5. Novel and non-traditional data for the study of sustainability transitions

Davide Beraldo, Nora Svensson Hahr, Stefania Milan, Lisa Braito, Veronica Ballerini, Chiara Bocci andEmilia Rocco

5.1 INTRODUCTION

Understanding sustainability transitions requires moving beyond traditional forms of analysis and data collection. As societies work toward more sustainable ways of living and governing, the forces driving these transformations are diverse and dynamic, unfolding across environmental, social, economic, and institutional domains. From a political economy perspective (see Chapter 3), capturing such processes demands data that can reflect change as it happens and make visible the interdependencies shaping transition pathways.

This chapter examines the opportunities and challenges of using novel, non-traditional data sources to study sustainability transitions. Such data, sometimes described as non-conventional or alternative, refer to information not originally produced for research but repurposed to generate insights into socio-economic or environmental dynamics. Examples include satellite imagery, social media traces, search engine queries, financial transaction data, and citizen-generated information. What distinguishes these sources is their origin outside traditional research or policy infrastructures and their ability to offer granular, context-rich perspectives on sustainability issues. They enable observation of complex systems and emergent dynamics that conventional sources such as surveys, statistics, or administrative records struggle to capture due to their static, fragmented nature. Yet, these innovative approaches also introduce methodological, ethical, and governance challenges.

The rising use of such data in sustainability research is both an opportunity and a necessity. Produced at high frequency and often in real-time, they reveal fast-changing processes and links between environmental pressures, behaviour, and institutional responses. These non-traditional data sources can also enable more participatory and responsive research, with citizens, institutions, and private actors jointly generating and interpreting data. For instance, satellite data can track ecosystem changes, while search data show shifts in public concern about sustainability.

Non-traditional data also play a crucial role in localising sustainability analyses. Many transition-related phenomena (think of climate change impact, energy poverty, pollution patterns, or exposure to environmental risks) are inherently place-based and unfold unevenly across regions and sub-regions. The increasing availability of geolocated non-traditional data allows researchers to capture these spatial dynamics with greater granularity, offering insights that remain invisible in national-level aggregates. Social media traces and remote sensing data, including satellite imagery, are especially valuable for generating fine-grained assessments of local vulnerabilities and capacities. Such localized evidence is essential for designing targeted, context-sensitive policies that address the specific needs of communities and territorial ecosystems.

Beyond their spatial granularity, non-traditional data strengthen the SPES analytical framework by illuminating dimensions that remain largely invisible in standard sustainability indicators such as the Sustainable Development Goals (SDG). Conventional metrics prioritize cross-country comparability and national averages, often obscuring local dynamics and the lived realities of transition processes. Non-conventional data help fill these gaps by capturing phenomena such as participation, empowerment, social cohesion, and peace; understood as human areas of security insufficiently measured by traditional survey-based or administrative statistics. Although cross-country comparisons can be difficult with such data, they are essential for within-country analyses that reveal territorial inequalities, community-level vulnerabilities, and differentiated rates of change. Combining innovative data sources with high-quality traditional measures is therefore crucial for developing a more comprehensive and context-sensitive understanding of sustainability transitions.

However, non-traditional data are not without significant challenges. First, many valuable datasets are privately controlled instead of being available in the public domain, raising concerns about unequal access, opacity, dependence on proprietary infrastructures and, consequently, power concentration. Second, ethical and legal questions also arise around privacy, consent, and data protection, particularly when data are derived from individual behaviours or digital platforms. Third, methodological challenges also persist, especially with regards to harmonization and integration with traditional sources; issues of reliability, representativeness, and interoperability remain central. Fourth, tracing methodological innovation in this area remains difficult, in part because it develops at the intersection of diverse disciplinary traditions, including data science, sustainability studies, digital humanities, and the social sciences. For these reasons, non-conventional data should not be viewed as replacements for conventional sources but *as complementary tools* that enrich the analytical and interpretative capacity of sustainability research.

This chapter explores the possibilities offered by novel, non-traditional data sources, referred to here interchangeably as non-conventional, alternative, innovative, or novel, and innovative methods for the study of sustainability transitions. The chapter is structured as follows. The next section maps the deployment of non-traditional data across multiple dimensions of sustainability, illustrating how different types of data are used to study environmental, social, and economic transformations and outlining the exploratory methodology developed for this purpose. Section 5.3 presents a case study based on Google Trends data, tracing public attention to sustainability across European countries. Section 5.4 examines approaches to integrating traditional and innovative data sources, with particular attention to validation. Section 5.5 reflects on the conditions and challenges surrounding data governance in sustainability research. The chapter concludes by advancing policy recommendations emerging from this work.

5.2 MAPPING NON-CONVENTIONAL DATA ACROSS THE DIMENSIONS OF SUSTAINABILITY

Understanding sustainability as a multifaceted phenomenon requires a pluralistic approach that integrates multiple perspectives, analytical frameworks, data sources, and methods. The increased availability of non-conventional data sources has opened new possibilities for analysis. However, tracing such methodological innovation is inherently difficult, as it develops at the intersection of diverse disciplinary traditions. To contribute to the debate, we leveraged Large Language Models (LLMs) to map out the novel data sources mentioned in a large sample of academic texts. Our analysis identifies the types of innovative data categories that have been employed to examine different dimensions of sustainability.

5.2.1 Using Large Language Models to Map Non-Conventional Data

Broadly speaking, our mapping of novel data sources across disciplines consisted of a) producing an expansive operationalization of the concept of sustainability to serve as a conceptual fishing net for capturing diverse data sources across disciplines; b) querying Google Scholar to produce a corpus of relevant academic sources; and c) using LLMs to extract and classify novel data sources. In this section, we introduce our exploratory methodology, omitting the technical details which can be found in Ballerini et al. (2024).

