Realising the potential of non-traditional data for research in Europe

Advancing access and re-use for improving health and wellbeing

Stefaan Verhulst

SUMMARY

Europe has invested heavily in data infrastructure through initiatives like the European Open Science Cloud (EOSC) and the European Health Data Space (EHDS), yet these frameworks remain largely focused on traditional, sector-specific data. This *CEPS R&D Perspective* argues that non-traditional data — generated by digital platforms, sensors, mobility systems, and consumption patterns — represents a largely untapped resource for health and well-being research. Drawing on evidence from over 290 studies, the note argues that non-traditional data can enable earlier detection of health risks, reveal nutritional and access inequalities, and link environmental exposure to disease outcomes. Grocery transaction records, wearable devices, satellite imagery, and mobility data have each shown measurable value when responsibly linked with conventional datasets.

Realising this potential at scale requires addressing persistent barriers: fragmented access, complex data linkage, fragile public trust, and short-term funding cycles. The paper proposes six integrated policy actions — expanding EOSC's mandate, broadening EHDS to cover health determinants, leveraging Horizon Europe for sustainable data reuse, activating DSA Article 40 for platform data access, professionalising data stewardship, and embedding social licence frameworks in governance structures. Together, these measures would transform Europe's data spaces from compliance infrastructure into genuine decision intelligence for health, well-being, and beyond.

Stefaan Verhulst is the Co-Founder of The GovLab (NewYorkCity) and The DataTank (Brussels), and Research Professor at the Tandon School of Engineering, New York University. He was a member of the High Level Expert Committee to the European Commission on Business-to-Government Data Sharing; and a member of the High Level Expert Committee to the Eurostat on Private Sector Data for Official Statistics.

CEPS ▪ Place du Congrès 1 ▪ B-1000 Brussels ▪ Tel: (32.2) 229.39.11

CONTENTS

INTRODUCTION: GIVING EUROPE'S DATA SPACES A PUBLIC PURPOSE

In recent years, Europe has invested heavily in advancing open science and building data spaces. The European Open Science Cloud (EOSC) and the European Health Data Space (EHDS) represent two of the most ambitious public data infrastructure projects within Europe. As impressive as these initiatives are, however, they remain largely focused on technical interoperability and legal compliance for traditional, sectorially specific forms of data. Like many similar data initiatives around the world, they tend to under-emphasise the potential of emerging forms of non-traditional data, for instance those generated through digital platforms, sensors, mobility systems, consumption patterns, and environmental monitoring.

This policy brief focuses on the potential of such data with regard to the domain of health and well-being. Our research in this sector demonstrates how non-traditional data sources already generate measurable value for understanding population risks, behavioural and environmental determinants, and structural inequalities. This research built upon evidence from more than 290 studies to illustrate that non-traditional data can support earlier detection of health threats, reveal nutrition and access disparities, and link environmental exposure to disease patterns. While this brief focuses on health and well-being, its usefulness and findings extend beyond a single sector: similar approaches can be used for a variety of domains, including agriculture, climate change, inequality, and more.

In what follows, we first present a brief overview of non-traditional data, discussing their sources, potential, and some associated risks (Section I). We then present empirical evidence showing how non-traditional data can be linked to specific health outcomes and risks (Section II), followed by a discussion of risks (III). Section IV examines the European strategic opportunity and presents some specific policy outcomes.

The overarching message of this brief is that the key policy challenge for European decision-makers is no longer whether non-traditional data can contribute to research and policy, but how Europe can move from fragmented pilots to a coordinated, legitimate, and sustainable system of data reuse. More specifically, this brief argues that EOSC, EHDS, and Horizon Europe should be aligned as mutually reinforcing levers to unlock non-traditional data for research and transform data spaces into decision intelligence for health and well-being. Further, policymakers and researchers should also explore how vetted researcher access under DSA Article 40 can enable responsible reuse of platform-generated data for other public-interest research.

1. NON-TRADITIONAL DATA: A GROWING LANDSCAPE

In recent years, it has become increasingly clear that non-traditional data sources can significantly expand the evidence base for understanding health, well-being, and various other social domains. Beyond their original or specifically defined contexts, these data are being repurposed to generate insights into population risks, behavioural patterns, environmental exposure, and structural inequalities. Broadly, the sources and types of non-traditional data include social media activity, mobility and geolocation data, digital transaction and consumption records, data from sensors and Internet-of-Things (IoT) devices, satellite imagery, and more. Advances in computational methods, machine learning, generative AI (which heavily relies on large unstructured datasets) and data linkage techniques have further increased the analytical value of these data, enabling their use at greater scale than previously.

