library services.

Practical data challenges: A Community Well-being Index will rely heavily on data from different data sources such as the National Health Service, local councils, or police crime statistics. The first substantial obstacle to using and integrating such data sources involves the data infrastructure within such organizations. This infrastructure creates complexities related to accessing and sharing data because of organizational size and system incompatibility. Siloed working environments further contribute to this obstacle. For instance, each unit in the same council may operate like a distinct business, using unique or incompatible data systems that complicate data sharing even within the same council. Finally, bureaucratic red tape and organizational complexities further hinder efficient data sharing and integration.

One notable internal solution implemented by some councils is the 'OneView' platform, which aggregates data from multiple sources to provide a comprehensive view of individuals or households. This aids frontline officers while delivering better support to residents and communities. However, this solution faces challenges in information governance and ethics since it compiles detailed and sensitive personal data. Due to the vast nature of councils, NHS departments and other organizations, governance frameworks must be in place to manage who has access to the collated data. Moreover, clear data-sharing protocols can ensure public benefit and address ethical considerations surrounding the secondary use of data initially collected for specific purposes. Indeed, a unified database presents magnified security concerns and vulnerabilities to hackers and actors with ill intent. However, tools that facilitate access and linkage across separate databases would be useful to facilitate this work.

2.2 Workshop 2 — Ethics and challenges of using synthetic data

The second workshop focused on the potential of synthetic data to navigate some of the ethical concerns identified in the previous workshop. Synthetic data—a novel dataset that maintains the underlying statistical structure of an original dataset—has the potential to address ethical concerns related to privacy and data permissions. Workshop discussions focused on three questions:

What are the potentials and pitfalls of synthetic data use?

How viable is the use of synthetic data by local authorities?

What implications for data ownership and stewardship arise in the use of synthetic data?

A summary of key discussion points and implications for this project and wider research in this area are below (see Appendix for full recommendations).

Technical shortcomings of synthetic data: Concerns about data quality, freshness, the ethical implications of data manipulation, and the potential perpetuation of biases from original datasets remain even with the use of synthetic data. A critical concern is the generation of synthetic data from outdated datasets, which could lead to misleading analytics and decisions. The complexity of translating qualitative data into synthetic formats is another concern. Ensuring high fidelity in synthetic datasets is another challenge. Natural language processing (NLP) approaches might help data translation yet may degrade fidelity in the dataset. The potential misuse of low-fidelity synthetic data and the ethical need for stringent governance frameworks to manage synthetic data applications also remain a concern. Synthetic data created today might become less relevant or potentially harmful over time if used to inform future policy decisions.

Real-world checks on synthetic data: To address concerns regarding the relevance and reliability of synthetic data over time, especially its ability to reflect real-world changes and complexities, policymakers should regularly compare synthetic datasets with real-world developments to ensure their continued relevance and accuracy. The dynamic nature of societal trends—such as shifts in how wealth is indicated—requires that synthetic data models adapt to reflect these changing realities. It is important to compare forecasted synthetic datasets to real-world developments regularly and advancements in AI might influence the future use of synthetic data. AI has the potential to reduce human biases in data generation but also requires continuous human oversight and accountability. Ethical guidelines and governance frameworks are essential to managing the risks associated with synthetic data, such as privacy breaches and bias perpetuation.

Implications for data ownership and stewardship: A key question for this project, and for wider debates on data ownership, is who should own data used in a Community Well-being Index? For the data used in this project, local councils and other government agencies are the data curators. Maintaining this system allows for the management and equitable use of data. This approach is also further justified by devolution, which will increase the powers and responsibilities of local authorities and city regions. Data is a potential source of income for local governments if sold to corporations. However local government ownership is not the only model. There is an argument that individuals own their data. Rights arguments in this vein suggest people should own their data and make choices about how it is used. Data has value, and this value should remain with the individual. However, there needs to be public education about data, its value, and use cases for a system of individual ownership to operate effectively. The value of data arguments also suggests that data could be a national resource, owned centrally by governments. Both central and local government ownership are theoretically controversial because of the potential for misuse, as is corporate ownership. A radical alternative is that data is not owned by anyone but is available to everyone or collectively owned. This utopian ideal still requires a data custodian to ensure data quality and has the potential to violate individual rights if data is used for a public good that conflicts with the individual's preferences.

2.3 Workshop 3 — Ethics of using AI to measure community well-being

This workshop focused on two issues:

Public perceptions of AI and use of a human-centred approach to AI

Discussion of the evaluation of the newly developed Community Well-being Index and comparisons with traditional self-report Subjective Well-being measures (see next section - Stage 4)

Public perceptions of AI: AI is not a homogeneous concept, yet the public holds a perception of AI that suggests its use by the government is dangerous and undesirable. The discussion focused on how this largely negative perception of government-use of AI influences the viability of a Community Well-being Index, particularly when such an Index relies on administrative data collected by the government for other purposes.

A human-centred approach to AI: AI trained on biased datasets will reproduce bias. Current principles of AI drafted in 2019 by the OECD and the G20, signed by 50 governments globally, suggest that AI should be human-centred - developed to preserve, not violate human beings' rights - fair, transparent, safe, and accountable. Given the bias in existing datasets, how should AI be designed to ensure it is human-centred and fair? Ethics is determined by society, and shapes society, suggesting that ongoing public dialogue is needed to ensure AI continues to adhere to these principles. This is consistent with the idea of having community members determine how community well-being is defined and measured, enabling community members to renegotiate the conditions under which AI can provide additional insight.

Government use of AI — rights and responsibilities: Governments were originally designed to offer protection in exchange for taxation. This social contract, however, has since evolved where governments are now expected to also guarantee rights to their citizens. Accountability requires governments to track and trace violations of rights, enforcing responsibility to entities that may violate them. In the context of AI use by governments, responsibility implies a peer relationship where these rights are constantly renegotiated between citizens and their government. Trust is also a peer relationship and works best when defection from the relationship is possible. As such, trust becomes a tradition of fair cooperation that a community builds with its government. In addition, correcting data and having appropriate redress for decisions based on data is also important for building trust.

