methods, data, analyses | Vol. 19(2), 2025, pp. 4-10

Editorial
An Overview of the Special Issue

Rachel Gibson[1] & Trent D. Buskirk[2]

[1] *University of Manchester*
[2] *Old Dominion University*

Across the quantitative social sciences, researchers increasingly face significant challenges and opportunities prompted by the arrival of new sources of very rich, highly granular, and often unstructured digital data. While traditional methods such as surveys and content analysis tools remain indispensable for measuring individual attitudes, behaviors, demographic characteristics, and media messaging online, they often struggle to capture the complex multimodal information streams and metadata generated by social media platforms, mobile devices, sensors, and tracking applications. Collecting and analyzing these diverse new forms of content, dynamic moment-to-moment behaviors, and naturally occurring interactions has become a pressing and exciting research task—one that holds real promise for answering longstanding research questions more completely and, in some cases, more accurately. Yet as access to these data grows, so too do the problems they pose in terms of representativeness, potential new sources of bias, complex preprocessing demands, and reproducibility.

One increasingly common response to these challenges is to link or “anchor” emerging data sources to more structured, researcher-designed forms of data, particularly surveys. Doing so offers two complementary benefits. First, traditional instruments can provide context, validation, and interpretability for new

The Special Issue is generated in part from the Digital Campaigning and Electoral Democracy (DiCED) Project. Grant agreement no. 833177 funded by ”European Research Council (ERC) European Union’s Horizon 2020 research and innovation programme

forms of behavioral or multimodal data. Second, emerging data sources can enrich surveys by extending coverage in time, modality, or behavioral detail, thereby filling gaps that conventional approaches alone may leave unaddressed.

We can frame the connections between the papers in this special issue using the motivating schematic depicted in Figure 1**.** Specifically, each study begins with a substantive research question for which traditional approaches and data sources—such as surveys, curated observational data or designed data, or conventional content analysis—could plausibly be used. However, in each case, the authors identify limitations in relying on these approaches alone, whether due to recall error, restricted temporal resolution, limited measurement scope, or difficulty capturing visual, behavioral, or contextual information. To address these limitations, the papers pursue two broad strategies. Some introduce **new analytical methods applied to existing data**, enabling researchers to model previously unaccounted-for sources of error or extract richer information from conventional inputs. Others incorporate **new or emerging data sources alongside established methods**, using digital traces, images, metadata, or real-time behavioral signals to improve the completeness, accuracy, or interpretability of the resulting analyses. In both cases, methodological innovation is driven not by novelty for its own sake, but by the goal of producing better answers to well-defined research questions.

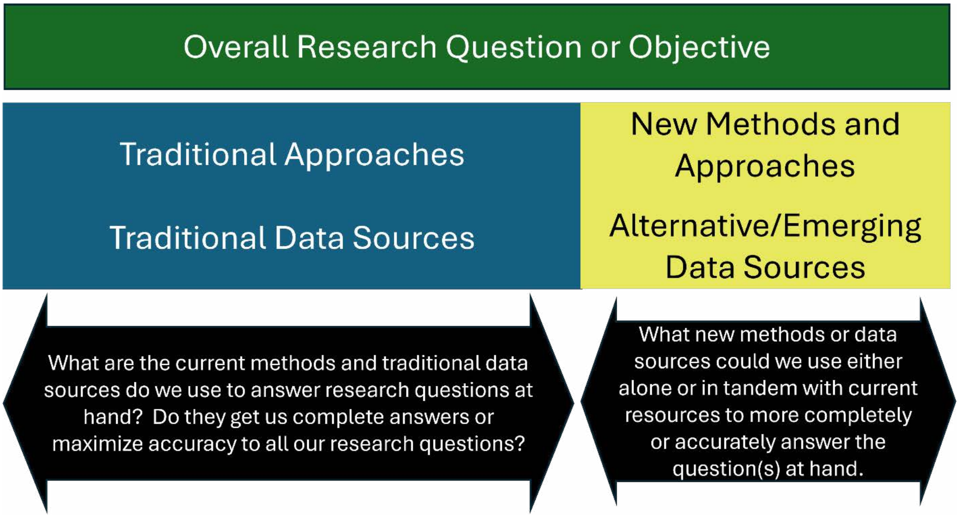
Overall Research Question or Objective
Traditional Approaches
Traditional Data Sources
New Methods and Approaches
Alternative/Emerging Data Sources
What are the current methods and traditional data sources do we use to answer research questions at hand? Do they get us complete answers or maximize accuracy to all our research questions?
What new methods or data sources could we use either alone or in tandem with current resources to more completely or accurately answer the question(s) at hand.

Figure 1 A motivating schematic for thinking about the use of new and emerging data sources in conjunction with more traditional methods and sources.

