Mobilising UK Data and AI for All with a National Grid of Civic Learning Systems

John Ainsworth, Claudio Angione, Kate Ardern, Katie Atkinson, Samuel Ball, Ben Barr, Ben Bridgewater, Iain E Buchan (editor), Camila Caiado, Helen Cole, Hannah Davies, Nigel Davies, Hedley C A Emsley, Alejandro F Frangi, Marta Garcìa-Fiñana, Hilary Garrett, Carole Goble, Mike Harding, Iain Hennessey, Matt Hennessey, Warren Heppolette, Andrew Hill, Thomas House, Jim Hughes, Owen Johnson, Dan Joyce, Axel Kaehne, Louise C Kenny, Jo Knight, Chris Kypridemos, Tom Lawton, Gary Leeming, Jims Marchang, Anthony Marson, Glen Martin, Simon Maskell, Mark Mon-Williams, Georgina Moulton, Annemarie Naylor, Farnaz Nickpour, P-J Noble, Boguslaw Obara, Annalisa Occhipinti, Kay O'Halloran, Séamus O'Neill, Reema Patel, Munir Pirmohamed, Roberta Piroddi, Rowan Pritchard Jones, Alan Radford, Emily Rempel, Sarah Rodgers, Malcolm G Semple, Sally Sheard, Alex Singleton, Matthew Sperrin, Tjeerd van Staa, Rob Tabb, Lauren E Walker, Paul Watson, Phil Waywell, Becca Wilson, John Wright, Xingna Zhang

Northern Health Science Alliance (NHSA) and partners, 24th January 2025

Foreword

This is a response from the Northern Health Science Alliance (NHSA) and partners to the Science, Innovation and Technology Committee inquiry into how research and innovation can boost regional economic growth and help to deliver the Government's Missions.[1,2]

We are an alliance of universities and NHS organisations plus wider civic partners from the North of England with experience in mobilising data, analytics and artificial intelligence (AI) to protect and improve health, optimise public services, and grow economies with industry.

Our response to the inquiry considers the interdependence of health and wealth and shows how a network of devolved regions, each with a critical mass of efficient data uses, with public trust, could form a national grid of innovation that advances the UK's global economic interests while reducing regional inequalities.

Compared to the South, people in the North of England die earlier, spend less of their lives in good health, and have just a third of the median household wealth. This results in a productivity gap that costs the UK around £13.2bn annually.[3] At the same time, the North has strengths in the life sciences and digital sectors, with considerable headroom to grow, fed by a vibrant talent pipeline, leading data assets, and foreign direct investment.[4] The North returns more outputs than the South for each pound invested in research. Yet, there are single buildings in the South with more infrastructure funding than the total for the North, e.g. £71M for the Francis Crick Institute vs £49M for the Northern Combined Authorities.[5]

The UK's AI economy needs data and public participation as well as funding to thrive. Here the North has mobilised data into science, public service innovation and technology advancement with public trust for decades. At the same time, national attempts to do the same have been thwarted by distrust. In this paper, we show how to build a national grid of civic learning systems - coupling public service optimisation with efficient advances in science and sustainable economic growth.

Summary

This paper sets out an opportunity for the UK to develop a world-leading infrastructure for data-driven public service innovation that feeds, and is fed by, science and engineering at national and civic scales, sustainably and with public trust.

The disruptive model is a national grid of civic learning systems where public services continuously improve their value and productivity with artificial intelligence (AI) assistance at population, provider and person levels, as shown in Figure 1.[6] In this civic AI model, data is foundational – collected, analysed, and acted upon in learning loops to continuously improve service delivery and outcomes.

At population level, the successful design and deployment of civic AI* would help target resources to maximise the value of services as a system. The inputs of one service (e.g. NHS support of the mental health of families with young children to avoid removal of a child when they are struggling) could assist the outputs of another (e.g. children's social care from local authorities facing extreme pressures from the rising costs of looked-after children[7]).

At provider or organisation level, civic AI would enable continuous quality improvement *across* agencies. For example, AI-gathered feedback from service-users, where a coordinated approach is needed across NHS, social care, and charity/voluntary sector agencies to deliver best outcomes and value for money.

At person level, civic AI would act as a concierge or ally, easing access to public services and harnessing data from the person's daily life, with their consent, to tailor experiences – predicting needs for early intervention to improve outcomes.

At data level, civic AI would orchestrate the construction of practical intelligence across multi-dimensional data landscapes, revealing hidden patterns and strategic opportunities that currently elude human analysis. It would empower decision-makers to intervene with unprecedented precision and impact, while preserving residents' privacy.[8]

The learning system flywheel of continuous improvement seeks to narrow the gap between data and action – in synergy across person, provider and population levels. To be practical, this requires AI support. The system needs to identify patterns, predict needs, and optimise service planning and delivery. Simultaneously, it must maintain data security and confidentiality, and nurture public trust in progressively more data and AI uses.

Each civic learning system would borrow strength from a national 'library' of data, metadata and algorithms, and its innovations would further develop the UK asset.[2] A network of these learning systems would enrich existing national data research. The UK would benefit not only from better data for scientific discovery but also from more efficient embedding of research and innovation into public services for agile, low-cost, well-governed trials.

This networked approach would transform scattered local assets into a powerful national engine for regional innovation and foreign direct investment. These investments could

* AI and related technology with functionality co-produced by residents in response to a place-based, regional challenge and developed and tested with regional data assets and governance.

improve data quality and curation, incubate new data flows, and accelerate technology evaluation and adoption in a trustworthy way – improving public services and feeding local / national economic growth equitably. A virtuous circle.

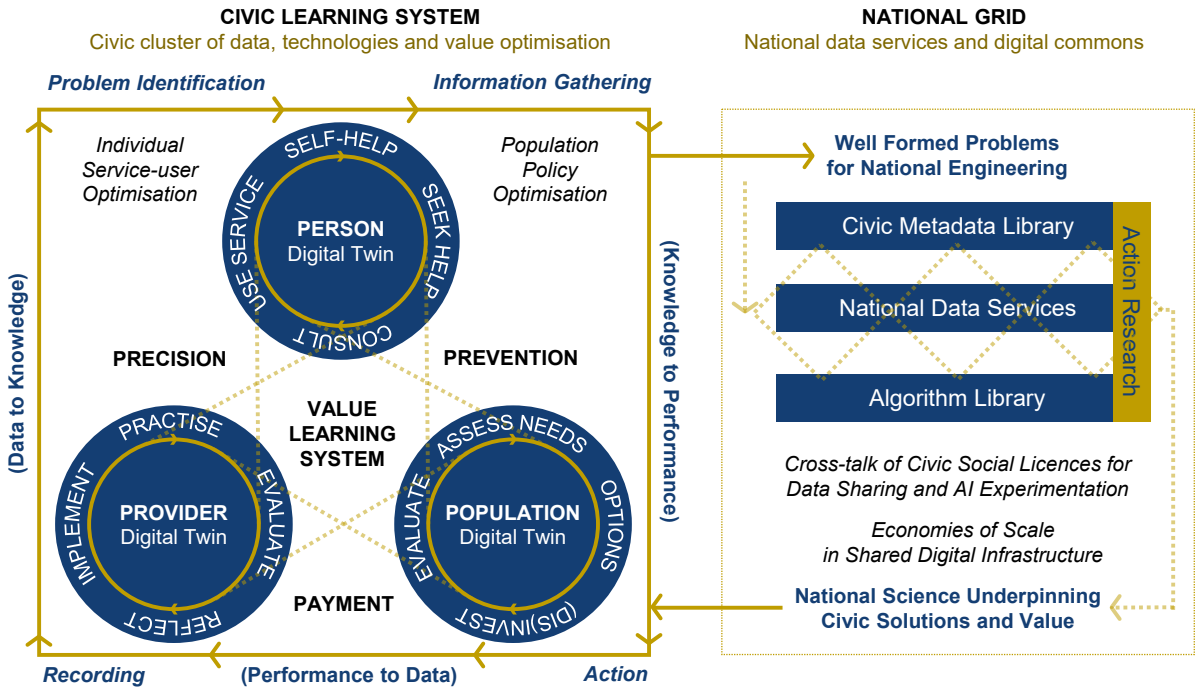

Figure 1: National grid of federated civic learning systems that synergises public service innovation across population, provider and person level needs, experiences and impacts – producing better data and digital workflows for research and innovation across the regions, driving national productivity and growth.