2.4 Ethical implications and best practice guidelines

The discussion of ethical concerns presented here does not resolve all the pertinent questions regarding local government's use of AI to generate community well-being insights. There are significant ethical, legal, and practical challenges that encompass data ownership, the use of personal data, public perceptions of AI, and trust in government, and this is a fast-moving area of science, policy, and technology. Workshop discussions identified a range of ethical and practical issues about the use of administrative data and AI, some of which are directly relevant to this project and have informed development of the Community Well-being Index. For example, concerns about data privacy and synthetic data. Other issues such as concerns about individual rights and data ownership, or the role of government in data stewardship and data reform, have much wider implications and connect with public and political debates about society and technology. While it is beyond the scope of this project to develop recommendations responding to all these concerns, we feel it is important to capture the discussions and proposals that emerged from the three workshops. A full summary of the ethical concerns, practical data challenges, and implications for best practice are included in Appendices A-D.

Stage 3

Developing a prototype: The Community Wellbeing Index

Key points

1 *Administrative data were converted to synthetic data by applying a transformer (a value) that changes the data points but preserves the underlying structure of the data.*

2 *A data-driven approach to weighting different indicators was adopted.*

3 *A functional proof-of-concept Index Dashboard, integrated with an algorithm to determine the Community Well-being score at local authority geographies in England, was created.*

This section describes the process of creating a proof-of-concept Community Well-being Index.

This stage of the project involved:

Working with Camden Council to identify potential sources of administrative data to measure community well-being

The creation of synthetic datasets to manage concerns about data privacy

Creating an algorithm

Creating a proof-of-concept data dashboard measuring community well-being at the local authority level

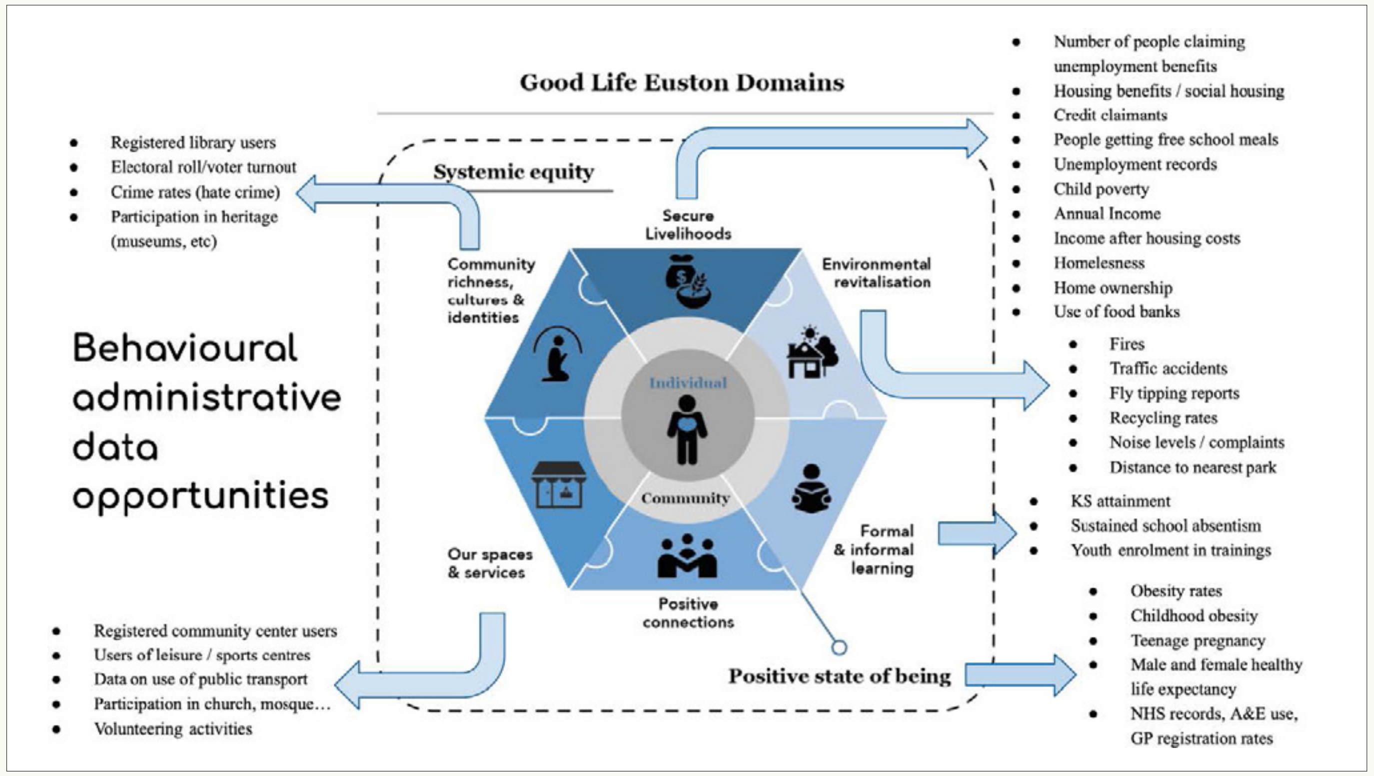
3.1 Mapping administrative datasets

As discussed in section 1.3, we adopted the Good Life Euston model as the definition of community well-being for this project. The Good Life Euston Index uses self-report survey data collected using a survey of 3,000 households in the Euston area to report on whether communities are thriving or struggling. The Good Life Euston Index reports on eight domains, 28 sub-domains, and over 60 indicators. For example, the Secure Livelihoods domain includes composite indicators reporting on real household disposable income, job satisfaction, debt burdens, ability to save and keep accommodation warm, and the Positive State of Being domain includes composite indicators reporting on perceptions of control and agency, and how secure people feel about the future.

To develop the Community Well-being Index for this project, we first worked with Camden Council officers to identify potential sources of administrative data routinely collected by Council services and contractors and to map this data against the domains and subdomains of the Good Life Index (see figure 3). The purpose of this exercise was to identify if sources of administrative data are available for all sub-domains of the Good Life Euston model and to explore the viability and challenges in accessing this data.

It was challenging to identify administrative data for some domains of the Community Wellbeing Index. Notable areas of missing data relate to various private and charitable enterprises, such as food bank usage and leisure/community centres. For the demonstrator Index, the team aimed to use as much administrative data as possible. Where administrative data was not available, the research team used publicly accessible secondary sources instead. In sum, 21% of the data sources, primarily those obtained from the ONS, were modelled on survey data. Such data, including annual income, income after housing, unemployment records, social housing, and home ownership, may be derived from sources other than surveys as novel approaches to data collection are implemented by the ONS. A full list of the data used in the Community Well-being Index is listed in Appendix E.