First, we produced an expansive, rather than exhaustive, operationalization of the notion of sustainability. Starting from the five pillars at the centre of the SPES framework, namely productivity, equity for all, sustainability, participation and empowerment, and human security (see Chapter 2; Biggeri et al., 2023), we associated three dimensions of sustainability with each pillar, including three cross-pillar dimensions. This resulted in a novel list of 23 dimensions: Environmental Sustainability; Climate Change; Biodiversity; Pollution; Productivity; Technological Innovation; Economic Growth; Labour Market; Equity; Poverty; Welfare; Inclusivity; Empowerment; Collective Action; Democratic Engagement; Social Cohesion; Human Security; Safety; Solidarity; Judicial System; Economy Cross; Energy; Research and Innovation. We then expanded each dimension into a corresponding set of keywords, arriving at 88 search terms overall. Whereas our operationalization does not have the ambition to cover all the semantic and empirical dimensions of sustainability, it provided us with a net which was able to collect a broad range of academic sources across fields and disciplines.

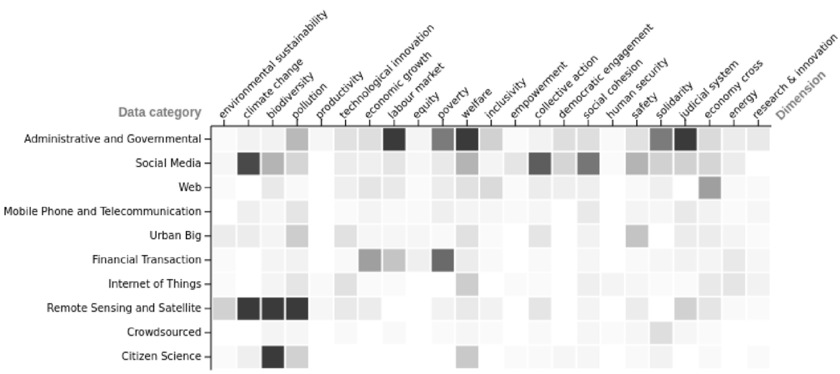
This list of keywords was subsequently used to query Google Scholar. Each query consisted of one of the keywords and several variations around the notion of non-conventional data (e.g., non-conventional data, novel data sources, or innovative measurement). After programmatically executing the queries, we collected up to 50 of the most relevant results, automatically downloading open-access sources, and manually downloading non-open access sources. After deduplication, removal of false-positives, and plain-text extraction, we obtained a final corpus of 882 texts. The corpus consisted of a wide range of examples of how non-conventional data are leveraged to study sustainability, such as the use of wearable sensors to model energy supply and demand (Happle, 2020), the integration of Earth Observation data in crop forecasting (Yli-Heikkilä et al., 2021), and the repurposing of Flickr images to map cultural ecosystem services (Mouttaki et al., 2022).

Using OpenAI's GPT-3.5-turbo via the Application Programming Interface (API), we programmatically extracted from each text the snippets referring to the data source and its associated phenomenon. This process produced 2,945 entries, each pairing a data source with the phenomenon it relates to. We then used GPT to classify each entry. Our coding scheme was based on a consolidation of the type of data mentioned in relevant publications (Del Río Castro et al., 2021; Del Vecchio et al., 2018; ElMassah & Mohieldin, 2020; Fritz et al., 2019; Ilieva & McPhearson, 2018; Kharrazi et al., 2016; van den Brakel et al., 2019; Weber et al., 2021). We instructed the LLM to classify text snippets that referred to data explicitly described as innovative in the articles, according to the categories outlined below.

5.2.2 Types of Data and Dimensions of Sustainability

The classification task produced the following distribution of data categories: Administrative and Governmental Data (362); Social Media Data (288); Remote Sensing and Satellite Data (248); Financial Transaction Data (131); Citizen Science Data (122); Web Data (120); Urban Big Data (115); Internet of Things Data (82); Mobile Phone and Telecommunication Data (77); Crowdsourced Data (40). Figure 5.1 shows how often different types of data have been used to examine each dimension of sustainability.

Administrative and Governmental data have been used in innovative ways, especially in the study of sociopolitical aspects of sustainability, such as the labour market, welfare, and the judicial system. Conversely, Social Media data are often adopted to study issues related to climate change and biodiversity, as well as societal issues such as collective action and social cohesion. Unsurprisingly, Financial Transaction data find most application in the study of poverty, economic growth, and the labour market. Remote Sensing and Satellite data are mainly used to study aspects of sustainability such as climate change, biodiversity, and pollution. Citizen Science data are mobilized to study issues as varied as biodiversity, pollution, and welfare. To demonstrate how novel data sources can be operationalization in practice, the next section presents a case study based on search data from Google Trends and its use in tracing public attention to sustainability.



Note: The darker the cell, the more frequently that type of data category has been used to study the corresponding dimension of sustainability.
Source: Ballerini et al. (2024).

Figure 5.1 Co-occurrence between innovative data categories and sustainability dimensions

5.3 CASE STUDY: GOOGLE TRENDS AND PATTERNS OF PUBLIC ATTENTION

In this section, we present the results of a study using Google Trends data, one of the data sources identified in the mapping above, which is particularly useful for analysing the dynamics and components of public attention to sustainability. The service operated by Google provides insights into the popularity of topics at specific times and locations by analysing the relative volume of searches on its search engine. Google Trends data have mostly been applied in health-related and epidemic studies (Brunori et al., 2022; Carneiro & Mylonakis, 2009; Ginsberg et al., 2009). The extent to which public opinion directs attention towards sustainability, and the specific issues that attract this attention, represents a crucial driver of sustainability transitions. Google Trends data have mostly been applied in health-related and epidemic studies (Brunori et al., 2022; Carneiro & Mylonakis, 2009; Ginsberg et al., 2009). However, we argue that they also offer valuable potential for monitoring public attention dynamics related to sustainability.