This special issue brings together a set of papers that advance the field along precisely this shared dimension—**the deliberate extension of traditional quantitative approaches through new data, new methods, or both**—to reach more complete, accurate, or informative conclusions about individual preferences, activities, and attitudes. Each article is accompanied by a "reflective methodological appendix," in which authors are asked to "lift the hood" on their research process and document the choices, constraints, adaptations, and tradeoffs encountered as their projects unfolded. These reflections make visible the iterative decision-making that typically remains hidden in published research and align closely with broader efforts to promote open and reflexive research practices within the social sciences. Importantly, the reflective appendices underscore a core motivation for this volume: emerging data sources come with significant qualifiers. None are genuinely "free." Across projects, researchers invested substantial effort in collection, cleaning, processing, linkage, and interpretation—often operating within constraints set by platform architectures, proprietary systems, device limitations, and finite computational infrastructure. Moreover, because these data depart from long-standing survey norms in structure, stability, and population coverage, authors repeatedly had to interrogate assumptions and adapt designs as the work unfolded. In these appendices we asked authors to go behind the veil of their papers—particularly the methods, data, and analyses components—to reflect on the pathway from inception to final publication. The result is a set of statements that make visible the decisions, routing, and particularly re-routing of research designs that would otherwise go unreported. Through this process, we hope this collection of papers and appendices can offer practical guidance to scholars embarking on similar projects and to contribute to a more transparent public dissection of the iterative dynamics that underpin research using new and emerging data sources.

Below we begin with an overview of the papers as a whole and how they provide alternative approaches to addressing a common challenge of linking established pre-digital data and methods with newer digital-exclusive versions. We also highlight their key findings as stand-alone pieces of research and their substantive value and contribution to their respective fields. We then turn our focus to the reflective appendices to identify the dilemmas and challenges that authors faced in investigating their research questions—and, importantly, how they addressed them in practice. We conclude by drawing out broader lessons learned from these experiences for future scholarship navigating the frontier of linked and augmented data analysis.

Emerging Data Types and New Modes of Observation

What unites the papers in this issue is a shared recognition that emerging data sources both complement and complicate survey-based research. Sometimes these sources can supplement surveys by filling coverage gaps or validating self-reports. In other cases, they introduce entirely new modes of measurement—visual, behavioral, or moment-triggered—that fundamentally reconfigure what social scientists can observe. In still others, they strain existing methodologies, requiring innovations in data processing, linkage, or research design. The five contributions showcased here span methodological innovations—from augmented Data Download Packages and real-time event-triggered surveys to mixture modeling for linkage errors and systematic coding of multimodal political appeals. Collectively, these studies address persistent challenges in data quality, representativeness, and analytical rigor by integrating diverse data streams, including digital trace data, visual content, and survey responses. Common themes include the pursuit of richer, more accurate insights into human behavior and communication, the development of tools to mitigate bias and error, and the expansion of research beyond traditional text-based and retrospective approaches.

The first paper by **Wedel, Ohme, and Araujo**, *Augmenting Data Download Packages—Integrating Data Donations, Video Metadata, and the Multimodal Nature of Audio-visual Content,* introduces Augmented Data Download Packages (aDDPs) as a novel way to enrich conventional digital trace data. By integrating survey responses, metadata, and multimodal content embeddings, aDDPs provide a more comprehensive view of user behavior. Using TikTok as a case study, the authors show how these enhanced packages enable nuanced analyses of engagement patterns and content classification, illustrating the potential of combining behavioral and self-reported data for social science research. Building on the theme of multimodality, the second paper by **Iglesias**, *Preferences, Participation, and Evaluation of Answering Questions About the Books Participants Have at Home Through Conventional and Image-Based Formats,* examines the role of visual data in survey design by comparing photo-based questions to conventional formats. Drawing on a large-scale mobile survey of Spanish parents, the study shows that while respondents generally prefer traditional questions, those who favor images engage more when given choice. Demographic and behavioral predictors of participation underscore the complexity of integrating visual tasks into surveys and highlight the need for adaptive designs that accommodate diverse respondent preferences. The third contribution by **Ochoa**, *Researching the Moment of Truth: An Experiment Comparing In-the-Moment and Conventional Web Surveys to Investigate Online Job Applications,* advances this discussion by exploring real-time, event-triggered surveys linked to metered data. Focusing on online job applications, the authors test whether surveys delivered immediately after detected events can improve data quality and reduce recall bias. The study finds strong acceptance of this approach and richer responses compared to conventional surveys, while also showing that memory-related errors may persist even under improved timing—highlighting both the value and limits of timely interventions for capturing accurate behavioral data.

While these papers focus on enhancing survey-based research through new data and measurement modes, the fourth study by **West, Slawski, and Ben-David**, *Improved Ensemble Predictive Modeling Techniques for Linked Social Media and Survey Data Sets Subject to Mismatch Error,* reminds us that linking data sources together, while expanding the scope of information, is not without error. This paper specifically addresses a critical methodological issue in linked datasets: mismatch errors introduced by probabilistic linkage. The authors propose a mixture modeling approach to adjust predictive modeling outputs when such errors occur, testing it on Twitter activity linked to survey-based ideology measures. Their method successfully recovers predictive performance, underscoring the importance of explicitly correcting linkage uncertainty in an era of increasingly complex data integration.