This approach can help realise the Government's AI Opportunities Action Plan[2] by unlocking data assets with public trust and supporting data-driven innovation with regional reciprocity. To do so would also build the civic infrastructure needed to realise the three big shifts the Government seeks from the NHS 10-year plan, from hospital to community, analogue to digital, and sickness to prevention.[9]

Background

Achieving the Government's five missions involves understanding complex, place-based systems. For example, combating antimicrobial resistance requires understanding of how to engineer antimicrobial stewardship systems that learn from the interplay of pathogen biology, changing environments, agriculture, medical testing and prescribing, organisation-level infection prevention and control, and patient behaviours around medication.[10] Another example is the sustainability of health and social care systems at a time when more people are living longer with more than one long-term condition. There is an escalating burden from combined mental and physical illness, which is worse in deprived areas already facing grave inequalities in health.[11] England's Integrated Care Systems (NHS-led civic health and care systems) are mandated to proactively improve population health but struggle to do so while reacting to the acute needs of sick patients. Policymakers see preventive civic partnerships, data, and AI as vital to NHS sustainability – as emphasised in Lord Darzi's recent report.[9]

Complexity cannot be ignored at the civic level. For example, the 8% of households with children in Cheshire and Merseyside (2.7m population) who face the most complex lives – those challenged by multiple types of adversity – consume 34% of the total NHS and social care resources spent on families with children.[12] This pattern is likely seen throughout the UK, and similarly across the life-course - where people are living longer with multiple conditions, discoordinated services and escalating use of healthcare resources.[13] When local communities debate different scenarios of data sharing for service improvement, scientific advancement or economic growth they usually demand more use of their data. This public expectation has been found consistently for more than a decade in the UK, from the #DataSavesLives movement, through local citizens juries for the National Data Guardian, Understanding Patient Data, and the NHSA-led Connected Health Cities initiative, to the participatory design work of the Civic Data Cooperative.[14-17] This social license can diminish in national data sharing debates when the place-based civic context of data-driven problem solving is lost. For example, the media sentiments around Care.Data, the GP Data for Planning and Research and Federated Data Platform,[18-20] were negative at the same time as strong public support was seen for use of the same data set in civic contexts, with tangible benefits to local communities.[21] Deliberative activities involving data sharing with residents from place-based communities have described a 'diameter of trust' – this must be large enough to achieve economies of scale in handling the data and achieving sufficient statistical power, yet small enough to earn public trust in using the data for good. These participatory activities with residents, researchers and innovators can identify new opportunities for local service improvement, local advancement of science, and local economic growth – building and sustaining opportunities for those living in the region. Residents gain greater access to information driven by their data, which they can use to lobby for, and enact, system change. In the civic setting, the leaders of key organisations taking decisions on system-wide data are also jointly accountable, which can further improve public trust that data insights will be acted upon well.

In the mid-2000s, through its e-Science initiative, the UK was at the forefront of engineering data systems for large-scale research.[22] This gave rise to principles such as the FAIR (findable, accessible, interoperable and reusable) framework for data sharing and architectures such as Research Objects encapsulating analytics for FAIR use between different e-Labs or Trustworthy/Trusted Research Environments (TREs) – with scope to check compliance (e.g. 5-Safes, SATRE).[23-29] This engineering approach is widely used to

support scientific collaboration and is compatible with International Data Spaces.[30] Similar engineering is needed across civic systems to help them to learn from data, supported by science and technology. The architectures and artefacts arising from e-Science have evolved in research domain-specific branches and in cross-cutting initiatives such as Data and Analytics Research Environments UK (DARE UK), with Administrative Data Research UK (ADR UK) and Health Data Research UK (HDR UK).[31-33]

The United Kingdom has established a robust foundation for data governance and sharing through strategic national initiatives. The Digital Economy Act 2017 (DEA) has created a comprehensive framework for federated data access, enabling secure and controlled data sharing across institutions.[34] DEA-accredited TREs, notably those operated by the Office for National Statistics (ONS), provide essential capabilities for sensitive data handling.[35] Infrastructure investments, such as those of the Science and Technology Facilities Council (STFC),[36] could enhance such capabilities, as outlined in the AI Opportunities Action Plan.[2] Administrative Data Research UK (ADR UK) additionally represents a significant advancement in this ecosystem, serving as a transformative partnership between government departments, research institutions, and national statistics authorities.[32]

Civic learning systems can maximise the value of national data and digital infrastructure with local contextual knowledge and actions. For instance, the national attempts to use data to optimise COVID-19 vaccination and testing were substantially improved by local initiatives: Liverpool used local community insights and neighbourhood profiles, including digital poverty and other factors to augment standard NHS intelligence in deploying a vaccine bus, leading to a 25% (95% CI 21% to 28%) increase in uptake.[37] This followed the world's first pilot of voluntary mass testing for people without symptoms of COVID-19 involving a quarter of Liverpool's population in a month - shaping the national COVID-19 Community Testing policy with civic engagement, multi-agency coordination and rapid, contextual evaluation.[38-42] Similarly, the use of local data and civic contextual knowledge improved the modelling of hospital burden from COVID-19.[43] If a 'national grid' of civic insights and actions had been used to support the recovery of public services and economies from the COVID-19 pandemic the profound inequalities and Long-COVID consequences reported by the NHSA[44] and Health Equity North[45] may have been lessened.

Civic systems can organise data, researchers and industry engagement in useful ways that are difficult to achieve nationally. For example, in action-oriented research (e.g. for service evaluation or planning), the service stakeholder owns the problem and asks the question. This complements academic research where the academic stakeholder owns the problem and asks the question. Governing these complementary data processing purposes together can maximise synergies. For example, the network of 11 regional NHS-led Secure Data Environments (SDEs) is part of the NHS Data for Research and Development Programme and relies upon NHS regional shared care records used for care, planning and research.[46,47] This exemplifies how different data processing purposes and governance tracks can share infrastructure for depth of curation, assurance of governance, consistent involvement of data rights-holders and economies of scale. Many NHS SDEs now seek linkage to non-NHS data for understanding wider/civic determinants of health and interactions with non-healthcare services that need to work in concert with healthcare. For example, social care data from local authorities, school pupil data from the Department for Education, benefits claims data from the Department for Work and Pensions, social prescribing data from charity and voluntary sector organisations, and housing and environmental data from various agencies.[48] The SDEs covering the North of England have a rich variety of priorities and

implementations that build on the Connected Health Cities initiative (2016-2020).[16] Connected Bradford prioritised linkage of broad civic data around the Born in Bradford birth cohort study in tandem with public service development uses of the same data – a civic approach now enriching the wider Yorkshire and Humber SDE. Lancashire and South Cumbria have an NHS-focused SDE driver programme where others such as Greater Manchester use SDE funding to support further development of a mature civic linked data infrastructure - both sit within the NW SDE programme alongside the Cheshire and Merseyside Data into Action SDE. All our SDEs and integrated care systems across the North of England are seeking to make their services more preventive at lower cost by integrating charity/voluntary sector organisations flexibly. A national grid of civic learning systems could support collaborative development and spread of tools (including social prescribing) for coordinating charity/voluntary inputs, as reflected in Greater Manchester's Creative Health Strategy.[49] Aligning industry partnership with population health needs and civic data/digital assets is key, as exemplified by the Eli Lilly obesity trial SURMOUNT-REAL in Greater Manchester, which comes with significant foreign direct investment.