Public attention towards an issue has traditionally been proxied through news media coverage (see, e.g., Baumgartner & Jones, 1991; Downs, 1991; Wollin, 1999). For example, Holt and Barkemeyer (2012) documented the increasing coverage of sustainability-related topics in European newspapers between 1990 and 2008, reflecting growing global awareness of environmental and social issues. Likewise, Hase et al. (2021) observed strong media attention to climate change in European countries between 2006 and 2018. However, media coverage represents a relatively indirect and top-down approach to measuring public attention dynamics. By contrast, Google Trends provides real-time data on Google's users online search behaviour, offering a more direct proxy of the bottom-up dynamics of public attention, both its quantitative evolution and its qualitative composition.

In our study, Google Trends data allowed us to operationalize and measure the dynamics of the public attention towards sustainability between 2013 and 2023, across 27 countries. We observed a notable growth in public attention, as well as a remarkable regional gap largely reproducing the historical Western vs. Eastern Europe cleavage. Moreover, we were able to reconstruct which topics are most closely associated with sustainability in the same timeframe in the 27 countries. The next section provides additional details about this methodological approach and the findings it yielded.

5.3.1 Google Trends through Application Programming Interface

Google Trends data are accessible through a web interface.[1] Three main classes of information are available: the quantitative evolution of the relative interest in an issue (InterestOverTime), the distribution of this index across (sub-) regions (InterestByRegions), and the topics and queries most frequently associated with a given search term (RelatedTopics and RelatedQueries). Relative interest represents the proportion of Google searches related to a given issue out of the total volume of searches.

One major limitation of the publicly accessible web interface is that relative interest is only expressed as a score normalized on a 1–100 scale within a specific combination of search terms, regions, and timeframes. While this enables ad hoc comparisons, it prevents more systematic aggregation, such as computing an EU-wide relative score.

To overcome this limitation and enable large-scale, programmatic data collection, we requested access to Google's Private Trends API. This approach allowed us to collect a non-normalized version of the InterestOverTime index[2] and to automate the retrieval of RelatedTopics across different timeframes and regions.

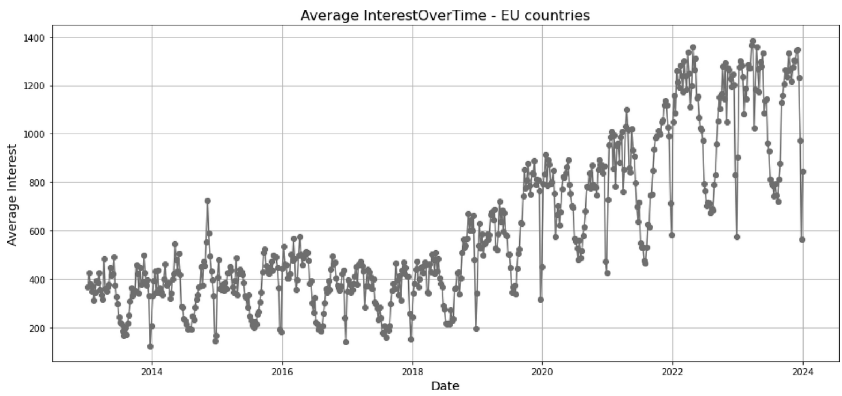
5.3.2 Temporal and Geographical Distribution of Attention Towards Sustainability

For each of the 27 EU countries, we retrieved the non-normalized InterestOverTime values for each week between 2013 and 2023. This dataset, based on the relative volume of Google searches, enables an analysis of public attention to sustainability, including its temporal dynamics and cross-country distribution.

The line graph in Figure 5.2 shows that, aside from monthly and seasonal fluctuations, the relative importance of Google searches related to sustainability has steadily increased between 2018 and 2021, after remaining almost invariant between 2013 and 2017, and seemingly plateauing after 2022. Further inspection of country-specific trends shows that this pattern is replicated in most EU countries, except for Finland, France, and the Netherlands, where the InterestOverTime dips between 2016 and 2018.

The distribution of the InterestOverTime across countries illuminates how attention towards sustainability is unevenly spread within the EU. Figure 5.3 plots the average values per country.

This distribution reveals a clear divide between Western and Eastern European countries, with nations such as the Netherlands, Denmark, and Portugal exhibiting search volumes an order of magnitude higher than countries like Slovakia, Poland, and Estonia.



Note: The values on the vertical axis represent the average across countries of the relative volume of Google searches associated with sustainability.
Source: Beraldo and Svensson (2025).

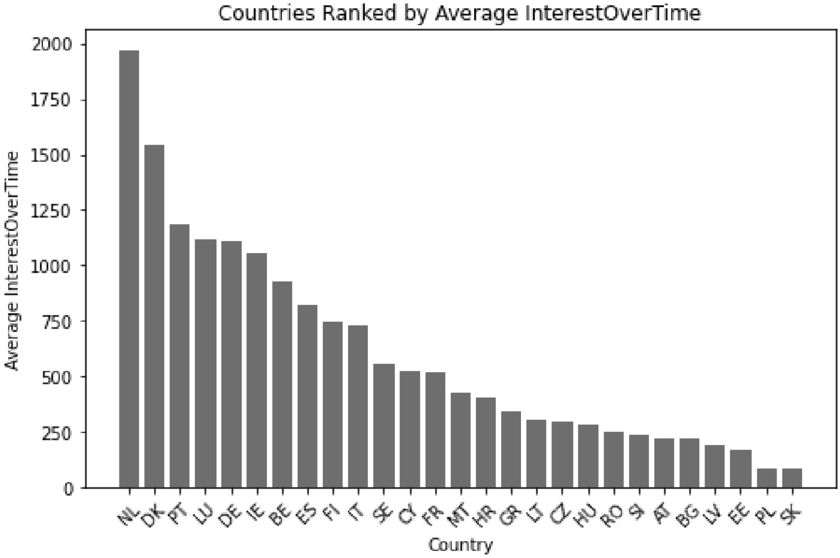
Figure 5.2 Evolution of global Google Search interest (InterestOverTime) in sustainability, 2013–2023

5.3.3 Bottom-up Associations Towards Sustainability

Google Trends reveal not only the quantitative dynamic of search behaviour but also which topics or specific queries most often appear alongside a given term. This helps identify the issues people commonly associate with sustainability. Our analysis tracks how the 15 topics most frequently co-occurring with "sustainability" in EU countries evolved between 2013 and 2023. Topics such as Sustainable Development, Tourism, Economy, Energy, and Environment show steady interest over time. Other, like Report, Sustainable Development Goal, Strategy, and Fashion, spike at specific moments. By contrast, topics such as Agriculture, Project, and Definition display a declining trend.