For the purposes of this brief, we define non-traditional data as: digital data generated through human interactions with technology platforms, devices, and systems. This definition encompasses a broad range of data types and sources, from wearable devices and smartphone sensors to social media applications, satellite imagery, and digitised administrative or legal records. In this sense, non-traditional data includes what is sometimes referred to as social, smart, born-digital, or digital footprint data, unified not by sector or content but by mode of production and original purpose outside conventional statistical or clinical reporting systems.

In what follows, we present empirical evidence linking non-traditional data to specific health outcomes and risks. Comparable evidence can be found in other domains, such as agriculture (e.g., satellite and sensor data for crop stress and food security), climate adaptation (mobility and social sensing data for disaster response), and social inequality (digital traces revealing structural disparities). Together, these examples demonstrate that the value of non-traditional data lies not in replacing traditional sources, but in complementing them, particularly where official data are delayed, incomplete, or blind to lived conditions.

At the same time, the use of non-traditional data raises a distinct set of challenges and risks. These include:

- Privacy and consent, particularly when data are reused beyond their original context;
- Access and fragmentation, as high-value data are often held by private actors under restrictive terms or otherwise siloed;

- Data quality and bias, reflecting uneven coverage, cultural and social exclusion, and errors or other forms of data corruption; and
- Public legitimacy and trust, which cannot be secured through legal compliance alone but requires what we have elsewhere called a "social licence."

Keeping these risks in view is essential in any effort to scale the use of non-traditional data. However, the appropriate response is not to forego such data altogether, but to enable their responsible, transparent, and purpose-bound use. Governance is therefore central to balancing the possibilities of misuse against the risks of missed use. In Section IV, we examine some key governance issues surrounding non-traditional data in the European health research space.

2. EVIDENCE FROM RESEARCH: THE POTENTIAL OF NON-TRADITIONAL DATA

Our recent research demonstrates that non-traditional data sources are no longer peripheral to health and well-being research. Across hundreds of studies, these data have been shown to generate actionable insights when responsibly reused and linked with traditional clinical, environmental, and administrative datasets. These findings are particularly true in areas where conventional data is delayed, incomplete, or structurally blind to lived conditions—for instance, the early detection of population-level risk, the measurement of behavioural and environmental determinants, and the analysis of structural drivers of health inequality.

In this section, we examine four specific types of non-traditional data and link them to health and well-being insights. These categories are intended to be illustrative rather than exhaustive: they highlight the potential of non-traditional data, without documenting every possible use case.

1. Grocery Purchase Data and Diet-Related Health Risks: A study using smartphone-based receipt collection services compared the purchasing behaviour of households with and without individuals living with diabetes. The analysis revealed systematic differences in nutritional choices, demonstrating how loyalty card and transaction data can surface diet-related risks and inform preventive policy design. This evidence suggests that consumption data—when governed appropriately—can complement clinical indicators and provide early signals of population-level nutritional inequalities. Unlike clinical records, which typically capture outcomes after diagnosis, consumption data makes it possible to observe dietary patterns earlier, and at population scale.

2. Wearable Devices and Mental Health Monitoring: Multiple studies in our research show that wearable and smartphone sensor data can detect patterns associated with depression, sleep disruption, and stress. Research using self-supervised learning on unlabeled wearable data demonstrated that mood disorder signals can be identified at scale, pointing to the potential of passive data collection for early detection, particularly in settings with limited access to mental health services. Once again, non-traditional data is able to supplement conventional data sources, in this case by enabling continuous, behaviourally grounded observation rather than episodic self-briefed and clinic-based measurements.

3. Environmental Exposure and Respiratory Health: Researchers have used satellite imagery and social media data to map pollen exposure and allergy symptoms across regions, linking environmental conditions to briefed health outcomes. Because traditional health datasets often lack fine-grained, real-time measures of environmental exposure, these approaches enable continuous (and sometimes real-time) monitoring of environmental risks and their health impacts. These approaches illustrate how non-traditional and environmental data can provide more accurate situational awareness for public health authorities in the context of climate change and air quality risks.