In some regions, politicians have called for civic integrated data for the dual purpose of improving public services and growing local economies with data-fluid innovation. For example, the Liverpool City Region Civic Data Cooperative (CDC) – with a Mayoral mandate – nurtures the social license for progressive data and AI uses across NHS, local authority, emergency services, academic and other data controllers in the Cheshire and Merseyside Data into Action programme.[17] The CDC is hosted by the University of Liverpool's Civic Health Innovation Labs (CHIL). CHIL embeds an NHS secure data analytic facility, supports SDE technical development and delivers a data action-research programme for civic partners.[50] Economic and Social Research Council (ESRC) investments support this work through the national Consumer Data Research Centre (CDRC) / Geographic Data Service (now part of UK Research and Innovation (UKRI) SmartData Research UK).[51,52] Following successes of this civic approach to data-driven innovation, for example in COVID-19 responses, the Liverpool City Region Combined Authority is planning an Office for Public Service Innovation to mobilise data in tackling issues faced by residents and services. Most parts of England and the devolved nations have (or are developing) similar data action research partnerships.

In this paper, we consider the core technologies and methods needed to enable civic clusters of data and innovation to interoperate as a national grid and become a world-leading asset for science, engineering and civic advancement.

Aim

We propose regional incentives and national infrastructure developments to build a national grid of civic learning systems – the Grid – harnessing multi-source data for public service innovation, science, AI engineering and inclusive economic growth.

Objectives

The following objectives establish the critical national structural factors required to enable a unified network of civic data and learning systems across the UK. These components will serve as the foundational architecture – analogous to power lines and pylons connecting civic data-action 'power stations' on a national grid. This infrastructure will facilitate secure data sharing, standardised practices, and collaborative innovation while ensuring local autonomy and flexibility to address regional needs within a diameter of trust (public trust and practical joint accountability and mutual aid between organisations). By creating these essential connecting elements, we can transform currently isolated data initiatives into an interconnected ecosystem that drives public service improvement, scientific advancement, and economic growth through data-driven innovation.

Social License for Data Uses

Public perceptions, expectations, and understanding of data uses will be key to realising the benefits of linked data and AI in the civic context. In the field of health data sharing and access for uses beyond direct care, extensive research has examined public perceptions highlighting three key findings:

1. Publics in all cases prefer their data to be used for **positive change** both for themselves and others.
2. Publics expect their data to **not** be used **solely** for commercial **profit-making**.
3. Publics require **transparent** and **accessible** information on **how** their **data** is **used** alongside understandable and flexible **opt-out** mechanisms.

Findings around public preferences for health data uses are strongly connected to the perceived trustworthiness of the NHS. Similar views arise in public participation activities over both NHS and non-NHS data uses by civic partners to advance scientific knowledge, improve public services and grow local economies.[21]

The Grid should build on these findings to develop a social license for its intended data uses. This social license must represent the expectations, hopes, and worries of diverse public groups for the use and application of data and AI. This means developing a participatory process for the Grid that is trustworthy, transparent, and iterative. Liverpool, has drawn upon the Ada Lovelace framework for Participatory Data Stewardship to plan and map participatory processes.[53] We highlight several potential activities for the Grid and allied policies (e.g. the proposed National Data Library):

- Ensuring active and transparent **communication** that focuses data stewardship conversations on data literacy and intended policy applications.

- Summarising and identifying **gaps** in **existing** data and AI **consultation** exercises held across the **UK** considering the extensive existing knowledge in this space.
- Identifying and promoting public involvement and collaboration in every new data and AI project making **public participation** a **standard** expectation.
- Developing and testing **participatory** and **innovative models** of empowerment to **shift decision-making to publics** in areas of controversy and where there is potential for change.
- Building and showcasing a diverse living portfolio of placed-based people-centred **success stories** where the use of data and AI has benefited communities in ways that matter to them – ensuring **inclusivity**, **equity** and **representativeness** of the successes within civic learning systems while avoiding exaggeration. And complementing these, **failure stories** and the **learning** to underpin future successes.

The Grid should signpost data controllers/custodians and innovators to best practices in participatory methods for progressing data uses with our publics. The Grid's metadata might usefully define links between data sources, data uses and public participation findings.

Framework for Civic Participation in Artificial Intelligence (AI)

There is a need for civic systems to manage the benefits and risks of AIs. To explore this the Liverpool City Region Civic Data Cooperative, Digital Good Network, and Elgon Social Research, brought together AI practitioners, members of the public, policymakers, researchers, innovators, and other specialists, to explore the intersection of participation, equity, and justice in AI systems.[17,54,55] Using a 3-horizons systems change model, participants identified key themes of:[56]

- **Adopting an Eco-Socio-Technical Systems Lens**
  – examining not only the technology but also the supply chains, organisational structures, cultural dynamics, and hidden networks of people and environmental resources that underpin AI systems and their implementations.
- **Inclusive Participation as a Lever for Systemic Change**
  – overcoming power asymmetries, cultural barriers and lack of transparency to use participation to drive innovation and systemic shifts.
- **Recognising Invisible Networks and Power Dynamics**
  – confronting invisible aspects of AIs such as the influence of capital, hidden labour, and environmental costs, and tackling inequalities (socio-economic, gender, regional etc.) in AI benefits and risks.
- **Mapping the Current System to Envision Change**
  – using systems-thinking to identify leverage points to drive more inclusive outcomes by thoroughly understanding the current state of AI and democratic systems, including their barriers and limitations, and co-defining ideal systems futures before designing pathways for change.

The Grid might usefully signpost innovators to uses of AIs called for by public participants in civic settings, which affect data capture and quality at source.

Novel Data Capture and Integration

Worn, mobile and frequently interacted with technologies are becoming ubiquitous. This raises the potential for citizens' avatars (transactional digital twins) to be used to personalise public services for individual benefit; improve public services for public benefit; and enable new research for science and society. The Grid could produce guidance and services for FAIR curation of data streams from emerging technologies including:[23]

- **Ambient / 'Nearable' Sensors**
  – sensing of person-level functioning and events such as gait, falls, activities of daily living, heart rate and respiratory rate from 4-dimensional radio image analysis in emerging home-based consumer technologies.

- **Worn Sensors**
  – sensing of biometrics and personal environmental exposome from devices worn in daily life, including sleep.

- **Voice Interaction**
  – voice-derived contextual information, including emotion, from the emerging shift in primary service access AIs from read/type to listen/talk technologies; plus, adaptive observation through 'conversations' that increase data resolution in specific contexts.

- **Mass Geolocation of Mobile Devices**
  – large-scale person-level geolocation and movement data, for example from accountable commercial aggregators of advertising user data from smartphones; plus, tracking of vehicle movements.

- **Smart Data (Mass Transactional Data)**
  – UK's Smart Data schemes across sectors like banking, energy, and telecommunications, ensuring that real-time, permissioned data flows can be effectively integrated into civic learning systems while maintaining appropriate governance and interoperability standards.

