Journal of Business Research 188 (2025) 115112

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Big data analytics democratized with clean collaboration and customer privacy choice

Koen Pauwels *, Zeynep Aksehirli

Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA

ARTICLE INFO

Keywords:
Big data
Analytics
Privacy
Digital transformation
Marketing Mix Model
Omnichannel

ABSTRACT

Digital technologies and platforms have started a data explosion, and big data analytics is moving beyond social media in retail media online and offline, integrating the seamless interaction between business and customers in omnichannel. These data allow better insights but also require a more holistic analysis to refine customer experience and corporate strategies.

While the business opportunities are exciting, we see three key challenges: organizational (silos), operational (in uncovering customer insights) and societal (respecting privacy). To this end, we propose the three solutions of (1) democratizing digital transformation, (2) clean rooms to collaborate with competing walled gardens, and (3) customer choice in privacy by design. Doing so, managers can improve insights actionability within robust frameworks of within-business leadership, cross-business data trade and customer privacy. Meanwhile, we offer business academics and analysts specific research questions to solve intriguing and impactful puzzles. Thus, the future of big data analytics looks bright.

1. Introduction

IBM defines *big data analytics* as "the systematic processing and analysis of large amounts of data and complex data sets, known as big data, to extract valuable insights" (Mucci and Stryker 2024). Referring to the Volume, Velocity, Variety, Veracity and Value (5 V's) of big data, Wamba et al. (2017) define big data analytics as "a holistic approach to managing, processing and analyzing the 5 V data-related dimensions to create actionable ideas for delivering sustained value, measuring performance and establishing competitive advantages. In their systematic review of big data in business research, Zhang et al. (2021) show a boost in 'big data analytics' and 'decision making', and a decline in 'social media' and 'knowledge' (e.g. Colicev et al. 2018). We observe the same in our academic and business engagements. In this paper, we reflect on recent developments moving big data out of the social media and knowledge boxes to improve decisions with holistic insights across our omnichannel business environment.

Working backwards from the customer, Verhoef et al. (2015) define omnichannel as the "synergetic management of the numerous available channels and customer touch points, in such a way that the customer experience across channels and the performance over channels is optimized." Thus, the big data promise is realized when business can interact seamlessly with their customers across a multitude of channels with relevant and timely communications to provide a single unified experience across channels. Omnichannel business emphasizes a unified customer experience rather than just facilitating transactional/relational exchanges as the central function of customer-facing channels (Hoyne 2022). Likewise, managers and retailers leverage omnichannel to gain a better understanding of the nuances of marketing effectiveness, from cross-channel synergy to supersaturation. For one, digital billboards allow brands to reschedule ad timing to avoid over-exposing consumers who have already seen the ads in previous days on their commute. For another, retail media promises to deliver not just the right message to the right customer, but also at the time this customer is actively looking for options in a category.

Despite these opportunities, most businesses have been slow to harvest the omnichannel potential for big data analytics. The first obstacle remains organizational silos of data and decision-making power. At a recent conference, the first author met a consultant colleague from a quarter century ago, who asked whether the data part was still over 80 % of any business analytics project. The answer was 'yes' in both practice and academia. Likewise, 'TV creatives', 'data scientists', 'social media gurus', 'business leaders', and 'digital marketers' hardly talk to each other- let alone produce comparable metrics

* Corresponding author.
E-mail addresses: k.pauwels@northeastern.edu (K. Pauwels), z.aksehirli@northeastern.edu (Z. Aksehirli).

https://doi.org/10.1016/j.jbusres.2024.115112
Received 6 July 2024; Received in revised form 17 November 2024; Accepted 25 November 2024
Available online 30 November 2024
0148-2963/

for customer communication budget and allocation decisions. The second obstacle is the insistence of some managers and analysts to track (prospective) customers at the individual level online, implicitly ignoring their actions and marketing exposure offline. While such 'personalization' and multi-touch attribution (MTA) held promise, the practice is a pallid picture – and both regulators and tech companies make it harder to follow customers around the web. Third, and relatedly, awareness has been limited as to how aggregate-level attribution has advanced beyond the classic marketing-mix models (MMM). While keeping the traditional advantages of data coverage (both personalized and mass media) and dynamic synergy (upper funnel ad exposure today increases bottom funnel ad effects tomorrow), modern MMM is much faster to yield actionable insights even while the campaign is ongoing.