Figure 3: Mapping behavioural administrative datasets to domains and sub-domains of the Good Life Euston model



3.2 Creating a proof-of-concept Community Well-being Index Dashboard

In this stage of the project the goal was to develop an Index that could offer added value insights about community well-being for policy makers. Existing community well-being indices were examined (see discussion in Stage 1), noting their use cases and ongoing employment for governmental entities, primarily regarding what elements could be added to their existing methodologies, and how a data driven perspective could improve their utility.
Next, we describe the development of the algorithm and dashboard.

Algorithm development: To manage ethical concerns about working directly with administrative data, indicators were converted to synthetic data by applying a transformer (a random value) to existing datasets that changes the data points but preserved the relationship between these data points, keeping the underlying structure of the data.

A key question when designing a community well-being algorithm is how to weight the behaviours/variables that contribute to each domain. Weighting determines the relative importance of a behaviour to a domain, and the relative importance of a domain to community well-being, affecting the overall community well-being score. In this project, we used a data-driven approach to weighting where we relied on the median of each variable to determine its weight, and all domains contributed equally to the overall well-being score. Each indicator is assigned a value, using a weighting, that is converted into a score based on the best and worst cases of that indicator across all councils. Domains are scored from an average of all indicator scores therein, with the Index itself being a sum of all domains present within the model. With the insight provided by established and currently used indices, including their methodology and logic, the algorithm consists of many layered arithmetic averages of scoring based on presently available information. We did not employ sensitivity testing of our weighting procedure for the index since we relied on a data driven approach to determine weighting. This is a value neutral approach that determines the relative importance of each variable to the overall index. Therefore, sensitivity tests are irrelevant given that we are not testing a specific hypothesis or theory about the relative contribution of each index. Given that this project was a demonstration of an approach to assessing community well-being, and the assignment of weights is a crucial step in the coproduction of such as assessment, we hope future analyses using this approach would employ sensitivity tests to confirm the weighting chosen for various variables.

A key question for the future development of community well-being measures based on administrative data and AI, is who should determine how this weighting occurs? As described above, we determined the weighting of differing indicators via a data-driven approach, using the values from the datasets provided to assign weighting to each indicator within a domain. Alternative approaches to weighting also have potential such as considering social variables from questionnaire responses external to the existing data, or working with community members to determine weights, either through a supervised process with data experts or employing a participatory or deliberative method that takes account of local needs, policy priorities, and best practices.

Dashboard development: The project team produced a functional proof-of-concept Dashboard, which can be used to view levels of community well-being for district and unitary authorities throughout England (see figure 4). This solution is scalable to the UK (including Wales, Scotland and Northern Ireland) as boundaries act as an overlay that can permit differing layers of granularity.

The dashboard is integrated with the algorithm, which determines the score for a given locality, in this case through the lens of the Good Life Euston model. Values are calculated via an external script that in turn updates the dashboard database. This is a flexible system – designed to enable councils to upload data and generate rapid feedback based on the current community well-being model. Alternative models can be added to the dashboard as new algorithms, allowing for place-specific definitions of community well-being to generate the Index.

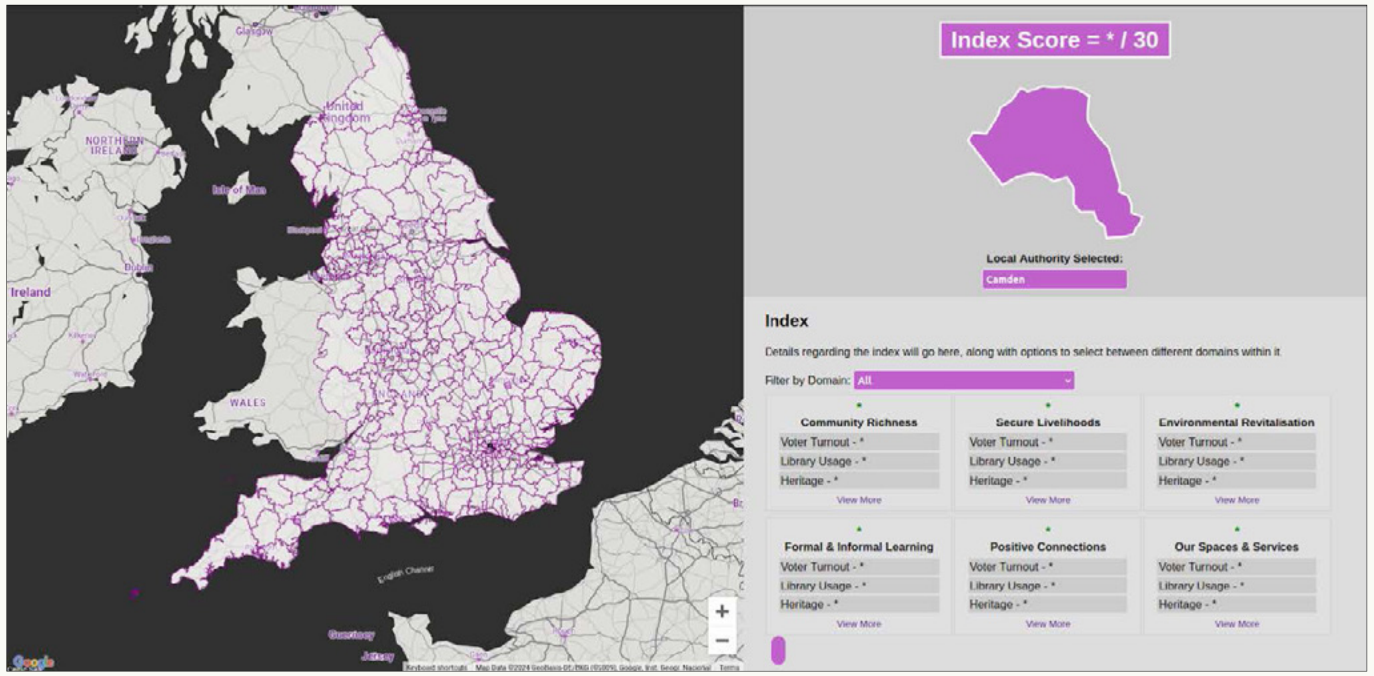
Data in its present form at aggregated local authority level provides insight into the broad aspects of well-being. However, the richness of this data is diluted by aggregation and a limiting factor in creating more granular and comprehensive metrics for understanding community well-being is the quality and depth of data accessible to researchers. Further release of Parish and Ward level information would greatly aid in granting further insight into community well-being.
When considering the intended application and usage of the Community Well-being Index Dashboard, and accompanying algorithm, the development adhered to and aimed to fulfil the following core concepts:

Usage of a dashboard should be accessible to non-technical people, focusing on visuals and transparency

Data used should be formatted using a template package that can be provided to councils, to allow ease of participation with the project

The dashboard is hosted securely and in a publicly accessible location online, to engage with the public and members of governmental bodies.