- **One Health Data Integration**
  – multi-species animal health data integration as a sentinel for shared environmental exposure and emerging disease risk.

- **Satellite, Drone and Fixed Camera Mass Imaging**
  – imaging of buildings (including thermal), traffic, crowds, natural environments (including weather) etc.

- **Environmental Sensors**
  – sensing of air quality, noise and other environmental factors from fixed installations.

- **Missing Sensor Data**
  – inference from patterns of missing data to understand the underlying reasons for missingness and help target innovations in sensors and sensing.

Together these data-streams can give new insights into person- and system-level dynamics at much higher resolution (in time, place and person) than most research projects can afford to capture. The complexity of the data demands resource-intensive fusion and wider

curation, which will intensify as more consumer technologies are linked to routine care. The Grid could achieve economies of scale in curation (similar to some of those current activities of the Geospatial Commission around personal mobility data or imagery), offering a better return on investment in research projects by leveraging FAIR research-ready data rather than funding silos of curation.[57]

Across the North of England alone there are six clusters of health technology research working on novel sensing and data capture. There is currently no national process for planning how the promising new data-streams can be fused with existing data to accelerate innovation without duplication. The Grid could help cohere the otherwise fragmented governance of integrated data, improving innovation processes and impacts.

Multi-Purpose Data Curation and Governance

At present, the separate governance of the same (or overlapping) datasets for different (but related) data processing purposes wastes time, money and opportunities for better governance. For example, NHS and linked data are typically curated and governed as separately for direct care, planning and research. This duplicates data systems, wasting scarce financial and human resources, and impeding whole-system insights. The Grid could increase efficiency and (cost) effectiveness for both UKRI and data controllers by offering:

- **Civic Data Governance for Research and Innovation Toolkit (GRIT)**
  – By analysing the data governance arrangements of civic clusters of data controllers across the UK – including understanding how they achieve joint accountability, economies of scale and offer mutual aid across constituent agencies – the Grid could produce a toolkit – GRIT – and use it to monitor the readiness of civic 'power stations' to join the 'grid' such that they can interoperate to borrow strength and contribute to national strength in global markets.

- **Privacy Preservation Tools/Services**
  – By promoting and/or sustaining software such as natural language processing tools to redact personally identifiable information from narrative records before research, and checking the outputs of research for the risk of disclosing individual identities (e.g. DARE SACRO) or adopting established tools with automated output checking (e.g. DataSHIELD), the Grid could improve both governance and agility of research with public service data. This increase in research activity embedded in civic systems would also improve data uses for service planning and evaluation.[58,59]

- **Structured Data Clusters and Complex Phenotyping for Reproducible Research**
  – By encouraging fuller use of provenance services around scientific claims, including metadata about algorithms used to derive structured data from un/semi-structured data, the Grid could improve research governance, reproducibility, and broader quality. In the mental health domain, for example, patterns of sleep, blood markers of inflammation and brain images may define an important syndrome (or phenotype) that has no specific diagnosis (and thus no diagnostic code in conventional structured records) but is key to understanding unmet need and measuring responses to a new treatment.[60] There is no current technical guidance for assuring the FAIR compliance and reproducibility of complex condition or multimodal research. As digital phenotyping services take root for simple conditions (e.g. in HDR UK), the Grid needs to address researchers' needs for finding, sharing and reusing more complex clusters of structured data.[61]

- **Entity Resolution across Persistent Longitudinal Record Domains**
  – By offering guidance and tools for privacy-preserving entity resolution (probabilistic identification of individuals across data silos where linkage of persistent longitudinal records is not feasible – commonly used in fraud detection), the Grid could enable valuable new discoveries and innovations.[62,63] For example, an AI for child/family protection may accelerate traditional safeguarding approaches that employ human insight generation across NHS, social care, police and other data sources without person-level linked records. Most inquiries into avoidable harms with safeguarding services cite late generation of insights from available data as a problem. Similar entity resolution might be applied to financial fraud or understanding antimicrobial resistance patterns across NHS, public health services, social care, farming and wastewater management. UKRI can leverage the UK's mandate to share data and workflows in this way for understanding complex systems of global importance.

Conversational Data Analytics and Capacity for Data Uses

The UK faces a tsunami of data, a blizzard of analytic outputs and drought of data scientists. This forecast is set to worsen in terms of the scarcity of human resource, yet commodity generative AIs (GenAI), Large Language Models and prompt engineering may be harnessed by the Grid to provide 'just-in-time' methodological training to public servants with key problem domain knowledge – securing and accelerating insights and upskilling topic experts in analysing their data sustainably.[64] This GenAI could also help align existing analytic capabilities within and across organisations.

- **Just-in-Time vs Just-in-Case Training in Methods and Tools**
  – The Grid could signpost researchers to resources that help novice data analysts, who are experts in the problem domain, to analyse their data robustly. These may include conversations with GenAI about research questions, input data, and modelling assumptions which generate analytic code - ideally to be discussed with expert methodologists. The tools might further signpost users to learning resources in the relevant methods – a just-in-time vs. conventional just-in-case training model, potentially earning professional development / educational credits.

- **Quality Assurance of Commodity AIs**
  – The Grid could collate a national gallery of best practice analyses of data and conversations with GenAI that generate the analytic code to reproduce the worked examples – if a GenAI subsequently fails 'inference safety' tests that it historically passed, then warnings can be issued.

- **UK Library of Research Objects for Training AIs to Teach Data Analytics**
  – The Grid could encourage civic clusters of data controllers to share data analyses that can be made public after disclosure checks – some may be worked up further into national training materials, others may give credit to service analysts for publishing their work. Most would have value for training AIs (including prompt engineering) to train humans in data analytic methodology, just-in-time.

Federated Discovery and Evaluation

The UK has rich public service data sources that can be used for research but fails to maximise this opportunity, particularly in clinical trials.[65] The Grid could work with research funding agencies to broker a 'market in data research excellence' between and across civic data clusters. This market would support high-profile federated studies with civic systems that meet data research readiness standards as follows:

- **Making System Heterogeneity Part of Good Research Design**
  – The Grid could curate best practice examples of where observational studies have been made more robust by incorporating important variation at system level – the national grid of civic, place-based systems of public services could be made visible as an internationally important asset in this way.
  – The Grid could broker the incorporation of federated analytic tools such as those developed by DataSHIELD, DARE UK TRE-FX, Bitfount, Vantage 6 and Lifebit into civic data analytic systems, supporting collaboration to attract inward investment in research studies and to spread the workload in challenging public service analyses across systems.[59,66-69]
  – The Grid could facilitate the dissemination of learnings after place-based innovation has been conducted, with place-based prioritisation of research.

- **Automating Trial Feasibility Assessment and Recruitment**
  – The Grid could signpost civic systems to tools that can run analyses to identify citizens who meet criteria for trials that may be of interest to them. When the tools are Grid compliant then a trial sponsor could have one system run a feasibility assessment on behalf of other civic 'power stations' on the Grid, thus offering a national variety of trial domain expertise as points of access to nationally consistent data/workflow systems.

- **Automating Long Term Follow-up from Trials**
  – The Grid could broker the fusion of observational study and trials services to make long-term follow up of trials participants easier, cheaper and more complete.

- **Disseminating Cost Models**
  – The Grid could publish best practice cost models that improve economic evaluation linked to trials and improve policy option appraisals for commissioners of public services – the more they are used for service planning the better they become for research.