4. Mobility Data and Pandemic Preparedness: During and after the COVID-19 pandemic, telecommunications and mobility data were used to develop epidemiological models to track and predict the spread of the disease. These studies demonstrated the value of such data for outbreak preparedness, for example by revealing how disease spreads through social and spatial networks rather than through clinical case counts alone. At the same time, these studies also revealed existing—and ongoing—governance gaps related to access, linkage, and public trust in the way data is collected, used, and reused.

Taken together, these examples provide powerful evidence that the scientific value of non-traditional data is no longer merely speculative. As noted above, the challenges remain institutionalising access, governance, and legitimacy so that such evidence can be generated systematically rather than through exceptional arrangements. We discuss some of these challenges below, followed by a positive roadmap for action for European policymakers.

3. CHALLENGES: FROM FRAGMENTATION TO TRUST AND SUSTAINABILITY

Despite demonstrated value, the adoption and implementation of non-traditional data has lagged the empirical evidence. A number of structural barriers—technical, legal, socio-cultural—currently stand in the way of systematically integrating non-traditional data into Europe's research and policymaking infrastructure.

In our research, we identified four interrelated barriers that played a significant role:

1. First, access to high-value non-traditional data remains fragmented and concentrated in a small number of institutions and countries. Most projects rely on bespoke agreements with private data holders that are time-limited and non-transferable.

2. Second, data linkage pathways are technically and legally complex. Integrating non-traditional data with clinical, environmental, demographic, and administrative datasets requires harmonised identifiers, interoperable standards, and secure infrastructures that are not yet widely available.

3. Third, public legitimacy remains fragile. Communities often lack visibility into how data about them is collected and reused, especially when commercial platforms are involved. Legal compliance alone has proven insufficient to secure trust, particularly in sensitive domains such as health and food systems. Without credible mechanisms for transparency, participation, and accountability, large-scale data reuse risks public resistance and political backlash; these represent significant barriers to adoption, especially in democratic states.

4. Finally, funding models privilege short-term experimentation over long-term infrastructure and governance. Many promising initiatives remain pilots without pathways to scale or sustainability. This reinforces a cycle in which data and data infrastructure are treated as project-specific costs rather than as shared, durable public goods that are deeply embedded into the fabric of democratic and social life.

These challenges are real, and if they remain unaddressed Europe risks reproducing a landscape of "innovation without integration," in which promising data-driven experiments proliferate but remain fragmented and short-lived. At the same time, the challenges are highly addressable, especially if governance, stewardship, and participation are treated as core components of data spaces rather than optional add-ons. We outline some concrete steps for action in what follows.

4. EUROPE'S STRATEGIC OPPORTUNITY: INTEGRATED POLICY ACTIONS FOR DECISION-MAKERS

In fact, Europe is uniquely positioned to address the challenges outlined above. Its regulatory frameworks emphasise fundamental rights and proportionality, its research system operates at continental scale, and it has already invested in the foundations of a data ecosystem through EOSC, EHDS, and Horizon Europe. Together, these assets provide the institutional, legal, and technical foundations for a responsible and effective approach to data-driven decision-making in health and well-being.

However, one important shortcoming is that these initiatives currently operate largely in parallel. EOSC focuses on research data interoperability, EHDS on clinical and administrative health data, and Horizon Europe on project-based innovation. While each of these is valuable on its own, a more integrated strategy is required. Such a strategy would explicitly recognise non-traditional data as core assets for health and well-being research, and embed governance, stewardship, and participation into Europe's data spaces.

Building on the empirical evidence and challenges identified above, six interlinked policy actions are proposed. These would allow Europe to move from fragmented experimentation with non-traditional data toward a coordinated system for decision intelligence and anticipatory/proactive policymaking in health and well-being—and in other social domains as well.

1. Integrating Non-Traditional and Environmental Data into EOSC: EOSC should be formally mandated to support the discovery and governed reuse of non-traditional and environmental data relevant to health and well-being, including mobility data, consumption data, satellite imagery, and sensor networks. Common metadata standards and ontologies linking health, environmental, and agricultural data should be developed, alongside federated access models and privacy-enhancing technologies such as trusted research environments and synthetic data services. Together, these measures would enable interdisciplinary research while reducing the concentration of capabilities in a small number of institutions.
2. Broadening the European Health Data Space to Include Determinants of Health: EHDS implementation should explicitly recognise that determinants of health extend beyond clinical and administrative data to include environmental exposure, mobility patterns, and consumption behaviours. Targeted pilot programmes focused on climate–health and food–health linkages would provide concrete pathways for integration, supported by technical interfaces between EHDS and EOSC to enable secure linkage of datasets across sectors and use cases.