Figure 4: A Screenshot of the Community Well-being Index Dashboard



Stage 4

Evaluating the effectiveness of the prototype Community Well-being Index

Key points

1 *Evaluation of the Community Well-being Index shows that it better predicts productivity compared to self-report Subjective Well-being measures.*

2 *Aggregate measures of community well-being that consider all relevant domains are better than single domains at predicting productivity.*

3 *These findings suggest that administrative data and AI can provide policy-relevant insights into community well-being.*

The final stage of the project evaluated whether a Community Well-being Index based on administrative data and AI can generate new insights that are relevant to government and local authority policy making. To do this, we compared the effectiveness of the Community Well-being Index with conventional self-report Subjective Well-Being measures (anxiety, life satisfaction, feeling life is worthwhile, and happiness) in explaining productivity differentials across local authorities in England.

4.1 Why use productivity to validate the Community Well-Being Index?

Community well-being encompasses health, educational outcomes, social inclusion, community cohesion, social infrastructure, and economic stability. Each of these elements can significantly impact on labour and economic productivity by influencing the physical and mental health, skill levels, social interactions, and overall quality of life of the workforce in a locality. In this sense, understanding and enhancing community well-being can lead to a more resilient and efficient workforce.

We used statistical methods to empirically assess whether the new Community Well-being Index can offer new insights into productivity differentials in the UK at the local authority level. Labour productivity is a measure of how much economic output is generated per hour worked for the whole economy and for each industry or sector. The Office for National Statistics calculates labour productivity by dividing gross value added (GVA) by hours worked for that industry or sector.[24]

We chose to explore the relationship between community well-being and productivity for two reasons: first, because of the strategic importance placed by the government on productivity and local economic growth in tackling regional inequalities; and second, because of ongoing debates surrounding productivity challenges in the UK and other developed economies. This is not to suggest that productivity gains should be the outcome of increasing community well-being, however, some form of assessment of the Index's explanatory potential is needed to address the question of whether administrative data and AI can generate insights. In the rest of this section, we discuss the statistical analysis and what the findings say about the value of using administrative data and AI to measure community well-being.

4.2 Correlation analysis

We used correlation analysis to investigate the strength of relationships in and between the newly developed Community Well-being Index and commonly used Subjective Well-Being measures.

There are strong correlations among self-report Subjective Well-being measures, indicating their interconnected nature. The numbers in table 1 represent Pearson correlation coefficients, which range from -1 to 1, with asterisks indicating significant correlations at a 1% level ($p<0.01$). For example, anxiety has a significant negative correlation with life satisfaction (-0.529), feeling life is worthwhile (-0.447), and happiness (-0.552), indicating that higher levels of anxiety are associated with lower subjective well-being. Life satisfaction, feeling life is worthwhile, and happiness are all positively correlated with each other.

However, the self-report Subjective Well-being measures tend to show varying correlations with the Community Well-being Index and some of its domains. For example, the Community Well-being Index Formal and Informal Learning domain is positively associated with the Subjective Well-being Life Satisfaction measure (0.112). Yet none of the Subjective Well-being measures shows a statistically significant relationship with the Community Well-being Index Our Spaces and Services domain. Many domains of the Community Well-being Index even show strong negative correlations with Feeling Life is Worthwhile and Happiness. Such results suggest that these Community Well-being measures capture different information and highlight the complexity of community well-being and its measurement at the aggregate level.

There is a relatively weak correlation between the Community Well-being Index and commonly used Subjective Well-being measures (anxiety, life satisfaction, feeling life is worthwhile, and happiness). This highlights the potential of the Community Well-being Index to offer valuable, additional insights information for policy development aimed at enhancing community well-being and productivity (see table 1 for all correlations).

Table 1: Pairwise correlations – subjective well-being measures versus community well-being indices[25]

Variables	Anxiety	Life Satisfaction	Life Worthiness	Happiness	Community Richness, Cultures and Identities	Our Spaces and Services	Secured Livelihood	Environmental Revitalisation	Formal and Informal Learning	Positive Connections	Community Well-being Index
Anxiety	1.000										
Life Satisfaction	-0.529*	1.000									
Life Worthiness	-0.447*	0.813*	1.000								
Happiness	-0.552*	0.754*	0.704*	1.000							
Community Richness, Cultures and Identities	0.119*	-0.154*	-0.145*	-0.125*	1.000						
Our Spaces and Services	-0.057	0.038	0.036	0.042	-0.003	1.000					
Secured Livelihood	0.144*	-0.196*	-0.195*	-0.112*	0.107*	0.142*	1.000				
Environmental Revitalisation	0.234*	-0.232*	-0.184*	-0.152*	0.130*	0.322*	0.280*	1.000			
Formal and Informal Learning	-0.191*	0.112*	0.047	0.013	0.018	-0.087*	-0.218*	-0.246*	1.000		
Positive Connections	-0.008	-0.090*	-0.085*	-0.072*	0.159*	-0.138*	-0.149*	-0.187*	0.406*	1.000	
Community Well-being Index	0.000	-0.135*	-0.154*	-0.123*	0.419*	0.590*	0.297*	0.259*	0.547*	0.521*	1.000

* indicates statistical significance at the 1% level ($p < 0.01$).

24 Office for National Statistics (2024). Labour productivity quality and methodology information (QMI). https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/methodologies/labourproductivityqmi#methods-used-to-produce-the-labour-productivity-data

4.3 Regression analysis

We used two-way fixed effects (TWFE) models to examine how different well-being measures explain, and can be applied to understanding, regional productivity differences in the UK.

The model is specified as follows:

$$y_{it} = \beta_0 + \beta_1 index_{it} + X'_{it}\beta + \alpha_i + \delta_t + \epsilon_{it}$$

y_{it} represents productivity at local authority level in the UK, measured by constant Gross Value Added (GVA) per hour.

$index_{it}$ represents the different well-being measures.

Control variables X_{it} include common factors that may also affect productivity, such as unemployment, investment, other human capital measures, and crime measures.

The longitudinal nature of our data—the fact that it contains information from a sustained time period—is well-suited for the TWFE models, as it helps to control for a range of location specific and time-specific factors that are not easily observable in a cross-sectional data setting. For example, location-specific factors could include regional economic policies and local climate in different parts of the UK. Time-specific factors might include national economic cycles, or significant events such as the Brexit transition or the COVID-19 pandemic.