In this paper, we aim to detail these challenges, and dig deeper into proposed solutions. Our experience is both academic and practical, having worked in retail media ads for 5 years. Moreover, both consultancy projects and conversations with managers resulted in books on data analytics and democratizing digital transformation, breaking down organizational silos in 'Break the Wall: Why and How to Democratize Digital in Your Business' (Aksehirli et al., 2022). Of course, this experience is not universal, so we hope this article sparks ideas and the sharing of both challenges and solutions. After specifying several of such promising research questions, we reflect on the bright future of big data analytics.

2. Challenges

2.1. Organizational Challenges: Siloed data and decisions

Written a dozen years ago at the start of the Big Data hype, the first author's premiere book was entitled 'It's not the Size of the Data, It's How You Use It' (Pauwels, 2014) because many businesses did not even make good use of the small data they had. Unfortunately, with the advent of omnichannel business, these challenges have only intensified. Businesses are now inundated with an increased number of customer data touchpoints, each providing valuable insights yet complicating the process of data integration and use for decision making. Moreover, many of these touchpoints occur in a 'walled garden', i.e. "a closed platform or ecosystem wherein the provider of the platform has total control over the content, applications, and/or media and restricts access as it sees fit" (Munro 2024). Facebook, Instagram, Amazon, Apple and Google's Play Store are key examples. The proliferation of such walled gardens, where customer data is siloed within specific platforms, further exacerbates the difficulty in tracking customers across diverse channels and touchpoints. Indeed, omnichannel brings a plethora of new data challenges (Cui et al. 2021):

1) Increased number of customer data touchpoints;
2) Difficulties in tracking customers across channels and touchpoints of walled gardens; and
3) Increased consumer concerns about tracking and privacy.

In traditional offline channels, information about customer needs was mainly obtained by conducting surveys and from retailers observing customers' final purchase decisions. Such information was shared upwards along the channel. Information about product fulfillment was transferred downwards mainly via inventory delivery along the channel. As firms increase their channel footprint, they have greatly expanded the flow of customer-related information. However, these behavioral metrics, such as clicks and visits, move faster than sales and don't answer the WHY of customer actions. Clicks may be by bots or in error, while website visitors may be current customers looking for fire busting without being interested to buy. Pauwels and van Ewijk (2020) show how surveys still fill this gap, and demonstrate that survey-based attitudes are superior in predicting brand sales several months out.

If only the big data V's would directly translate into more actionable business insights! Unfortunately, they don't (Seggie et al. 2017, Soyer et al. 2019), because:

(1) Higher volume means more opportunity for cherry-picking with confirmation bias;
(2) Higher variety means more challenges in comparability and communication of insights;
(3) Higher velocity means more chasing of fast-moving but unimportant vanity metrics;
(4) Higher veracity means more blind confidence in the results, even when meaningless;

As remedies, the authors propose to start from the decision for which the data is collected and analyzed. What is your business goal and what are your available options to reach it? How exactly would more data help you make a better decision? Which variety of data helps you explain your decision to different stakeholders? When are insights needed for decisions, and how does that translate into correct time intervals (not everything has to be 'in real time')? How will results from one analysis be validated by other data and insights? The latter question is crucial when it comes to next challenge, attribution of business outcomes to our actions.

2.2. Operational challenges: Attribution in a world of walled gardens

Attribution is key in both marketing science and business practice, where it is used both to justify spending and to improve its effectiveness. Traditional attribution models, developed before the advent of the Internet, typically adopt a strategic, macro-level perspective. These models aim to estimate marketing-sales elasticities and optimize the allocation and sizing of marketing budgets. As discussed in "Demonstrating the Value of Marketing" (Hanssens and Pauwels, 2016), this research stream often uses marketing mix models (MMM), which are increasingly validated through field Experiments, in an iterative fashion (Model-Experiment-Model-Experiment; MEME). Beyond assessing the short-term isolated effects of individual marketing actions, the literature has also explored methods to measure synergy (e.g., Naik and Raman, 2003) and long-term impacts (e.g., Dekimpe and Hanssens, 1999). Intermediate metrics, such as attitudinal (survey-based) and behavioral (online) data, have proven valuable in enhancing predictive accuracy and diagnostic power (Srinivasan et al., 2010; Hanssens et al., 2014).

Online cookies gave attribution prominence to multi-touch attribution (MTA) with the benefits of unobtrusive, individual-level tracking of customer online interactions. The past decade has seen a surge in academic research on this topic, revealing the limitations of simplistic last-click and first-click attribution models (e.g., de Haan et al., 2016). Online behavior research has also mapped and visualized various consumer journeys in the digital space (Ghose and Todri, 2016) and developed more efficient techniques for analyzing and storing big data (Bradlow et al., 2017). However, MTA faces challenges when companies also use traditional marketing channels such as TV advertising, radio, print, and billboards, as seen with Amazon and Kayak.com (Pauwels et al, 2016).