- **Supporting Progressive Regulation**
  – The Grid could help regulators (e.g. the Medicines and Healthcare products Regulatory Agency (MHRA) use civic data and digital workflows to accelerate regulatory processes and make them more inclusive of the people needing the products. In this way, UK regulation could become more efficient and proportionate, thereby enhancing innovation and economic growth. The Manchester-led InSilicoUK is working on how to incorporate real-world data into in-silico trials/studies to produce regulatory grade evidence systematically.[70-72]

- **Systematising Health Technology Assessment**
  - The Grid could help health technology assessment bodies, including the National Institute for Health and Care Excellence (NICE) in their Transformation Programme to use civic data and digital workflows to incorporate and maintain NICE Guidance systematically within the network of learning systems.

- **Facilitating Synthetic Control Analyses for Evaluating Natural Experiments**
  – The Grid could publish best practice guidance in synthetic control analyses and broker federation across civic systems where natural experiments (e.g. early implementation of a policy option) are of mutual interest. Advances in machine learning for control matching with complex data structures may be presented alongside state-of-the-art statistical modelling approaches.

- **Improving Understanding and Quality of Routinely Collected Data**
  – The Grid could support the culture change needed to improve the quality of routinely collected data. This starts with a shared understanding of data quality between those who input the data and those who process the data for different purposes across the civic system. Workforce education and provision of the appropriate tools and time for high quality data generation are key. The improved understanding is needed to maximise the value of increasingly ubiquitous tools, such as natural language processing to extract concepts from the text clinicians type and organise them with common data models producing the datasets analysts use for service evaluation, planning and research. The Grid can use its digital commons to maximise the value of open-source tools in terms of education, systems integration and economical maintenance of code. In addition, more efficient use of computationally expensive resources brokered by the Grid would reduce the environmental impact of these AI solutions.

- **Mobilising Citizen Research Avatars**
  – The Grid could broker the formation of national cohorts for specific studies from residents of civic systems ‘donating their digital twins’ for broad ranging research. The underpinning digital services would be used in civic systems to crowdsource insights that require topic-specific interaction with participants – the Grid would convene developers and run challenges to maximise information while minimising participant burden through their interactive digital twin or research avatar, which is habit/preference/context sensitive.[60,71,73]

- **Integrating Citizen Research Avatars with National Studies**
  – The Grid could help national studies such as Our Future Health to connect with participants via the civic learning system grid and its interaction with residents.[74] This could make collection of biological samples, experiences or personal exposomes etc. more adaptive, yielding more discovery power and better value for money. It could also tap into civic pride in mobilising under-represented groups into these studies, better reflecting those most affected by the results.

- **Fostering and Facilitating Economic Growth**
  – The Grid could help lower the barriers to testing technological innovations with relevant data helping innovators to get a detailed understanding of the diversity of contexts in which their innovations may be deployed. This information is crucial to robust business cases that remain credible when they are executed at scale. The

Grid will also provide a focal point for progressive regulation (e.g. clear guidance, training and civic sandboxes for advancing regulatory processes) – safely removing barriers that hold back start-ups from flourishing. It would also help to coordinate the upskilling of regional workforces in data science and AI. The Grid would thereby generate growth that makes a tangible difference to the UK's economy.

Personalised and Networked Prediction

The early recognition of key patterns from data (e.g. medical diagnoses or consumer trends) and prompt intervention to prevent harm and improve outcomes is key to maximising the value of most public services and many commercial operations. For example, the NHS 10-year plan consultation seeks ways to become more predictive and preventive – delivering a health service that costs less and maximises health in the long run. This cannot be achieved without radical improvements in data uses and methods of prediction. The Grid could support this progress as follows:

- **Promote Best Practice in Causal, Counterfactual Prediction**
  – The Grid could improve prediction research by signposting researchers to best practice examples of causally robust prediction models where the UK has leading methodology and data; the current literature is littered with poor models that underpin decision support systems, affecting individuals in ways that were often not intended or anticipated.[75]
  – The Grid could showcase an 'implementation-first' viewpoint to prediction modelling, where implementation considerations of the model are done as part of study design, not an afterthought post development. Similarly, The Grid could showcase best practice model development, evaluation, and deployment, for different purposes such as population risk stratification in commissioning services and counterfactual prediction in choosing service options for individuals.[76] Predictions should consider multiple outcomes in order to give holistic views of individuals' lives. Emerging methods can be explored on these causally robust foundations, for example transfer learning AI approaches to borrowing strength for predicting specific groups' outcomes from wider population data.

- **Facilitate Routine Maintenance of Predictive Models with Evolving Data**
  – The Grid could mitigate the risks of key predictive models drifting in calibration (where a forecasting model or decision support tool becomes less reliable and accurate after its deployment) or other aspects of predictive performance as contexts change over time and data drifts. The UK's world class data and methodology can be harnessed to maintain predictive models after publication using AI augmented surveillance of predictive performance and dynamic modelling approaches. The Grid could convene methodologists and civic data/problem holders to nurture data science in this field. This will ensure that as new data becomes available, models are maintained after publication and continue to deliver value.[77-79]

- **Progressive Regulation and Implementation of Predictive Models**
  – Most clinical prediction models are not implemented following development. This does not mean they are not useful. The barriers to implementation are high, including regulation, and implementation is often a post-development afterthought. Current regulatory frameworks focus on rigorous pre-implementation assessment based on

old data. The Grid would enable prospective surveillance of model performance with contemporary data, which could a) increase the safety-in-use of predictive models, b) lower barriers to adoption, and c) improve quality assurance of algorithms because they evolve with cumulative data and use.

- **Facilitate Brokerage Between Digital Twins, Predictive Models and Decisions** – The Grid might work with innovation partners in civic clusters to map libraries of predictive models to individuals' interactive digital twins (or avatars) – in similar ways to how investigators map potential participants to trials. This would use graph-based integration of person-level data to link relevant predictive models to decisions over service needs and options, as illustrated in Figure 2.

- **Promote Advances in Ideographic Searches of Complex Longitudinal Data** – The Grid could set engineering challenges with research funders, leading to new ways of visualising citizen journeys in complex systems. This would complement probabilistic stratification of populations with tools that allow the researcher to ask the AI, "show me the people in this population with a journey like this person's…"