Results using self-report Subjective Well-being measures: The TWFE regression results indicate that among various self-report Subjective Well-being measures only anxiety shows a statistically significant positive relationship with productivity, with a coefficient of 0.028 (see table 2 for all coefficients). Other measures such as the composite Subjective Well-being Index, which includes feeling life is worthwhile, happiness, and life satisfaction do not exhibit significant relationships with productivity. A substantial proportion of the variation in productivity is explained by the fixed effects rather than the explanatory variables. The findings underscore the complexity of the relationship between well-being and productivity, suggesting that traditional self-reported measures may not fully capture the aspects of well-being that influence productivity.

The relationship between anxiety and productivity is complex and varies by context. While high levels of anxiety can impair productivity by inducing stress and reducing cognitive function, moderate levels may enhance focus and motivation, as suggested by the Yerkes- Dodson law[26]. However, Corbett (2015)[27] argues that empirical support for a positive link between stress and performance is questionable. Therefore, policymakers should approach such findings cautiously. Folkman and Moskowitz (2000)[28] highlighted the importance of recognizing individual differences in coping with anxiety, which, in turn, affects emotional resilience and overall well-being. People with effective stress management strategies might leverage anxiety to enhance their productivity, whereas those without such strategies might experience a decline in productivity. Further research and more granular data are required to better understand these relationships at the aggregate level.

25 The rows and columns contain both the domains of the community well-being index and self-reported subjective well-being measures.

26 Yerkes, R.M. and Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 18(5), pp.459-482. doi:10.1002/cne.920180503.

27 Corbett, M., 2015. From law to folklore: work stress and the Yerkes-Dodson Law. Journal of Managerial Psychology, 30(6), pp.741-752. doi:10.1108/JMP-03-2013-0085.

28 Folkman, S. and Moskowitz, J.T., 2000. Positive affect and the other side of coping. American Psychologist, 55(6), pp.647-654.

Table 2: Two-way fixed effects models using self-reporting measures to predict productivity

	(1)	(2)	(3)	(4)	(5)
	Productivity	Productivity	Productivity	Productivity	Productivity
Personal wellbeing index	-0.006				
	(0.014)				
Anxiety		.028**			
		(0.011)			
Life worthiness			-0.009		
			(0.05)		
Happiness				-0.005	
				(0.045)	
Life satisfaction					0.066
					(0.055)
Constant	3.448***	3.3***	3.35***	3.341***	3.198***
	(0.066)	(0.012)	(0.103)	(0.09)	(0.11)
Observations	2219	4200	4200	4200	4200
Within R^2	0.673	0.744	0.744	0.744	0.744
Year Dummy	YES	YES	YES	YES	YES
Local Authority Dummy	YES	YES	YES	YES	YES
Standard errors are in parentheses					
**** p<.01, ** p<.05, * p<.1*					

Results using other behavioural indices: In addition to the commonly used self-report Subjective Well-being measures there are socioeconomic variables frequently collected through survey questions by the Office for National Statistics (ONS) at the local level (see table 3). While these indicators are not explicitly designed to measure well-being, they can be seen as behavioural measures of well-being because they either directly or indirectly influence and reflect the physical and mental health, security, and overall quality of life of individuals in a community. In this sense, they can provide critical insights about community well-being. For instance, the level of reported personal crime in an area is a vital indicator of community safety and security. High crime rates lead to increased stress, anxiety, and fear among residents, negatively affecting their mental health and overall well-being[29]. Conversely, reduced crime rates enhance feelings of safety and stability, contributing to better mental health and a higher quality of life.

Regression analysis using a group of selected behavioural measures reveals that there is a statistically significant relationship with local productivity. Notably, higher levels of reported personal crime and higher levels of the population with cardiovascular conditions consistently show a statistically significant negative relationship with regional productivity across all models. While increased levels of job-related training, physical activity, and workplace safety are positively related to productivity. While anxiety remains positively correlated with productivity after controlling for a wide range of behavioural variables in model six (see table 3, model 6), its statistical significance level has declined from the 5 percent level to the 10 percent level.

29 See Lorenc et al. (2014) for comprehensive review. Lorenc, T., Petticrew, M., Whitehead, M., Neary, D., Clayton, S., Wright, K., Thomson, H., Cummins, S., Sowden, A., Renton, A., 2014. Crime, fear of crime and mental health: Synthesis of theory and systematic reviews of interventions and qualitative evidence. Public Health Res 2. https://doi.org/10.3310/phr02020

Table 3: Two-way fixed effects models using selected behavioural measures

	(1)	(2)	(3)	(4)	(5)	(6)
	Productivity	Productivity	Productivity	Productivity	Productivity	Productivity
Crime	-.161***	-.158***	-.159***	-.16***	-.156**	-.156**
	(0.061)	(0.061)	(0.061)	(0.061)	(0.061)	(0.061)
Cardiovascular disease		-.088***	-.087***	-.084***	-.082***	-.084***
		(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Job related training			.031**	.031**	.03**	.027**
			(0.013)	(0.013)	(0.013)	(0.013)
Physical activity				.044**	.044**	.041**
				(0.019)	(0.019)	(0.019)
Workplace safety					.065*	.063*
					(0.036)	(0.036)
Anxiety						.021*
						(0.011)
Constant	4.162***	4.558***	4.412***	4.198***	3.878***	3.901***
	(0.282)	(0.314)	(0.331)	(0.322)	(0.383)	(0.385)
Observations	2219	2219	2219	2219	2219	2204
Within R^2	0.676	0.679	0.681	0.682	0.683	0.683
Year Dummy	YES	YES	YES	YES	YES	YES
Local Authority Dummy	YES	YES	YES	YES	YES	YES
Standard errors are in parentheses						
**** p<.01, ** p<.05, * p<.1*						

Results using Community Well-being Index: Introducing the composite Community Well-being Index into our regression analysis, we observe several significant findings (see table 4). The Community Well-being Index remains positively correlated with productivity across all models, with a coefficient around 0.07, significant at the 1% level. This consistent positive relationship underscores the usefulness and potential of using a multi-dimensional measure of community well-being in understanding key policy questions, such as the productivity differentials across the UK.

The behavioural variables previously discussed are retained as control variables to assess the robustness of our results on the Community Well-being Index. Most of these variables remain statistically significant in the model.