Unfortunately, moving outside of the organization reveals more data silos. Indeed, the primary challenge of successful omnichannel marketing lies in the ability to *track customers across diverse touchpoints—both digital and non-digital, and those owned by the company and external entities* (Cui et al. 2021). For instance, a promising source of big data is retail media, which involves retailers helping advertisers reach shopper audiences through a variety of options, from bottom-funnel (eg Sponsored Products) to mid-funnel (eg Sponsored Brands) and upper funnel (e.g. connected TV) online to offline (e.g. personalized ads and coupons in-store). However, retail media data is available in walled gardens, which compete for the same business dollars, and therefore guard 'their' data and the secret sauce of their analytics and metrics. Business leaders such as P&G's Chief Brand Officer Marc Pritchard, has consistently lamented the lack of transparency and external validation over the last 7 years (Sluis 2017, Neff 2024). Beyond ecommerce, retail media now also comes to brick-and-mortar stores, requiring business analytics and attribution to integrate offline with online touchpoints.

2.3. Societal challenges: Customer privacy

Last but certainly not least, renewed consumer concerns about tracking and privacy may trump the promised benefits of personalization and convenience (Acquisti et al. 2020) Current evidence is lacking beyond self-reported privacy importance (stated preferences), which are not reflected in revealed preferences of low privacy price in consumer market choices (Johnson et al 2020,Lin, 2020[1]). As a result, we need better and quantifiable evidence on how consumers trade off privacy with price and convenience (Bleier et al 2020). Current papers only address the former, showing the willingness-to-pay for privacy is context dependent (Eggers et al 2023, Tomaino et al 2023).

Big data analytics require business to accurately collect and analyze data across various technology and channel touchpoints. Nowadays, businesses use two primary methods to track consumers across devices and media within their control: deterministic tracking and probabilistic tracking. Deterministic tracking involves using persistent login identifiers, such as email addresses, to accurately identify consumers across different platforms. For example, a Washington Post subscriber who logs into both the website and the app with the same email address can be consistently identified, enabling precise tracking of their interactions and preferences. Deterministic tracking provides high accuracy and thus give users a seamless and personalized experience, but only when they log in.

Conversely, probabilistic tracking helps when persistent identifiers are unavailable. This approach uses statistical methods to estimate the likelihood that various data points correspond to the same user. Elements such as IP addresses, device types, browser configurations, and usage patterns are analyzed to make informed assumptions about user identity. Although probabilistic tracking is less precise than deterministic tracking, it is essential in scenarios where login data is not consistently available. By applying machine learning models, companies can improve the accuracy of probabilistic tracking, using training data from deterministic tracking to enhance their predictions. Despite the inherent challenges and risk of misclassification, probabilistic tracking remains a vital tool for understanding consumer behavior across multiple touchpoints.

Key privacy questions for businesses include (Cui et al. 2021):

Are customers willing to have their data across devices and touchpoints collected, synced and used for marketing by the firm itself?
Are customers willing to have their data shared with other firms that control touchpoints along their consumer decision journey?
Are governments willing to allow firms to share, sync and collect data in this way?

3. Solutions

3.1. Solution 1: Democratizing digital transformation to overcome inside silos

While digital transformation has touched organizations for several decades, the current challenge is to democratize digital in your business, ensuring that both employees and customers are empowered to leverage digital tools and insights. In our book about the topic, we observe that the digital revolution has sharpened silos, with the data science analysts and the decision makers often residing in different buildings or floors and speaking different languages.

To address these issues, we propose a model that treats the organization as a collective of integrated parts of a complex system. In every organization, change is a constant phenomenon, occurring at various levels and at different speeds. This dynamic can be likened to a biological model, encompassing stages of birth, growth, maturation, death, and renewal. We adopt a Nesting Adaptive Change Model (Fig. 1), which can be applied to organizational change realities to aid with developing organizational resilience (Aksehirli et al., 2023).