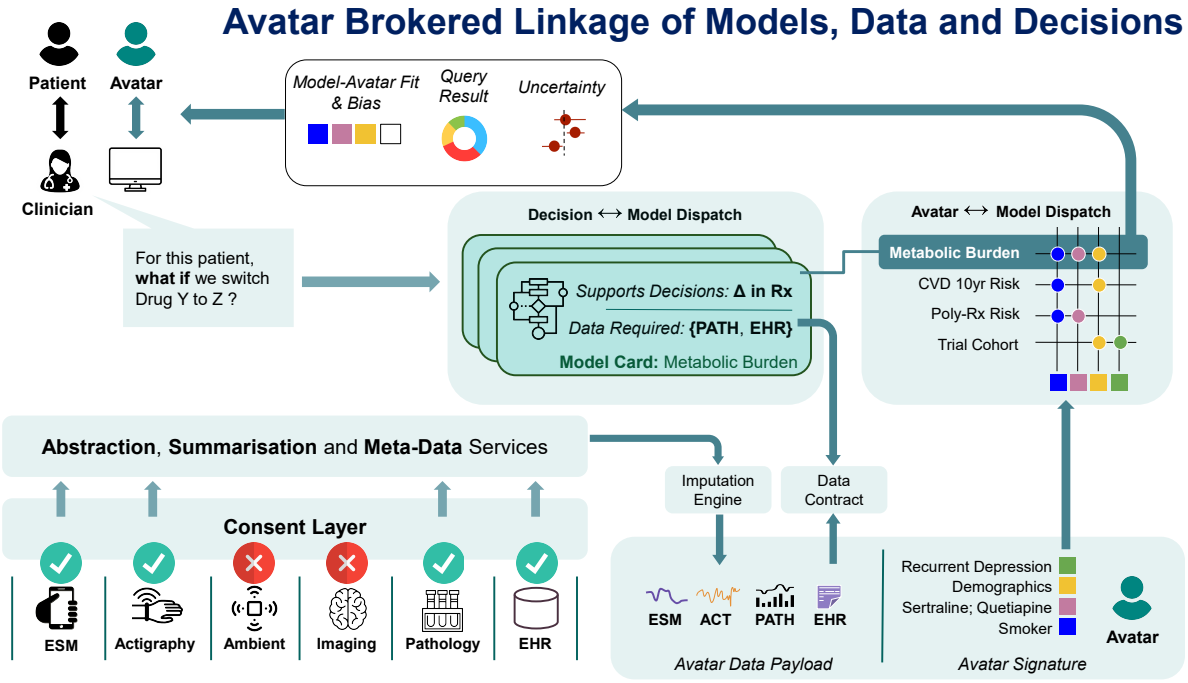

Figure 2: Systematising uses of predictive models for service needs and options using a graph-based representation of patient data, personal preferences and adaptive interaction – a transactional digital twin or avatar.

Synthetic Data and Scenario Simulators

Civic clusters of person-level and small-area-level linked longitudinal data that offer increased precision also risk deductive disclosure. The statistical structure in the data holds potential value for researchers and innovators. Advances in machine learning can surface useful amounts of the inherent joint probability distributions in multi-variable data (e.g. the patterns of relationships across all possible combinations of individual data items) that are key to making discoveries and training AI models. For instance, attention-based

architectures like transformers, the core of language models, can elucidate relationships among the input features, even when they originate from different data sources on the same individuals. Building foundational (large, reusable and repurposable) models of voluminous data representing substantial aspects of public services in civic systems is a computationally expensive problem. Furthermore, it can yield models that can leak information they are trained on and consequently, may identify individuals. The Grid would be well placed to broker between funders, data controllers, application domain researchers / innovators, and methodologists to advance this field as follows:

- **Synthetic Data Quality Tests**
  – The Grid could curate a library of algorithms for data controllers to test the utility of synthetic data trained on the real data they control.

- **Synthetic Data Disclosure Risk Management**
  – The Grid could signpost data controllers to vulnerabilities and the required intrusion testing algorithms and services and provide best practice guidance on managing the disclosure risks of models that generate synthetic data.

- **Economical High-Performance Computing for Synthetic Data Generation**
  – The Grid could work with research funders and STFC to establish the most cost-effective and environmentally sustainable way a civic system (or collaborative of systems) could consume national high performance computing resources to provide (a national grid) of synthetic civic data laboratories for science and technology engineering (in public research or commercial uses). This would ensure not only that synthetic data generation is cost-effective but also aligned with scientific and innovation priorities.

- **Synthetic Data as a platform for 'What If' scenarios**
  – Synthetic data can prime individual-level simulations (i.e., microsimulations or agent-based models) and substantially accelerate the development of close-to-reality in-silico policy experiments. In-silico policy experiments allow policymakers to simulate various scenarios, assessing potential outcomes without the cost and ethical considerations of real-world trials and can revolutionise the policymaking process by making it more data-driven, proactive, and efficient.[80-82]

- **Preparation of Research Objects for Publication**
  – The Grid could provide best practice guidance, and possibly services, for researchers to replace the real data in their Research Objects with plausible substitute synthetic data producing the same answers, with further disclosure checks ahead of open publication.[24]

- **Supporting International Collaboration**
  – The Grid could provide services to help civic data controllers to consistently and safely generate synthetic datasets, facilitating international collaboration and investment.

- **Researching Agent-Based Civic Futures**
  – The Grid could help researchers to explore the potential impacts of the increasing use of agent-based systems acting autonomously on behalf of human users and how they interact with other people, services and environments where they live.

Conclusions

The UK stands at a pivotal moment on its journey toward data-driven innovation and scientific excellence - poised to unlock opportunities for national economic growth. This paper proposes a transformative vision for a national grid of civic learning systems, enabling unprecedented, trusted access to public sector datasets while fostering continuous improvement in public services. By establishing a robust national infrastructure for data sharing and analysis, the Grid would serve as a catalyst for scientific discovery and an engine for public service innovation. This approach would transform currently siloed data initiatives into an integrated ecosystem where local insights strengthen national capabilities, and national resources empower local solutions. It would also improve the quality of research by providing fuller reflections of civic systems in nationally available data, alongside contextual metadata that is key to the best modelling and interpretation of results. The proposed framework would make public sector datasets more accessible to researchers and would also create a sustainable model where scientific advancement and civic improvement become mutually reinforcing goals. This would position the UK as a global leader in data-driven public service innovation and attract inward investment in research. This approach can help to deliver the Government's Missions in economic growth, health and opportunity. In particular, the Grid would help to realise the AI Opportunities Action Plan by unlocking data assets with public trust and supporting data-driven innovation with regional reciprocity.

References

1. Science, Innovation and Technology Select Committee. How can science and tech innovation boost regional economic growth? https://committees.parliament.uk/committee/135/science-innovation-and-technology-committee/news/204260/how-can-science-and-tech-innovation-boost-regional-economic-growth/ Accessed 16/1/25.
2. Department for Science, Innovation and Technology. AI Opportunities Action Plan. https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan Accessed 18/1/25.
3. Bambra C, Munford L, et al. Health for Wealth: Building a Healthier Northern Powerhouse for UK Productivity. Northern Health Science Alliance 2018. https://www.thenhsa.co.uk/app/uploads/2018/11/NHSA-REPORT-FINAL.pdf Accessed 21/1/25.
4. NP11 and Northern Health Science Alliance. A Northern Life Sciences Supercluster: The Economic Potential of a Systemwide Approach. 2021. https://www.thenhsa.co.uk/app/uploads/2021/09/A-Northern-Life-Sciences-Supercluster_NP11-NHSA-report-FINAL-FOR-PUBLICATION-1.pdf Accessed 21/1/25.
5. Northern Health Science Alliance. NHSA Analysis of the UK Clinical Research Landscape in 2022. 2024. https://www.thenhsa.co.uk/app/uploads/2024/02/2022-Health-data-analysis.pdf Accessed 21/1/25.
6. Ainsworth J, Buchan I. Combining Health Data Uses to Ignite Health System Learning (2015) Methods Inf Med; 54(06): 479-487 https://pubmed.ncbi.nlm.nih.gov/26395036/
7. Bennett D, Lee P, et al. Children in Care in the North of England. 2024: A report prepared for the Child of the North All-Party Parliamentary Group. Northern Health Science Alliance 2024. https://www.healthequitynorth.co.uk/app/uploads/Children-in-Care-Report-2024-FINAL-2.pdf Accessed 21/1/25.
8. Policy Connect. Our Place Our Data: Involving Local People in Data and AI Based Recovery. https://www.policyconnect.org.uk/research/our-place-our-data-involving-local-people-data-and-ai-based-recovery Accessed 2/12/24.
9. Department of Health and Social Care. Summary letter from Lord Darzi to the Secretary of State for Health and Social Care. https://www.gov.uk/government/publications/independent-investigation-of-the-nhs-in-england/summary-letter-from-lord-darzi-to-the-secretary-of-state-for-health-and-social-care Accessed 5/12/24
10. Howard A, Aston S, Gerada A, Reza N, Bincalar J, Mwandumba H, Butterworth T, Hope W, Buchan I. Antimicrobial learning systems: an implementation blueprint for artificial intelligence to tackle antimicrobial resistance. Lancet Digit Health. 2024 Jan;6(1):e79-e86. https://doi.org/10.1016/S2589-7500(23)00221-2.
11. Mair FS, Nickpour F, Nicholl B, MacDonald S, Joyce DW, Cooper J, Dickson N, Leason I, Abbasi QH, Akin IF, Deligianni F, Camacho E, Downing J, Garrett H, Johnston Gray M, Lowe DJ, Imran MA, Padmanabhan S, McCowan C, Clarkson PJ, Walker LE, Buchan I. Developing SysteMatic: Prevention, precision and equity by design for people living with multiple long-term conditions. J Multimorb Comorb. 2024 Sep 30;14:26335565241272682. https://doi.org/10.1177/26335565241272682