Table 4: Two-way fixed effects models using the Community Well-being Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Productivity	Productivity	Productivity	Productivity	Productivity	Productivity	Productivity	Productivity	Productivity
Wellbeing Index	.072***	.073***	.074***	.071***	.07***	.071***	.072***	.068***	.06***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.019)	(0.02)	(0.02)	(0.02)	(0.019)
Crime	-.153***								
	(0.056)								
Cardiovascular disease		-.091***	-.089***	-.087***	-.085***	-.086***	-.085***	-.086***	-.089***
		(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.03)	(0.03)	(0.029)
Job related training			.031**	.031**	.03**	.027**	.03**	.029**	0.021
			(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Physical activity				.037*	.038**	.034*	.037**	.036*	.04**
				(0.019)	(0.019)	(0.019)	(0.019)	(0.018)	(0.019)
Workplace safety					.065*	.063*	.065*	.062*	.067*
					(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Anxiety						.021*			
						(0.011)			
Unemp records							-0.001		
							(0.013)		
Transport								.098*	
								(0.052)	
Investment									.024**
									(0.01)
Constant	3.871***	3.574***	3.423***	3.248***	2.949***	2.97***	2.953***	2.425***	2.905***
	(0.265)	(0.16)	(0.17)	(0.191)	(0.245)	(0.244)	(0.266)	(0.38)	(0.249)
Observations	2149	2149	2149	2149	2149	2134	2134	2149	2030
Within R^2	0.675	0.675	0.677	0.677	0.679	0.679	0.677	0.681	0.684
Year Dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES
Local Authority Dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES
Standard errors are in parentheses									
**** p<.01, ** p<.05, * p<.1*									

Multicollinearity is a common issue in linear regression, arising when explanatory variables are highly correlated. This makes it difficult to isolate their individual effects, leading to inflated standard errors, unstable coefficients, and reduced statistical significance. While the overall model fit may remain high, multicollinearity can distort inference and undermine the reliability of policy conclusions. Addressing it is crucial for producing robust and interpretable results.

In this report, multicollinearity is addressed by calculating the Variance Inflation Factor (VIF) for each independent variable, with only those variables having a VIF below 10 retained in the panel regression model. For instance, the reported level of personal crime was removed after the first regression model (see table 4). Overall, these results reaffirm the usefulness of administrative data in understanding productivity differentials at the local authority level. To improve productivity and achieve sustainable development within a locality, policymakers may want to focus on improving one or more domains of the Community Wellbeing Index where an area lags behind other localities. Such a targeted approach could contribute to the overall enhancement of both community well-being and productivity.

Interestingly, when examining individual domains of the Community Well-being Index, none shows a statistically significant relationship with productivity. This seems to suggest that only the combined effect of various well-being factors, rather than individual components, can drive productivity. Such findings create space to explore the synergistic and interdependent effects of these well-being domains, where improvements in one area might only yield significant productivity gains when complemented by advances in others. This finding highlights the importance of further research into understanding the interactions between each domain within the Community Well-being Index and their synergistic effects.

4.4 Limitations of this analysis

This research is still at an early stage and the findings are still subject to several limitations.

First, the use of behavioural administrative data is not immune to measurement errors or biases. The complex and multifaceted nature of community well-being suggests that no single measure can easily capture all aspects of well-being.

Second, our TWFE regression models are static and cannot infer causal relationships, so more sophisticated panel regression models should be adopted for future research.

Third, our results may also be affected by missing data in some domains of the Community Well-being Index, for example, in the Our Spaces and Services domain.

Community well-being is complex and context specific. Further research is needed to investigate the multiple dimensions of community well-being and to better understand the dynamics between community well-being and other socio-economic variables.

Conclusions and future directions

This project is the beginning of an effort to integrate administrative data and AI as a measurement tool to provide insights into community well-being.

Preliminary analysis shows it is feasible to use administrative data and AI to generate policy-relevant insight into community well-being; evaluation of the Community Well-being Index shows that it better predicts productivity compared to self-report Subjective Well-being measures, and multidimensional well-being measure is better than single domains at predicting productivity.

This work identifies ethical and practical challenges in the use of administrative data and AI for local authorities seeking to measure community well-being and proposes recommendations addressing data privacy and ownership, public engagement with AI, and investments in local government data infrastructure (see Appendix for details).

In this concluding section we summarize four areas where future research is needed to progress the use of AI tools in general and the Community Well-being Index as a specific case.

5.1 Debiasing and AI approaches

Regardless of the consensus on the ethical use of synthetic data, ethical concerns remain with AI approaches if debiasing—removing inherent systemic and structural biased inherent in the dataset—existing data sets is not possible. Administrative datasets, the material that AI learns from, inherently reflect existing societal biases. AI learning from these data sets will learn these biases as well, threatening to enshrine them in decision-making. This presents major challenges to the future development of AI tools - not only must decisionmakers be aware of these biases, but AI must be able to identify and quantify biases, removing their influence from training and subsequent performance of the AI. Debiasing tools represent the next advance in data science and will be critical for the future development of tools like the Community Well-being Index, both to refine the approach and provide maximize its utility as an aid to decision making.

Beyond debiasing, using AI to speed up the preparation of datasets for analysis by the algorithm would facilitate wider adoption of the approach and the algorithm. Currently, datasets come in many formats, and harmonizing these for AI learning is time-consuming. But as this work continues, the AI can learn the formats and extract relevant data, saving time and effort. Further computer science research focused on such challenges would drive the project forward.

5.2 Shock and resilience prediction

The research approach we have developed in this project requires that the algorithm is validated once constructed to ensure that it indeed holds predictive power. The predictive nature of the algorithm makes it well-suited to estimate the resilience of communities in the face of potential economic shocks. Predicting how communities will respond in the face of economic uncertainty offers valuable insight to policymakers. For instance, such information about community resilience that could be acquired independently of self-report surveys during a time of disaster would have been useful in shaping responses during the pandemic that attempted to balance public health concerns with economic concerns.

Not only can the algorithm predict community resilience in the face of economic shocks, but it can also suggest which domains and sub-domains drive community well-being in a given locality. As such, it provides information policymakers can use to target investment opportunities and other resources. Drawing such benefits from the community well-being index requires further research, more test cases and local governments, and further advances in data science such as black-box or generative AI.