As to time, we differentiate 4 phases:

Initiation of Change: This is the starting point where the need for change is recognized. Often, the impetus for change comes from the smallest and most agile parts of the organization. These units, being closer to external influences, can detect emerging needs and trends earlier.
Implementation of Change: Once change is initiated, the organization must implement it effectively. This involves allocating resources, modifying processes, and ensuring that all members of the organization are aligned with the new direction.
Building Resilience: As changes are implemented, it is crucial for the organization to develop resilience. This means creating systems and processes that can absorb and adapt to new changes without significant disruption.
Reconsideration and Renewal: After the initial changes have been embedded, there should be a continuous cycle of feedback and reassessment. The organization needs to review the impact of changes, learn from any challenges, and be ready to initiate further changes if necessary.

As the interconnected units of the organization adapt and develop resilience, several expected outcomes emerge. Adaptive changes often begin at the smallest units of the system and then propagate upwards and outwards from this point of impact. Consumer or industry touch points are often such units (e.g. sales, marketing, and service departments often notice consumer trends; research teams become aware

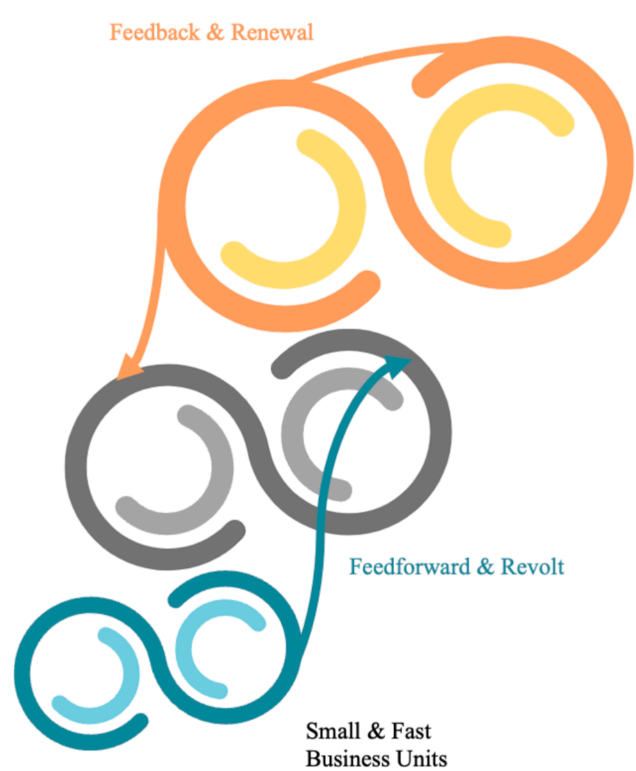

Fig. 1. Nesting Adaptive Change Model.

[1] "Though consumers express strong privacy concerns in surveys, we find that only 0.23% of American ad impressions arise from users who opted out of online behavioral advertising.".

of product trends; systems analysts spot and adopt technology trends). The segments of the organization interacting with the external environment are usually the first to perceive the need for change, attempt to address it, and as these needs expand, these smaller segments can instigate broader changes and growth cycles in adjacent or higher organizational levels. This cascading effect ensures that the organization builds both value and resilience. When lower levels successfully harness the value of these changes, the strategic perspective from higher levels provides a feedback loop, enabling continuous regrouping, reconsideration, and renewal of how the lower levels operate.

Big data analytics is the current wave that is making its way across the organizational levels. To make the complex system of interactions work smoothly, businesses must foster a culture of inclusivity and integration, where analytics literacy is not confined to specific departments but is a shared competency across the organization. This involves investing in training programs that equip employees with the necessary skills to utilize digital tools effectively and encouraging cross-functional teams to work together seamlessly. AI, in its current state, is becoming an efficacious tool in this transition (Hermann and Puntoni, 2024). Generative AI can be an effective tool in integrating skill-building and decision support for novice data analytics users. In fact, the anonymized and depersonalized nature of the generative AI conversation would alleviate technology adoption obstacles such as learning anxiety and feelings of low efficacy that many studies caution about (Iyer and Bright, 2024; Dwivedi et al., 2019).

In our Nested Adaptive Cycles (NAC) framework, there are two main directions that a change can flow. Imagine the small wheels spinning first, setting the larger wheels in motion. Think of specialists, those who know the nitty–gritty of their tasks, igniting the digital transformation journey. Picture the IT team revamping the customer onboarding process or the call center streamlining complaint management. These initiatives, though focused and immediate, often lack the foresight of long-term planning. This is where bottom-up approaches shine, driving incremental yet impactful digital changes.