12. Piroddi R, Baker W, Daras K, Buchan I, Rafferty J, Astbury A, Barr B. System-wide health needs segmentation: innovating integrated care for complex needs households. European Journal of Public Health, Volume 34, Issue Supplement_3, November 2024, ckae144.036, https://doi.org/10.1093/eurpub/ckae144.036.
13. Tran PB, Kazibwe J, Nikolaidis GF, Linnosmaa I, Rijken M, van Olmen J. Costs of multimorbidity: a systematic review and meta-analyses. BMC Med. 2022;20(1):234. https://www.doi.org/10.1186/s12916-022-02427-9.
14. Health e-Research Centre. Data Saves Lives. https://www.herc.ac.uk/ Accessed 2/12/24.
15. NHS Confederation, London. Understanding Patient Data. https://understandingpatientdata.org.uk/ Accessed 2/12/24.
16. NHSA Connected Health Cities Impact Report. Northern Health Science Alliance 2020. https://www.chc-impact-report.co.uk/ Accessed 19/1/25.
17. Civic Data Coop. Liverpool City Region Civic Data Cooperative https://civicdatacooperative.com/ Accessed 2/12/24.
18. BBC, London. Care.Data: How did it go so wrong? https://www.bbc.co.uk/news/health-26259101 Accessed 2/12/24.
19. BBC, London. NHS data sale 'an invasion of privacy', campaigners say. https://www.bbc.co.uk/news/uk-england-somerset-57568711 Accessed 2/12/24.
20. Foxglove, London. Legal action launched: no legal basis for the £330 million Palantir NHS Federated Data Platform. https://www.foxglove.org.uk/2023/11/30/legal-action-palantir-nhs-federated-data-platform/ Accessed 2/12/24.
21. Civic Data Coop. Round 'Ere. https://civicdatacooperative.com/round-ere/ Accessed 2/12/24.
22. Wikipedia. e-Science. https://en.wikipedia.org/wiki/E-Science Accessed 2/12/24.
23. GO FAIR. FAIR Principles. https://www.go-fair.org/fair-principles/ Accessed 2/12/24.
24. Bechhofer, S., De Roure, D., Gamble, M. et al. Research Objects: Towards Exchange and Reuse of Digital Knowledge. Nat Prec (2010). https://www.nature.com/articles/npre.2010.4626.1
25. researchobject.org. Enabling reproducible, transparent research. https://www.researchobject.org/ Accessed 2/12/24.
26. Ainsworth J, Buchan I. e-Labs and Work Objects: Towards Digital Health Economies. In: Mehmood, R., Cerqueira, E., Piesiewicz, R., Chlamtac, I. (eds) Communications Infrastructure. Systems and Applications in Europe. EuropeComm 2009. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2009, vol 16. Springer, Berlin, Heidelberg. https://link.springer.com/chapter/10.1007/978-3-642-11284-3_22
27. Health Data Research UK. Trusted Research Environments. https://www.hdruk.ac.uk/access-to-health-data/trusted-research-environments/ Accessed 2/12/24.
28. UK Data Service. What is the Five Safes framework. https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/ Accessed 2/12/24.

29. DARE UK. SATRE: Standardised Architecture for Trusted Research Environments. https://dareuk.org.uk/how-we-work/previous-activities/dare-uk-phase-1-driver-projects/satre-standardised-architecture-for-trusted-research-environments/ Accessed 2/12/24.

30. European Council. European Health Data Space: Council adopts new regulation improving cross-border access to EU health data. 2025. https://www.consilium.europa.eu/en/press/press-releases/2025/01/21/european-health-data-space-council-adopts-new-regulation-improving-cross-border-access-to-eu-health-data Accessed 22/1/25.

31. DARE UK. https://dareuk.org.uk Accessed 2/12/24.

32. ADR UK. Administrative data for public good. https://www.adruk.org/ Accessed 2/12/24.

33. Health Data Research UK. Federated Analytics. https://www.hdruk.ac.uk/research/research-data-infrastructure/federated-analytics/ Accessed 2/12/24.

34. Wikipedia. Digital Economy Act 2017. https://en.wikipedia.org/wiki/Digital_Economy_Act_2017 Accessed 2/12/24.

35. Office for National Statistics. Secure Research Service. https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice Accessed 2/12/24.

36. Science and Technology Facilities Council (STFC) https://www.ukri.org/councils/stfc/ Accessed 2/12/24.

37. Zhang X, Tulloch JSP, Knott S, et al Evaluating the impact of using mobile vaccination units to increase COVID-19 vaccination uptake in Cheshire and Merseyside, UK: a synthetic control analysis BMJ Open 2023;13:e071852. https://doi.org/10.1136/bmjopen-2023-071852

38. The University of Liverpool. Covid-SMART Community Testing Evaluations. www.liverpool.ac.uk/coronavirus/research-and-analysis/covid-smart-pilot/ Accessed 18/1/25.

39. García-Fiñana M, Hughes DM, Cheyne CP, Burnside G, Stockbridge M, Fowler TA, Fowler VL, Wilcox MH, Semple MG, Buchan I. Performance of the Innova SARS-CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot: population based cohort study. BMJ. 2021 Jul 6;374:n1637. https://doi.org/10.1136/bmj.n1637.

40. Zhang X, Barr B, Green M, Hughes D, Ashton M, Charalampopoulos D, García-Fiñana M, Buchan I. Impact of community asymptomatic rapid antigen testing on covid-19 related hospital admissions: synthetic control study. BMJ. 2022 Nov 23;379:e071374. https://doi.org/10.1136/bmj-2022-071374.

41. Barr B, Zhang X, Green M, Buchan I. A blueprint for synthetic control methodology: a causal inference tool for evaluating natural experiments in population health. BMJ. 2022 Nov 23;379:o2712. https://doi.org/10.1136/bmj.o2712.

42. Green MA, García-Fiñana M, Barr B, Burnside G, Cheyne CP, Hughes D, Ashton M, Sheard S, Buchan IE. Evaluating social and spatial inequalities of large scale rapid lateral flow SARS-CoV-2 antigen testing in COVID-19 management: An observational

study of Liverpool, UK (November 2020 to January 2021). Lancet Reg Health Eur. 2021 Jul;6:100107. https://doi.org/10.1016/j.lanepe.2021.100107.

43. Vekaria, B., Overton, C., Wiśniowski, A. et al. Hospital length of stay for COVID-19 patients: Data-driven methods for forward planning. BMC Infect Dis 21, 700 (2021). https://doi.org/10.1186/s12879-021-06371-6

44. Munford L, Khavandi S, et al. A year of COVID-19 in the North: Regional inequalities in health and economic outcomes. Northern Health Science Alliance 2021. https://www.healthequitynorth.co.uk/app/uploads/2023/03/A-Year-of-COVID-in-the-North-report-2021.pdf Accessed 18/1/25.