But what priorities emerge across EU countries? Figure 5.4 addresses this by showing the most frequently co-occurring topics across countries. The list reflects the aggregation of each country's top 10 topics.

We can observe how some topics have more of a generalist character, being highly relevant in many countries, while others seem to be more country-specific. Predictably, Sustainable Development is a topic consistently associated with sustainability in every EU country. Interest in Environment is widespread across countries, though it appears particularly strong in Germany and Italy. Energy is also a prominent concern in most EU countries, especially Denmark and Ireland, except for Luxembourg, Slovakia, Malta, and the Baltic states. By



Note: This graph represents the distribution of InterestOverTime for the topic Sustainability across EU countries, averaged over the timeframe 2013–2023.
Source: Beraldo and Svensson (2025).

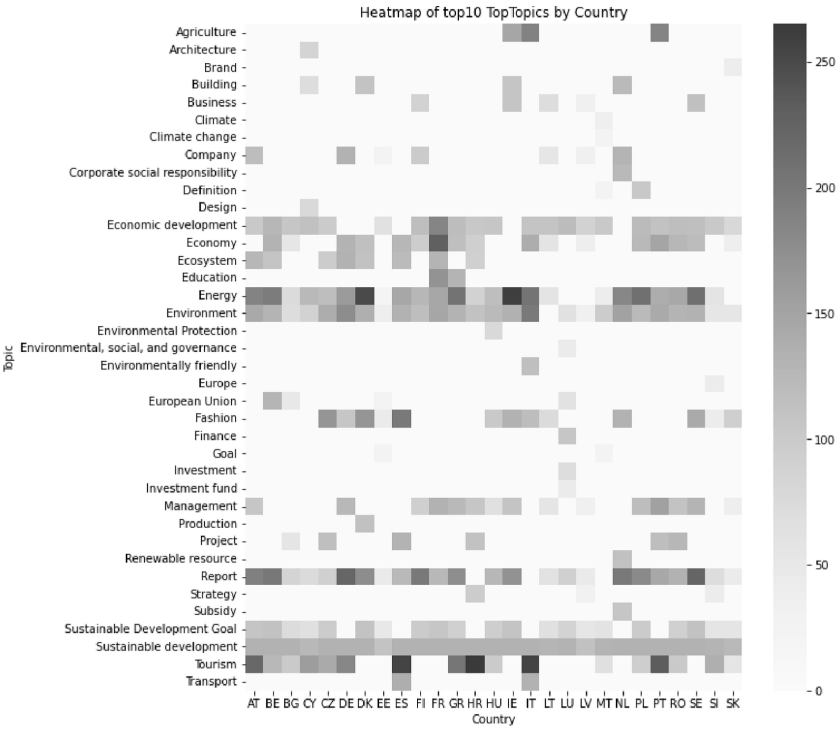
Figure 5.3 Distribution of Google Search interest in sustainability across EU countries (2013–2023)

contrast, topics such as Agriculture appear prominently in only a few countries, namely Italy, Portugal, and Ireland. Likewise, Transport ranks highly in Spain and Italy but is far less salient across the rest of the EU.

5.3.4 Potential and Limitations of Google Trends Data

These findings underscore that public concern for sustainability is not a monolithic phenomenon but a multifaceted one, with its meanings and emphases shifting across time and place. Google Trends data have the potential to accommodate this complexity by aggregating, quantifying, and classifying the spontaneous associations towards the keyword sustainability that results from everyday search practices. This methodological framework could be integrated into policymaking to monitor societal trends, to evaluate policy effectiveness, and to develop sector-specific policies.

Google Trends data also come with important limitations, some of which are common to other types of non-conventional data. First, even the non-normalized



Note: This heatmap shows how often each TopTopic appears in association with each EU country in 2013–2023 (darker colours indicate more frequent associations).
Source: Beraldo and Svensson (2025).

Figure 5.4 Frequency of country–TopTopic associations in the EU (2013–2023)

version of InterestOverTime reflects the relative prominence of a query within the total number of searches, rather than absolute search volumes. This implies that factors such as population size and internet penetration, while not affecting the index mathematically, could still affect its substantial interpretation. Second, another major limitation concerns the classification of queries into topics – used both to compute InterestOverTime and to retrieve related topics – is the outcome of undisclosed, black-boxed machine learning procedures that are prone to inconsistencies and may change over time without notice. Finally, operationalizing a broad concept such as public attention to an issue through web searches raises not only questions of statistical representativeness but also – perhaps more importantly in this context – issues of validity: search behaviour may be driven by diverse, context-specific motives, complicating its interpretation as an indicator of a single phenomenon. This underscores the need to view the relationship between non-traditional and traditional data in terms of integration rather than substitution, a perspective developed in the next section.

5.4 INTEGRATING TRADITIONAL AND INNOVATIVE DATA SOURCES AND ITS VALIDATION

Throughout this chapter, we have demonstrated how the proliferation of non-conventional data sources marks a turning point in the study of sustainability transitions, enabling researchers to capture complex and near-real-time phenomena that were previously difficult to observe using traditional data collection approaches. However, the promise of these innovative data sources must be evaluated critically. As noted earlier, non-traditional data typically do not guarantee statistical representativeness and therefore limit the generalizability of results. Leveraging the full potential of these data sources requires robust approaches to validation, often achieved by calibration, or integration with conventional, survey-based, or official statistics.