Finally, the fifth paper by **Cashell**, *Improving Assessments of Group-Based Appeals in Political Campaigns by Systematically Incorporating Visual Components of Ads*, extends the conversation on multimodality by introducing a coding scheme for visual and textual group-based appeals in political campaigns. Applying the schema to thousands of images from U.S. House races, the study reveals that indirect visual cues are as prevalent as direct mentions, often used in combination. By systematically capturing visual indicators, the framework reduces bias in measuring group targeting strategies and provides a more complete picture of modern political communication than using the current text-only approaches alone.

Taken altogether, these studies illustrate a shared focus on methodological innovation and data enrichment. They show how linking and augmenting data sources—whether through multimodal content, real-time triggers, or error-adjusted models—can overcome limitations of traditional approaches and open new avenues for research. As digital environments continue to evolve, the approaches showcased in this special issue chart a path toward more robust, adaptive, and multimodal research designs.

Reflections on the Reflective Appendices

The reflective appendices reveal a set of shared methodological themes that reinforce the central motivation of the special issue: the promise of emerging data comes tightly coupled with practical and inferential challenges that often surface most sharply after the initial research design is on paper. Across the five projects, authors repeatedly emphasize that key constraints emerged during data collection, processing, and integration—through lower-than-anticipated participation, uneven data completeness, technical constraints imposed by platforms or devices, and greater-than-expected demands for infrastructure and manual labor. These experiences reinforce that emerging data sources are rarely “plug-

and-play" alternatives to surveys; instead, they require substantial investment, flexibility, and explicit feasibility assessment as projects unfold. The appendices also underscore that some of the most consequential challenges—such as low uptake in burdensome tasks, unstable triggering conditions, linkage uncertainty, or inconsistencies in human coding—cannot be "fixed" purely in the analysis stage but instead shape what can credibly be claimed and what analyses are ultimately possible.

At the same time, the appendices highlight distinctive challenges that are specific to particular data types and integration strategies. Image-based approaches raise issues of respondent burden, privacy sensitivity, and the labor-intensity of classification, whether manual or automated. Behavioral trace and in-the-moment designs are highly dependent on platform architectures, URL stability, and operating-system capabilities, requiring sustained technical attention and ongoing monitoring of the data-generating process. Data donation work points to platform governance and user verification processes as practical bottlenecks, while also raising the likelihood of self-selection and the need to think carefully about what populations and phenomena can be studied credibly given participation patterns. Linked survey–social media analyses emphasize that even when linkage appears strong, mismatch error and linkage quality remain central threats to inference and must be modeled explicitly rather than treated as a secondary issue. Together, these reflections complement Figure 1 by showing that "adding" new data or methods does not simply expand what can be studied; it also introduces new constraints that must be documented and weighed as part of any fitness-for-use assessment.

Emerging Frameworks for Fitness-for-Use

Several papers demonstrate that emerging data may serve as enhancements rather than replacements for surveys. Visual data enriches communication research; multimodal DDPs broaden analytic possibilities; image-based responses provide detailed, ecologically grounded measurement for otherwise difficult-to-measure household inventories; and triggered surveys align measurement more closely with real-world behavior. Others underscore the need for caution and explicit adjustment when new forms of error are introduced, as in linkage mismatch error and predictive modeling. Across the issue, the resulting stance is pragmatic: emerging data sources neither uniformly surpass nor simply replicate surveys. Their value depends on context, construct, and research design—and on careful evaluation of fitness-for-use relative to the inferential goals at hand.

Developing a Culture of Methodological Reflexivity and Transparency

Beyond the substantive and methodological contributions of the five studies, the special issue advances open scientific practice by promoting “reflexive praxis”—researchers documenting the choices, complications, and adjustments shaping their approach and analysis. Each paper includes a peer-edited appendix offering insights into data access challenges, technical constraints, processing bottlenecks, linkage and measurement vulnerabilities, and the iterative redesign that often accompanies work with emerging sources. These reflections are intended as practical guides and as a step toward making the methodological pathway of digitally enabled social science more transparent and cumulative.

Conclusion

Together, the contributions demonstrate a future in which ‘old’ and emerging data sources and methods of data collection operate in tandem. Structured forms of data collected through surveys and established content analysis tools remain essential for representativeness, comprehensiveness and intentional measurement; emerging digital data provide new contextual and visual richness, temporal precision, and behavioral grounding. Yet integration requires methodological imagination, transparency, and rigor—and an honest accounting of the constraints and trade-offs involved. This special issue aims to contribute to that agenda by showing not only what these approaches can achieve, but also what is required to implement them responsibly and interpret them appropriately.