45. Scott S, Salvatelli R, et al. Navigating the Long Haul: Understanding Long Covid in Northern England. Health Equity North 2024. https://www.healthequitynorth.co.uk/app/uploads/LONG-COVID-REPORT-2024-FINAL.pdf Accessed 18/1/25.

46. NHS England. The NHS Research Secure Data Environment Network. https://digital.nhs.uk/data-and-information/research-powered-by-data/sde-network Accessed 2/12/24.

47. NHS England. Research, powered by NHS Data. https://digital.nhs.uk/data-and-information/research-powered-by-data Accessed 2/12/24.

48. Rodgers SE, Geary RS, Villegas-Diaz R, Buchan IE, Burnett H, Clemens T, Crook R, Duckworth H, Green MA, King E, Zhang W, Butters O. Creating a learning health system to include environmental determinants of health: The GroundsWell experience. Learn Health Syst. 2024 Oct 10;8(4):e10461. https://doi.org/10.1002/lrh2.10461

49. Greater Manchester Integrated Care Partnership. The Greater Manchester Creative Health Strategy. 2022. https://gmintegratedcare.org.uk/wp-content/uploads/2022/11/gm-creative-health-strategy-low-res.pdf Accessed 21/1/25.

50. University of Liverpool. Civic Health Innovation Labs (CHIL) https://www.liverpool.ac.uk/civic-health-innovation-labs/ Accessed 2/12/24.

51. Consumer Data Research Centre. https://www.cdrc.ac.uk/ Accessed 2/12/24.

52. Smart Data Research UK. Let's do good things with data. https://www.sdruk.ukri.org/ Accessed 2/12/24.

53. Ada Lovelace Institute. Participatory Data Stewardship. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/09/Participatory-data-stewardship_Final-report.pdf Accessed 2/12/24.

54. ESRC Digital Good Network. https://digitalgood.net/ Accessed 2/12/24.

55. Elgon Social Research. https://elgon.social/ Accessed 2/12/24.

56. Wikipedia. Three Horizons. https://en.wikipedia.org/wiki/Three_Horizons Accessed 2/12/24.

57. Geospatial Commission. National Underground Asset Register. https://www.gov.uk/government/organisations/geospatial-commission Accessed 2/12/24.

58. DARE UK. SACRO: Semi-Automated Checking of Research Outputs https://dareuk.org.uk/how-we-work/previous-activities/dare-uk-phase-1-driver-projects/sacro-semi-automated-checking-of-research-outputs/ Accessed 2/12/24.

59. DataSHIELD. https://datashield.org/ Accessed 2/12/24.

60. Mental Health Research for Innovation Centre. Our research. Mental Health Avatar. https://mric.uk/our-research/mental-health-avatar/ Accessed 5/12/24.

61. HDR UK Phenotype Library https://phenotypes.healthdatagateway.org/ Accessed 5/12/24.

62. Wikipedia. Record linkage. https://en.wikipedia.org/wiki/Record_linkage Accessed 2/12/24.

63. Quantexa. What is Entity Resolution and How Does It Transform Data Into Value? https://www.quantexa.com/resources/entity-resolution-guide/ Accessed 2/12/24.

64. Cabinet Office; Central Digital & Data Office. Generative AI Framework for HMG. https://www.gov.uk/government/publications/generative-ai-framework-for-hmg/generative-ai-framework-for-hmg-html Accessed 2/12/24.

65. Association of British Pharmaceutical Industries. NHS patients losing access to innovative treatments as UK industry clinical trials face collapse. 2022. https://www.abpi.org.uk/media/news/2022/october/nhs-patients-losing-access-to-innovative-treatments-as-uk-industry-clinical-trials-face-collapse/ Accessed 19/1/25.

66. DARE UK. TRE-FX: Delivering a federated network of trusted research environments to enable safe data analytics. https://dareuk.org.uk/how-we-work/previous-activities/dare-uk-phase-1-driver-projects/tre-fx-delivering-a-federated-network-of-trusted-research-environments-to-enable-safe-data-analytics/ Accessed 2/12/24.

67. Bitfount. Data science and AI without borders. https://www.bitfount.com/ Accessed 2/12/24.

68. Vantage. Vantage6. https://vantage6.ai/ Accessed 2/12/24.

69. Lifebit. Get the Data Now. In the World's Largest Federated Data Network. https://www.lifebit.ai/ Accessed 2/12/24.

70. KPMG (United Kingdom): KPMG Life Science Regulatory Solutions and InSilicoUK Pro-Innovation Regulations Network 2024. In Silico Regulatory Evidence Utilisation within the Life Science Sector. https://doi.org/10.5281/zenodo.12735158

71. Frangi A, Denison T, Myles P, et al. Unlocking the power of computational modelling and simulation across the product lifecycle in life sciences: A UK Landscape Report. InSilicoUK Pro-Innovation Regulations Network 2023. https://doi.org/10.5281/zenodo.8325274

72. Redrup E, Mitchell C, Myles P, Branson R, Frangi A F. Cross-Regulator Workshop: Journeys, experiences and best practices on computer modelled and simulated regulatory evidence— Workshop Report. InSilicoUK Pro-Innovation Regulations Network 2023. https://doi.org/10.5281/zenodo.10121103

73. Buchan IE, Winn JM, Bishop CM: *A unified modeling approach to data-intensive healthcare*. The Fourth Paradigm 2009: 91-97. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/4th_paradigm_book_part2_buchan.pdf Accessed 2/12/24.

74. Our Future Health. Let’s prevent disease together. https://ourfuturehealth.org.uk/ Accessed 5/12/24.

75. Prognosis Research. https://www.prognosisresearch.com/ Accessed 2/12/24.

76. Prosperi M, Guo Y, Sperrin, M, Koopman J, Min J, He X, Rich S, Wang M, Buchan I, Bain, J. Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nature Machine Intelligence 2020;2(7):369-375. http://doi.org/10.1038/s42256-020-0197-y.

77. Lin L, Sperrin M, Jenkins DA, Martin GP, Peek N. A scoping review of causal methods enabling predictions under hypothetical interventions. Diagn Progn Res. 2021 Feb 4;5(1):3 http://doi.org/10.1186/s41512-021-00092-9.

78. Hughes DM, Komárek A, Bonnett LJ, Czanner G, García-Fiñana M. Dynamic classification using credible intervals in longitudinal discriminant analysis. Statistics in Medicine 2017; 36:3858-3874. http://doi.org/10.1002/sim.7397

79. Hughes DM, Komárek A, Czanner G, Garcia-Fiñana M. Dynamic longitudinal discriminant analysis using multiple longitudinal markers of different types. Statistical Methods in Medical Research 2016; https://doi.org/10.1177/0962280216674496.

80. Squazzoni F, Bianchi F. Exploring Interventions on Social Outcomes with In Silico, Agent-Based Experiments. In: Damonte, A., Negri, F. (eds) Causality in Policy Studies. Texts in Quantitative Political Analysis. 2023. Springer, Cham. https://doi.org/10.1007/978-3-031-12982-7_9

81. Reshef Kera D, Navon E, Wellner G, Kalvas F. Governance in Silico: Experimental Sandbox for Policymaking over AI Agents, in Gray C, Ciliotta-Chehade, E, Hekkert P, Forlano L, Ciuccarelli P, Lloyd P. (eds.), DRS2024: Boston, 2024, 23–28 June, Boston, USA. https://doi.org/10.21606/drs.2024.200

82. SyntheticMass - Synthetic patient and population health data for the state of Massachusetts. https://synthea.mitre.org/ Accessed 20/1/25.