Since the mid-twentieth century, probabilistic sample surveys,[3] conducted primarily by national statistical agencies, have been the gold standard for gathering high-quality, representative, and regularly updated information on populations and key phenomena. Survey data underpin public policy, academic research, and business decisions by providing reliable and comparable figures over time. Nowadays, the field faces challenges such as declining response rates, rising costs, and the growing demand for more timeliness and granularity; however, survey data retain their importance as trustworthy benchmarks. In this context, traditional, high-quality survey and census data play a pivotal role as reference points to calibrate and validate findings derived from innovative sources.

Validation of innovative data sources can take several forms. For instance, if web searches or social media mentions are hypothesized to reflect public attitudes or behaviours regarding sustainability, probabilistic survey data on related attitudes or reported behaviours can serve as a yardstick to test the strength, direction, and stability of those associations. Similarly, administrative or observational data – such as energy consumption records or pollution sensor readings – might be compared with probabilistic survey-reported behaviours to assess measurement validity and reveal systematic biases. In more technical domains, satellite or sensor data used to estimate environmental phenomena can be compared against established field survey or census data. For instance, by integrating sample survey results with satellite-based night-time illumination data, it is possible to estimate population distribution across inhabited areas (Edochie et al., 2025).

Such validation can reveal both the strengths and the limitations of innovative data sources. For example, while Google Trends or social media data can supply granular, real-time signals of emerging topics in sustainability, their interpretation often benefits from triangulation against representative attitudinal surveys or longitudinal datasets. This process may help clarify the extent to which spikes in online interest reflect changes in actual knowledge, concern, or behaviour within the broader population (European Union, 2022).

5.4.1 Data Integration for Improved Measurement

Integrating traditional and innovative data sources holds the promise of improved statistical power and new analytical possibilities. Advanced methods such as record linkage, statistical matching, and model-based data fusion can be leveraged to this aim (for detailed method descriptions and definitions, see Ballerini et al., 2024). For example, the longstanding practice of linking household surveys with administrative registers (such as income or education records, as in the EU statistics on income and living conditions – EU-SILC framework) enables more precise measurement of social dynamics such as mobility or poverty. Similarly, the integration of the national enterprises' census, or ad hoc surveys, with information from web scraping data, where text analysis is performed on the text on their website, can be useful to analyse enterprises and their innovation and sustainable practices (see, e.g., van den Brakel et al., 2019). When a probability survey is available for the target population of enterprises (e.g., a survey on start-ups), one can use methods to combine this information with the data arising from web scraping to make inferences about the determinants of innovation for this company. Finally, by leveraging probabilistic municipal-level surveys, such as the Italian urban environmental data survey, which collects information on urban green areas, waste production, and the presence and activity of waste-disposal plants, researchers can construct reliable and representative environmental indicators (UNOOSA, 2018). These survey-based measures can be integrated with widely available non-traditional data sources, including satellite imagery providing land-used information (e.g., the CLMS project[4]) and daily pollution data generated by sensor networks (e.g., the CAMS project[5]). This combination enables more comprehensive and fine-grained assessments of local environmental conditions. Further examples of the use of satellite imagery and remote sensing data can be found in Anderson et al. (2017), Donaldson and Storeygard (2016), and Paganini et al. (2018).

Yet, effective data integration presents significant methodological and operational challenges. The development of new statistical methods for integrating probabilistic and non-probabilistic data – and their dissemination beyond specialist communities – remains an urgent frontier. Given the diversity of available data and analytical purposes, the selection of integration and validation strategies must be informed by the specific research question, available datasets, and intended uses. Flowcharts and decision aids can guide researchers in selecting appropriate methods depending on the structure, quality, and compatibility of their data sources (see Ballerini et al., 2024, p. 78 for an example).

As a final note, integrating diverse data sources requires not only technical solutions for harmonization, comparability, and representativeness, but also careful attention to privacy, data protection, and ethical governance. These issues are central to developing an ethical framework for data use in sustainability research. The next section reflects on these considerations.

5.5 TOWARDS AN ETHICAL DATA GOVERNANCE

To fully leverage the potential of novel data sources for studying sustainability transitions, the governance of data in sustainability research requires careful attention. Our analysis highlights four central themes: the opacity of data infrastructures, the need for robust ethical data governance, the emerging role of data donation frameworks, and the question of what counts as data in the first place.

The first theme relates to data opacity and the challenges posed by black-boxed infrastructures, which is the preferred approach of most stakeholders and systems in the contemporary tech industry (cf. Pasquale, 2015; Willim, 2024). Working with innovative data sources such as Google Trends exposes the profound opacity of today's data infrastructures. Many are privately owned, poorly documented, and shaped by algorithmic processes that remain beyond public scrutiny (Seaver, 2017). For instance, Google provides only limited public documentation on how Google Trends data are generated and processed, which often leaves researchers inferring key parameters in order to interpret results (see also Springer et al., 2025). This illustrates a wider imbalance in the data ecosystem: while public institutions and researchers depend on privately controlled infrastructures, the design choices and biases embedded in these systems are rarely transparent. Such opacity undermines core principles of reproducibility, accountability, and trust in scientific inquiry, making the call for ethical data governance both methodological and political. Some go as far as to call for public data commons as a way to safeguard European values and rights and enhance increasing data sharing (cf. Zygmuntowski et al., 2021).

Second, ethical data governance must move beyond compliance to embed values of transparency, inclusion, participation, and accountability throughout the data lifecycle. It is not limited to protecting personal data but extends to ensuring the integrity, quality, and legitimacy of data used for public-interest research. This approach recognizes that data governance shapes what can be known, whose voices are heard, and how sustainability transitions are managed (Ponti et al., 2024).