5.3 Continuous monitoring of public perceptions of AI

The psychological research described earlier in this report was collected before the launch of ChatGPT, which changed the way the public thought about AI, making its terminology more prevalent in public discourse and increasing the number of people who intentionally interact with AI. It is not far-fetched to consider that other AI advancements may also change public perception. Therefore, it is important to monitor public perception of AI as it moves along a continuum from a technological tool to a vital interaction partner to facilitate human behaviour. As such, future research should collect public perceptions of AI at regular intervals so that the most up-to-date perceptions can inform ethical debates around data and AI.

The final future direction of this research tracks public perception of AI longitudinally, across multiple generations. This approach relies on self-report surveys for people to report their perception of AI along the dimensions already outlined in the literature and behavioural measures of AI use. Moreover, participants would be able to report perceptions of novel AI technologies and applications, allowing a nuanced understanding of how the technology develops and is deployed.

Appendices

Appendix A: Guidance on combating data challenges

Taking advantage of AI technology to provide insight that can guide policy requires overcoming many data challenges. We suggest the following ways of doing so:

1. Data Library Platform: There is a consistent problem of siloed working in local authorities due to the huge extent of the business. This makes data transfers difficult even within one organization. There is a tested solution to improve data management practices called the "OneView platform" which aggregates data from multiple sources, providing a comprehensive view of individuals or households. This system aids frontline officers in better supporting residents by offering a holistic view of an individual and their family. However, aggregation platforms like this are likely to reinforce public data security and privacy concerns.

2. Standardization of Data: The collected data needs to be stored in a standardized format to develop a holistic approach across council services. For instance, the definition of a 'child' varies across services, creating gaps in data and service delivery.

3. Fostering Ethical Practices: Ethics is broader than legal obligations. Some information provided by individuals is legally safe to be used for some policymaking purposes, but consent around use is ethically tricky. Thus, fostering ethical practices in data usage, which goes beyond legal compliance to consider broader ethical implications and public perception is essential.

4. Training and Education: Educating elected officials on these issues to help prioritize data-related challenges and solutions. The role of data scientists is evolving, requiring them to engage more with the social aspects of data usage, including consent and public education.

Appendix B: Guidance on involving community stakeholders

There is a disparity between who currently gets to define community well-being and who should ideally be defining it. Several recommendations were developed around the definition of well-being and community stakeholder involvement. Here are the key recommendations:

1. Diverse Stakeholder Representation: Community well-being should be defined in a process that actively involves the community members whose well-being is being measured. The involvement of citizen scientists along with experts, academicians and policymakers is essential. This should go beyond consultation to include active participation in defining what constitutes well-being for community members. A broad portfolio of stakeholders should be engaged in defining community well-being. This includes residents, local businesses, government agencies, NGOs, and other community organizations that contribute to the community's social fabric.

2. Sensitivity to Local Contexts: Recognize that community well-being is context dependent. Definitions and measures of community well-being should be adaptable to reflect the specific needs, cultures, and circumstances of different communities, rather than imposing a one-size-fits-all model.

3. Multi-dimensional Measures: Adopt multi-dimensional measures that capture various aspects of community well-being beyond economic indicators. This includes social, environmental, and cultural dimensions that contribute to a holistic view of what it means to thrive.

4. Feedback and Adaptation: Implement mechanisms for continuous feedback from community members on community well-being initiatives. This allows for ongoing adaptation and refinement of community well-being measures to better serve the community's needs.

5. Focus on Marginalized Groups: Special attention should be given to including marginalized and often voiceless groups in the community. Ensuring their participation in the definition process is crucial for a comprehensive and inclusive approach to community well-being.

Appendix C: Guidance on data privacy and permissions challenges

Addressing data privacy and permissions requires a multifaceted approach that involves improving public understanding, ensuring transparency and ethical data use, and actively engaging with the public to rebuild trust in data-sharing practices. We suggest the following recommendations:

1. Enhanced Public Information Campaigns: As outlined in one of the sections above, we support ADR UK's recommendation for public information campaigns to spread awareness and educate the public about their rights around data privacy and permissions. Building trust in data use requires a proactive approach to engaging the public about how data is used and ensuring inclusivity and addressing the impact of digital inequality on data accessibility.

2. Campaigns to Communicate the Purpose of AI use: Psychological research indicates that if the installation of CCTV is justified with specific reasons, such as addressing a spate of robberies in a neighbourhood, public support for it increases substantially. Conversely, the absence of a communicated rationale for surveillance leads to a lack of public approval. Thus, it is important to transparently convey the purpose of using AI as public reactions are significantly rooted in the stated purpose.

3. Legal and Policy Reforms: There is inconsistency in data sharing and a lack of data infrastructure. It is therefore necessary for legal and policy reforms to facilitate easier data sharing and use while respecting privacy and permissions.

4. Building Trust and Transparency: Certain departments and institutions face additional challenges due to their reputation, impacting their willingness to share data. This highlights the broader issue of trust in public institutions and the need for those institutions to demonstrate trustworthiness in their data-handling practices.

5. Inclusion in Education: Integrate data privacy and protection concepts into the national curriculum, aiming to equip younger generations with a better understanding of data privacy from an early age, potentially changing the narrative around data use and privacy.

6. Engage with Affected Demographics: When defining what constitutes 'public good', organizations should consult with the demographics likely to be affected by data-sharing practices, ensuring their perspectives and concerns are considered.

Appendix D: Guidance on combating technical shortcomings of synthetic data

Synthetic data holds the potential to overcome many of the ethics and privacy concerns involved in using real-world administrative datasets for insight into community well-being. However, technical shortcomings of synthetic data undermine this potential. Following are recommendations for combating these challenges:

1. Need for Ethical Standards and Governance: There is optimism about the capacity of synthetic data to change the landscape of data-driven research. However, this optimism is tempered by the necessity for rigorous ethical standards and robust data governance frameworks to mitigate risks associated with synthetic data use.

2. Critical Need for High-Fidelity Synthetic Datasets: It is important to create high-fidelity synthetic datasets that accurately mirror real-world complexities, driven by data quality and freshness. High-fidelity synthetic data ensures that even unique or rare occurrences within datasets are accurately represented, facilitating comprehensive and nuanced insights essential for informed decision-making and policy formulation.

3. Importance of Data Quality Standards: Before utilizing any real or synthetic data, establishing data quality standards is crucial. Insurmountable data issues can undermine a project's legitimacy, despite the enhanced public credibility through terms like "algorithms" and "AI".[30]

4. Charting a Path Forward for Public Governance: Collaboration among stakeholders, including data scientists, policymakers, and community leaders, is crucial to responsibly harness the benefits of synthetic data while addressing privacy and ethical standards.