3. Using Horizon Europe as a Leverage Mechanism for Sustainable Data Reuse: Horizon Europe should require funded projects in health to demonstrate how data access and reuse will be sustained beyond the life of individual grants and how resulting datasets and tools will align with EOSC and EHDS infrastructures. Evaluation criteria should include interoperability, stewardship arrangements, and societal benefit. Clear incentives should also be provided for partnerships with private data holders, under transparent governance frameworks.

4. Leveraging Article 40 of the Digital Services Act (DSA) for Research Access: Europe should explore stronger alignment between health and research data spaces and emerging access mechanisms under Article 40 of the Digital Services Act (DSA), which enables vetted researchers to request access to platform data for the study of systemic risks. While Article 40 was not designed specifically for health research, many platform-generated datasets - including behavioural signals, mobility patterns, and information exposure dynamics - are highly relevant to understanding determinants of health and wellbeing. In addition, the infrastructure for vetting researchers could also be leveraged for other data re-use cases.

5. Institutionalising Data Stewardship: Recent research highlights the importance of data stewards, intermediary roles capable of bridging legal, technical, ethical, and scientific domains. Europe should professionalise data stewardship through dedicated fellowships, regional hubs linked to EOSC and EHDS, and common curricula and certification frameworks. In health and agriculture, data stewards would align data use with public interest objectives and societal values, not merely regulatory compliance.

6. Embedding Social License and Participation in Data Space Governance: Public legitimacy must be treated as an operational requirement. Europe should pilot social licence frameworks that allow communities to define acceptable uses of their data, integrate citizen panels into EOSC and EHDS governance structures, and require large-scale data reuse projects to demonstrate mechanisms for engagement and transparency.

Each of these actions is important on its own, but together, they are mutually reinforcing. Without stewardship, a vetted research infrastructure and participation, access will lack legitimacy; without EOSC and EHDS integration, reuse will remain fragmented; without Horizon Europe leverage, sustainability will remain elusive. Considered in an integrated manner, this agenda offers Europe a realistic pathway to turn its existing data assets and data investments into durable public capacity for health and well-being--and, generally, more informed and evidence-based policymaking.

CONCLUSION: REALISING THE R&D POTENTIAL OF NON-TRADITIONAL DATA

Non-traditional data represents a largely untapped R&D resource for Europe's health and well-being systems. At a high level, non-traditional data expands the frontier from treatment toward prevention and resilience. It enables earlier identification of mental health distress, links air quality and climate exposure to respiratory outcomes, and reveals nutritional and access inequalities that remain invisible in medical records alone. Together, the gathering empirical base points to a new generation of interdisciplinary R&D that integrates biomedical, environmental, and social evidence.

The strategic importance of this shift lies not only in access to more data, but in the capacity to ask new research questions. Non-traditional data allows European researchers to move from static indicators to dynamic models of risk, behaviour, and vulnerability. It supports experimentation with privacy-preserving analytics, federated learning, and synthetic data generation, thereby advancing methodological innovation alongside substantive discovery. In this sense, non-traditional data are not simply inputs to existing research paradigms but drivers of new scientific capabilities.

Realising this potential requires that Europe's data spaces become part of a stack for responsible data re-use. In particular, we have argued that aligning EOSC, EHDS, and Horizon Europe can transform episodic experimentation into a durable R&D ecosystem in which non-traditional data is discoverable, linkable, and governed for public interest purposes. Data stewardship and participatory governance are essential to ensure that this expansion of research capacity remains legitimate and socially anchored.

Health and well-being offer a particularly compelling test case for this transformation. They are domains in which environmental change, social behaviour, and biological outcomes are inseparable, and where the returns to integrated data-driven research are high. By embedding non-traditional data into its R&D strategy for these sectors, Europe can accelerate innovation in prevention, sustainability, and resilience while demonstrating that digital infrastructures can serve collective well-being.

The central opportunity is therefore not merely to build larger data spaces, but to mobilise them for discovery. If Europe succeeds in integrating non-traditional data into its research architecture for health and well-being, it will strengthen its scientific leadership and provide a model for how data governance can enable, rather than constrain, public-value innovation.

Centre for European Policy Studies
Place du Congrès 1
1000-B Brussels
Belgium