A robust ethical framework for data governance should rest on at least three principles.

- Transparency: data providers, especially privately-owned platforms, must disclose collection and processing details, including sampling procedures and algorithmic logics, and ensure stable, auditable access (cf. Helberger et al., 2018).
- Proportionality: data use should be aligned with research aims while minimizing risks of harm, misrepresentation, or surveillance (cf. European Data Protection Supervisor, 2020).
- Reciprocity: data practices should generate public value, supporting shared learning and democratic oversight rather than reinforcing corporate or institutional asymmetries.

Ethical governance also demands methodological integrity; that is, ensuring that novel forms of data are validated and used responsibly. Ethical and methodological considerations are thus deeply intertwined: without transparency and open access, research quality and legitimacy are compromised.

Third, emerging data donation initiatives illustrate how citizens can play an active role in ethical data governance. These frameworks allow individuals to voluntarily share their data for scientific and public-interest purposes under conditions of informed consent, transparency, and collective benefit (Boeschoten et al., 2020; see also Araujo et al., 2022). At the policy level, the EU Digital Services Act (DSA) marks a major step toward institutionalizing responsible data access. It requires very large online platforms to provide vetted researchers with data relevant to systemic risks, including those affecting democracy, public health, and sustainability (Selling et al., 2025). Together, these developments signal a gradual redistribution of data power, fostering a more equitable relationship between citizens, researchers, and digital platforms.

Finally, ethical data governance entails expanding what counts as data. Innovative data are not limited to big data repurposed from platforms but also encompass citizen-generated and qualitative forms of evidence that capture lived experience and local knowledge. These include participatory approaches such as citizen sensing, deliberative mapping, and experimental games (as discussed in Chapter 9 of this book) that engage non-experts and laypersons in data generation and interpretation. Integrating these practices with institutional and traditional data sources requires recognizing the 'legitimacy of people as knowers' (Milan, 2024, p. 112), and the 'right to contribute information considered by appointed institutions' (Berti Suman et al., 2023, p. 4), thus legitimizing non-credentialed forms of knowledge. Pushing the boundaries of current regulatory frameworks is necessary to enable the recognition on an equal footing of data that originates outside formal institutional settings and mechanisms. This recognition has the potential to enrich sustainability monitoring, enhance legitimacy, and ground innovation in social realities rather than abstract algorithmic metrics. Ethical data infrastructures and equitable data governance frameworks are thus not merely technical issues: they are strategic enablers of a just, inclusive, and trustworthy sustainability transition.

5.6 CONCLUSIONS AND RECOMMENDATIONS

This chapter explored how novel, non-traditional data, ranging from social media traces to search queries, financial transactions, and citizen-generated information, can deepen our understanding of sustainability transitions by providing granular, timely, and context-sensitive insights that traditional data often miss. Using LLMs, we mapped the diverse types of innovative data employed across 23 sustainability dimensions, revealing both their analytical potential and the challenges surrounding their use. To demonstrate how non-conventional data can be mobilized to effectively understand sustainability transitions, we deployed Google Trends to show how online search behaviour can uncover temporal, geographical, and thematic patterns in public attention to sustainability within the EU. The chapter also assessed the methodological, ethical, and governance issues inherent to non-conventional data, including questions of integration with traditional data, representativeness, data opacity, and dependence on private digital infrastructures. Ultimately, we argued that integrating innovative and conventional data sources, supported by robust validation and ethical frameworks, is essential for more responsive and equitable sustainability policymaking.

Taken together, these findings underscore several strategic priorities that policymakers must address to effectively leverage non-conventional data for sustainability transitions. The following recommendations build directly on the empirical evidence and conceptual discussions developed throughout the chapter, translating them into actionable guidance for strengthening sustainability measurement, data governance, and policy design. They align with the three policy imperatives identified by Lengfelder et al. (2025, p. 446) regarding the role of AI in fostering human development: invest, inform, and include.

1. Build integrated and inclusive data ecosystems

   Policymakers should support infrastructures that enable integration of survey data, administrative records, geospatial data, digital traces, and citizen-generated inputs, enabling more granular and locally relevant insights. Such integration must also prioritize equity and inclusion to ensure that data-driven policies do not reinforce structural disparities in digital access or representation.

2. Strengthen validation, transparency, and ethical governance

   To address uneven data quality, accessibility gaps, and biases in innovative data, sustainability analyses should embed robust validation frameworks, triangulating non-conventional sources with surveys and administrative benchmarks. Ethical governance must also ensure transparency, accountability, and proportionality through clear documentation of algorithms, sampling, and processing practices.
3. Expand responsible, equitable access to platform and proprietary data for public-interest research

   Meaningful public interest research requires stable, well-documented access to platform data. Regulators should strengthen and expand frameworks such as the DSA to reduce asymmetries between public institutions and private platforms, enabling auditable, reliable, and fair access for researchers working on sustainability challenges.
4. Invest in capacity-building and participatory data practices

   Working with innovative data demands interdisciplinary expertise. Policymakers should support training and cross-sector collaboration to help researchers and public administrations use these sources responsibly. They should also promote participatory and citizen-generated data initiatives (e.g., community sensing) to diversify evidence bases and democratise data production.

In conclusion, while innovative data can greatly enhance our understanding of sustainability transitions, their effective use depends on rigorous methodological standards and ethical data governance. This cannot be separated from the broader technological ecosystem in which such data are generated, processed, and increasingly analysed; one that is now deeply shaped by AI. As AI tools such as LLMs become embedded in data collection, interpretation, and decision-making, they amplify both the opportunities and the risks associated with non-traditional data. These systems are largely controlled by a small number of private companies, whose concentration of technological and data assets can exacerbate inequalities and reduce people to objects of optimization rather than agents of change.