5. Reliance of Qualitative Data: Although there may be less trust in synthetic qualitative data, we also recommend converting this data to quantitative data and then producing synthetic data.

6. Data Testing is Important: Creating synthetic data based on small data and then using the remaining data to test the generated data was one of the effective solutions to avoid errors in data generation. In this way, different levels are created to filter out biases and get opinions from the original data about the new data generated.

7. Judgement Calls: Some datasets are updated less frequently or updated with lags compromising the freshness of the data. In some cases, the data is still useful for decision-making, while in cases like school data, the demographics change each year, so it becomes out of date quickly. Thus, it is important to make a judgment call on the original data and decide whether to use it to generate synthetic data or not.

8. Public Perception of Research: Leveraging the psychological tendency of the public to believe in scientific things is effective while communicating about the research work and findings. Science heuristics plays a role in people believing results. This heuristic must not be exploited without ensuring the underlying data quality.

9. Use of AI in Synthetic Data: The synthetic data created by generative AI algorithms can be a fairer and richer version of the original data. However, it is mainly done by finding patterns in the original data. Thus, avoiding complete reliance on AI is essential. It is crucial to ensure that decision-making systems are not fully automated and that there is human intervention in AI decisions.

10. Limiting Synthetic Data Generation: Synthetic data produced for one purpose but used for another purpose might generate issues. To avoid misuse, there is a significant need to limit the generation of synthetic data. The synthetic data generation process should be made available, rather than the data itself, to prevent misuse.

11. Partnerships with Data Controllers: Encouraging partnerships with data controllers and councils to spread awareness about the benefits of data sharing is essential.

12. Exploring Legal and Ethical Data Use: It is crucial to continue to explore legal and ethical ways to use real data to ensure decisions are not always based on potentially less relevant and low-fidelity synthetic data. It is important to foster a culture of ethical data use, such that synthetic data can be a tool for public good under principles of fairness, transparency, and accountability.

30 Modhvadia, R. (2023). How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain. Ada Lovelace Institute and The Turing Institute. https://www.adalovelaceinstitute.org/report/public-attitudes-ai/

13. Promoting Intra-Organization Data Sharing: Facilitating data sharing within organizations under legal constraints could reduce the unnecessary use of synthetic data. Organizations having full transparency within their departments from the generation of synthetic data, sharing of data and storage of the data will define the evolving landscape of synthetic data.

14. Reforming Data Sharing and Privacy Laws: Updating data sharing and privacy laws to fit the evolving technology landscape.

15. Building Repositories and Highlighting Key Use Cases: Synthetic data and its application in policy decision-making is recent and thus very new. This creates a need for maintaining repositories of how synthetic datasets were created, what was the main purpose of creating them, and what were the limitations of the datasets. Outlining the limitations of indices built using synthetic data and highlighting relevant and irrelevant use cases to minimize negative impacts is essential.

Appendix E: Data sources used in the Community Well-being Index Dashboard

Free School Meals	https://explore-education-statistics.service.gov.uk/data-tables/permalink/50ebbb55-4157-4635-29bd-08dc6efa3562
Child Poverty	https://www.gov.uk/government/collections/children-in-low-income-families-local-area-statistics#latestrelease
Hate Crimes	https://data.london.gov.uk/dataset/mps-monthly-crime-dahboard-data
Homelessness	https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness#full-publicationupdatehistory
Parks	https://www.ons.gov.uk/economy/environmentalaccounts/datasets/accesstogardensandpublicgreenspaceingreatbritain
Recycling	https://www.gov.uk/government/statistics/local-authority-collected-waste-management-annual-results
Life Expectancy	https://fingertips.phe.org.uk/search/life%20expectancy
Fly-Tipping	https://www.gov.uk/government/statistics/fly-tipping-in-england#full-publication-update-history
Traffic Accidents	https://www.gov.uk/government/statistical-data-sets/reported-road-accidents-vehicles-and-casualtiestables-for-great-britain#historic-trends-ras01
Noise Complaints	https://fingertips.phe.org.uk/search/noise%20complaints
Absenteeism	https://fingertips.phe.org.uk/search/absenteeism
Child Obesity	https://fingertips.phe.org.uk/search/child%20obesity
Obesity	https://fingertips.phe.org.uk/search/obesity
Youth Training	https://explore-education-statistics.service.gov.uk/data-tables/permalink/b7cb5ba0-2a2c-4e30-29c2-08dc6efa3562

Teen Pregnancy	https://fingertips.phe.org.uk/profile/sexualhealth/data#page/9/gid/8000036/pat/15/par/E92000001/ati/402/are/E09000007/iid/20401/age/173/sex/2/cat/-1/ctp/-1/yrr/1/cid/4/tbm/1/page-options/car-do-0
Annual Income	https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/smallareaincomeestimatesformiddlelayersuperoutputareasenglandandwales
Income After Housing	https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/smallareaincomeestimatesformiddlelayersuperoutputareasenglandandwales
Transport	https://www.gov.uk/government/collections/uk-local-authority-and-regional-greenhouse-gas-emissionsnational-statistics
KS Attainment	https://explore-education-statistics.service.gov.uk/data-tables/permalink/de5c7b7e-44ba-42f2-1903-08dc6f2a7f9d
Unemployment Records	https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/modelledunemploymentforlocalandunitaryauthoritiesm01/current
Unemployment Benefits	https://fingertips.phe.org.uk/search/unemployment%20benefits
Social Housing	https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/housinginenglandandwales/2021comparedwith2011#tenure-by-accommodation-type
Home Ownership	https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/housinginenglandandwales/2021comparedwith2011#tenure-by-accommodation-type
Voter Turnout	https://www.electoralcommission.org.uk/research-reports-and-data/electoral-registrationresearch/electoral-registration-great-britain-2022#background

Institute for Global Prosperity

Research at the UCL Institute for Global Prosperity aims to generate new insights about sustainable and inclusive prosperity and provide new models for developing and interpreting evidence.

Underlying our research is a rethinking of what we mean by prosperity. Prosperity must mean enabling people to flourish in ways beyond financial growth – and doing so equitably and sustainably, for humankind and the planet. We work with businesses, non-governmental organisations and citizens to produce interdisciplinary methodologies and problem-focused research.

For more information about our wide range of current projects and our innovative Master's and PhD programmes please see: https://www.ucl.ac.uk/bartlett/igp/

Contact

www.prosperity-global.org
www.ucl.ac.uk/bartlett/igp/

igp@ucl.ac.uk
@glo_pro