Emerging governance frameworks offer guidance for navigating these challenges. For example, Simoncini and Biggeri (2025) advance five principles – digital loyalty, human-in-the-loop inclusivity, explicability, fairness, and entitlement – that can help steer the responsible use of data and AI. Embedding these principles into data governance can ensure that AI-enhanced analytical tools strengthen human capabilities, democratic agency, and public value rather than undermining them, and contribute to realising the full promise of innovative data for sustainability transitions.

NOTES

1. https://trends.google.com/
2. The hypothesis that the specific version of InterestOverTime available through Google's Private API (accessible via the getTimelineForHealth method) has been empirically validated through ad hoc tests (see Appendix 1 in Ballerini et al., 2024).
3. That is, every unit in target population has known and nonzero probability of being selected into the sample through a random mechanism.
4. https://atmosphere.copernicus.eu/air-quality
5. https://land.copernicus.eu/en

REFERENCES

Anderson, K., Ryan, B., Sonntag, W., Kavvada, A., & Friedl, L. (2017). Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-Spatial Information Science*, *20*(2), 77–96. https://doi.org/10.1080/10095020.2017.1333230

Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Velde, B., de Vreese, C., & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, *4*(2), 372–387. https://doi.org/10.5117/CCR2022.2.001.ARAU

Ballerini, V., Beraldo, D., Bocci, C., Braito, L., Milana, R., & Trans, M. (2024). *Report on mapping, harmonising and integrating novel data sources for research purposes* (SPES Report no. 4.1; SPES Project – Sustainability Performances, Evidence and Scenarios). University of Florence. https://www.sustainabilityperformances.eu/wp-content/uploads/2024/01/WP4-Del-4_1_Pat-I.pdf

Baumgartner, F. R., & Jones, B. D. (1991). Agenda dynamics and policy subsystems. *The Journal of Politics*, *53*(4), 1044–1074. https://doi.org/10.2307/2131866

Beraldo, D., & Svensson, N. (2025). Public attention towards sustainability in the EU: An exploration of Google Trends data (SPES Project – Sustainability Performances, Evidence and Scenarios No. SPES Working Paper no. 4.2). University of Florence. https://www.sustainabilityperformances.eu/wp-content/uploads/2025/03/SPES-D4.2_Public-Attention-Towards-Sustainability-in-the-EU-_layout.pdf

Berti Suman, A., Balestrini, M., Haklay, M., & Schade, S. (2023). When concerned people produce environmental information: A need to re-think existing legal frameworks and governance models? *Citizen Science: Theory and Practice*, *8*(1), 10. https://doi.org/10.5334/cstp.496

Biggeri, M., Ferrannini, A., Lodi, L., Cammeo, J., & Francescutto, A. (2023). The "winds of change": The SPES framework on Sustainable Human Development (No. SPES Working Paper 2.1; SPES Project – Sustainability Performances, Evidence and Scenarios). University of Florence. https://www.sustainabilityperformances

.eu/wp-content/uploads/2023/10/SPES-Working-paper-2.1_29th-September-2023_FINAL.pdf
Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T. B., & Oberski, D. L. (2020). A framework for digital trace data collection through data donation. *CoRR*, *abs/2011.09851*. https://arxiv.org/abs/2011.09851
Brunori, P., Resce, G., & Serlenga, L. (2022). Searching for the peak: Google Trends and the first COVID-19 wave in Italy. *International Journal of Computational Economics and Econometrics*, *12*(4), 445–458. https://doi.org/10.1504/IJCEE.2022.126323
Carneiro, H. A., & Mylonakis, E. (2009). Google Trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*, *49*(10), 1557–1564. https://doi.org/10.1086/630200
Del Río Castro, G., González Fernández, M. C., & Uruburu Colsa, Á. (2021). Unleashing the convergence amid digitalization and sustainability towards pursuing the Sustainable Development Goals (SDGs): A holistic review. *Journal of Cleaner Production*, *280*, 122204. https://doi.org/10.1016/j.jclepro.2020.122204
Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Open innovation and social big data for sustainability: Evidence from the tourism industry. *Sustainability*, *10*(9), 3215. https://doi.org/10.3390/su10093215
Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, *30*(4), 171–198. https://doi.org/10.1257/jep.30.4.171
Downs, A. (1991). Up and down with ecology: The "Issue-Attention Cycle". In D. Protess & M. E. McCombs (Eds.), *Agenda setting: Readings on media, public opinion, and policymaking* (pp. 38–50). Routledge.
Edochie, I., Newhouse, D., Tzavidis, N., Schmid, T., Foster, E., Hernandez, A. L., Ouedraogo, A., Sanoh, A., & Savadogo, A. (2025). Small area estimation of poverty in four West African countries by integrating survey and geospatial data. *Journal of Official Statistics*, *41*(1), 96–124. https://doi.org/10.1177/0282423X241284890
ElMassah, S., & Mohieldin, M. (2020). Digital transformation and localizing the Sustainable Development Goals (SDGs). *Ecological Economics*, *169*, 106490. https://doi.org/10.1016/j.ecolecon.2019.106490
European Data Protection Supervisor (2020). *The EDPS quick-guide to necessity and proportionality.* https://www.edps.europa.eu/sites/default/files/publication/20-01-28_edps_quickguide_en.pdf
European Union (2022). *Data innovation in demography, migration and human mobility.* Publications Office of the European Union. https://doi.org/10.2760/027157
Fritz, S., See, L., Carlson, T., Haklay, M. (Muki), Oliver, J. L., Fraisl, D., Mondardini, R., Brocklehurst, M., Shanley, L. A., Schade, S., Wehn, U., Abrate, T., Anstee, J., Arnold, S., Billot, M., Campbell, J., Espey, J., Gold, M., Hager, G., … West, S. (2019). Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability*, *2*(10), 922–930. https://doi.org/10.1038/s41893-019-0390-3
Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014. https://doi.org/10.1038/nature07634
Happle, G. (2020). *Data-driven occupant presence models for urban building energy simulation* [ETH Zurich]. https://doi.org/https://doi.org/10.3929/ethz-b-000438825
Hase, V., Mahl, D., Schäfer, M. S., & Keller, T. R. (2021). Climate change in news media across the globe: An automated analysis of issue attention and themes in

climate change coverage in 10 countries (2006–2018). *Global Environmental Change*, *70*, 102353. https://doi.org/10.1016/j.gloenvcha.2021.102353

Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, *34*(1), 1–14. https://doi.org/10.1080/01972243.2017.1391913

Holt, D., & Barkemeyer, R. (2012). Media coverage of sustainable development issues – attention cycles or punctuated equilibrium? *Sustainable Development*, *20*(1), 1–17. https://doi.org/10.1002/sd.460

Ilieva, R. T., & McPhearson, T. (2018). Social-media data for urban sustainability. *Nature Sustainability*, *1*(10), 553–565. https://doi.org/10.1038/s41893-018-0153-6

Kharrazi, A., Qin, H., & Zhang, Y. (2016). Urban big data and Sustainable Development Goals: Challenges and opportunities. *Sustainability*, *8*(12), 1293. https://doi.org/10.3390/su8121293

Lengfelder, C., Tapia, H., & Biggeri, M. (2025). Navigating AI with a human development compass – shaping tomorrow's capabilities. *Journal of Human Development and Capabilities*, *26*(3), 439–448. https://doi.org/10.1080/19452829.2025.2520011

Milan, S. (2024). Citizen-generated data within the Green Deal data space. In S. Thabit Gonzalez & G. Maccani (Eds.), *Unlocking Green Deal data: Innovative approaches for data governance and sharing in Europe* (pp. 99–113). Publications Office of the European Union.

Mouttaki, I., Bagdanavičiūtė, I., Maanan, M., Erraiss, M., Rhinane, H., & Maanan, M. (2022). Classifying and mapping cultural ecosystem services using artificial intelligence and social media data. *Wetlands*, *42*(7), 86. https://doi.org/10.1007/s13157-022-01616-9

Paganini, M., Petiteville, I., Ward, S., Dyke, G., Steventon, M., Harry, J., & Kerblat, F. (2018). *Satellite earth observations in support of the Sustainable Development Goals*. European Space Agency. https://eohandbook.com/sdg/files/CEOS_EOHB_2018_SDG.pdf

Pasquale, F. (2015). *The Black Box Society: The secret algorithms that control money and information*. Harvard University Press.

Ponti, M., Portela, M., Pierri, P., Daly, A., Milan, S., et al. (2024). *Unlocking Green Deal data: Innovative approaches for data governance and sharing in Europe*. Publications Office of the European Union.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 205395171773810. https://doi.org/10.1177/2053951717738104

Selling, L. K., Keller, C. I., Ohme, J., Klinger, U., & de Vreese, C. (2025). Data access for researchers under the Digital Services Act: From policy to practice (Weizenbaum Policy Paper No. 14). Weizenbaum Institut, Berlin.

Simoncini, A., & Biggeri, M. (2025). Agency and the role of international regulation: Reconciling AI technologies and human development. *Journal of Human Development and Capabilities*, *26*(3), 449–460. https://doi.org/10.1080/19452829.2025.2517743

Springer, S., Strzelecki, A., Holmes, B., Ziermann-Canabarro, J. M., Kaatz, M., & Zieger, M. (2025). Setting trends with Google: Limits and perspectives when utilising search engine data. *Quality & Quantity*. https://doi.org/10.1007/s11135-025-02431-0

UN Office for Outer Space Affairs (UNOOSA) (2018). *European Global Navigation Satellite System and Copernicus: Supporting the Sustainable Development Goals*

– *Building Blocks Towards the 2030 Agenda*. Office for Outer Space Affairs, United Nations. https://www.unoosa.org/res/oosadoc/data/documents/2018/stspace/stspace71_0_html/st_space_71E.pdf

van den Brakel, J. A., Buelens, B., Curier, R. L., Daas, P., Gootzen, Y., de Jong, T., Puts, M., Tennekes, M., & Willems, R. (2019). *Aspects of existing databases, traditional and non-traditional data sources and collection of good practices* (The Makswell Project). Statistics Netherlands (CBS). https://www.makswell.eu/attached_documents/output_deliverables/deliverable_2.1.pdf

Weber, I., Imran, M., Ofli, F., Mrad, F., Colville, J., Fathallah, M., Chaker, A., & Ahmed, W. S. (2021). Non-traditional data sources: Providing insights into sustainable development. *Communications of the ACM*, *64*(4), 88–95. https://doi.org/10.1145/3447739

Willim, R. (2024). Opacity. In R. Willim (Ed.), *Mundania: How and where technologies are made ordinary* (pp. 80–93). Bristol University Press.

Wollin, A. (1999). Punctuated equilibrium: Reconciling theory of revolutionary and incremental change. *Systems Research and Behavioral Science*, *16*(4), 359–367. https://doi.org/10.1002/(SICI)1099-1743(199907/08)16:4%3C359::AID-SRES253%3E3.0.CO;2-V

Yli-Heikkilä, M., Wittke, S., Kokkinen, M., & Partala, A. (2021). *CropYield—Towards pre-harvest crop yield forecasts with satellite remote sensing. Final report* (No. 80/2021; Natural Resources and Bioecomony Studies). Luke: Natural Resources Institute Finland. https://jukuri.luke.fi/server/api/core/bitstreams/f64ec0ca-d9ed-4237-bfd1-e8506b220320/content

Zygmuntowski, J. J., Zoboli, L., & Nemitz, P. F. (2021). Embedding European values in data governance: A case for public data commons. *Internet Policy Review*, *10*(3). https://doi.org/10.14763/2021.3.1572