For instance, at a leading tech company, a bottom-up initiative saw the support team develop a new ticketing system that cut response times by 50 %. The project began with ground-level insights, leading to high employee engagement and motivation. However, the team had to secure senior management's blessing, particularly when extra resources were needed, or when the change rippled across other departments. In top-tier organizations we've consulted, employees propose initiatives through well-documented processes, outlining customer problems and necessary inter-team dependencies. Senior management reviews these, consults with relevant teams, and provides constructive feedback and approval.

Contrast this with a top-down management style, where the larger wheels set the smaller ones in motion. Here, digital transformation projects start with managerial backing but might suffer from low employee involvement and practical challenges. Take, for instance, a retail giant switching to a new payment system. Senior leadership pushes the initiative, convinced of its strategic importance, and communicates the vision to the employees. These employees, with their hands-on experience, critique and refine the execution plan. Motivation skyrockets when employees understand their role and the broader benefits. Managers should articulate how the transformation not only enhances customer experience but also enriches the employees' work and career prospects. It's crucial to adapt project specifics based on employee feedback to ensure success. See Table 1 for the benefits and drawbacks of top-down and bottom-up approaches.

In reality, organizations seldom use these approaches in isolation. A hybrid model, blending top-down and bottom-up strategies, harnesses the strategic vision of senior managers and the innovative, technical prowess of specialists. This synergy often results in a successful and sustainable digital transformation.

A recent *meta*-analysis of big data analytics shows that such social factors have a stronger impact on enhancing firm performance than technical components in digitalization (Oesterreich et al., 2022). The interaction between different gears/levels of the organization as depicted in NAC is akin to the Dynamic Capabilities Theory (DCT). In DCT, an organization's success in changing environments depend on the robustness of their systems to sense, seize and transform relevant information into innovation (Teece, 2007). Sensing and seizing steps are more readily imitable, making the *transforming* step the actual sustainable competitive capability (Cadden et al., 2023). The foundational skills in *transforming* include knowledge integration, developing coordination skills, and establishing an innovative culture emphasizing psychological safety (Teece, 2007, Edmondson and Lei, 2014). All these require breaking down organizational silos as we discussed. Breaking down both decision making and data access silos in organizations can harness the full potential of digital transformation, driving innovation and enhancing customer experiences.

Table 1
Pros and Cons of Top-Down and Bottom-Up Approaches.

Top-Down Approach (Project Initiator: Senior Managers)	Bottom-Up Approach (Project Initiator: Specialists, Knowledge Workers)
Pros: • Comprehensive knowledge and big-picture vision of project initiators • Alignment with overall business strategy and vision • Early determination of strategic objectives and goals Cons: • Lack of flexibility • Poor responsiveness to ground-level issues • Higher implementation costs • Potential for low employee participation and motivation • Employees may feel undervalued	Pros: • Comprehensive knowledge and big-picture vision of project initiators • Alignment with overall business strategy and vision • Early determination of strategic objectives and goals Cons: • May lack long-term vision • Difficulty in seeing all factors affecting the problem or company's performance • Requires strong managerial support

3.2. Solution 2: Clean rooms to overcome outside silos

A recently popular solution to walled garden issues are clean rooms: secure, privacy-focused environments designed for the safe sharing and analysis of data between organizations without compromising the confidentiality of the underlying information (Smolinsky, 2024). The clean room setup allows different parties to combine their datasets and perform joint analyses (such as audience overlaps) without actually sharing the raw data with each other. A key example is the Amazon Marketing Cloud (AMC) in retail media. The retail media landscape in 2024 is complex and fast evolving, with the first day of Cannes Golden Lions seeing partnership announcements between Instart and YouTube, Omnicom and Amazon, and CVS pharmacy and The Trade Desk (Fig. 2). Recently, QuoVadis (n.d.) offered the Gravity Theory of Walled Garden to predict which and why hundreds of these walled gardens will trade data with each other through clean rooms.

Starting with the Supply of retailers, hundreds of retail media players have emerged after observing the success of Amazon, with tech providers such as Flywheel and The Trade Desk leveraging data. As to Demand, brand managers have been spooked by 3rd party cookie depreciation by Safari (default since 2016) and the four year year long saga of Google, who reversed its depreciation decision in July 2024. Many brands moved to first-party data with consumer opt-in consent to use their phone number of email into an identifier, of which e.g. Proctor & Gamble reportedly have over 1 billion. Next, streaming and connected TV (CTV) providers such as Netflix and Disney offer new advertising opportunities for the upper-funnel advertising, such as home screen takeovers and product placements. And while publishers such as the New York Times and Meta used to rely on third party cookies to connect advertisers with audiences, they are now operating their own walled

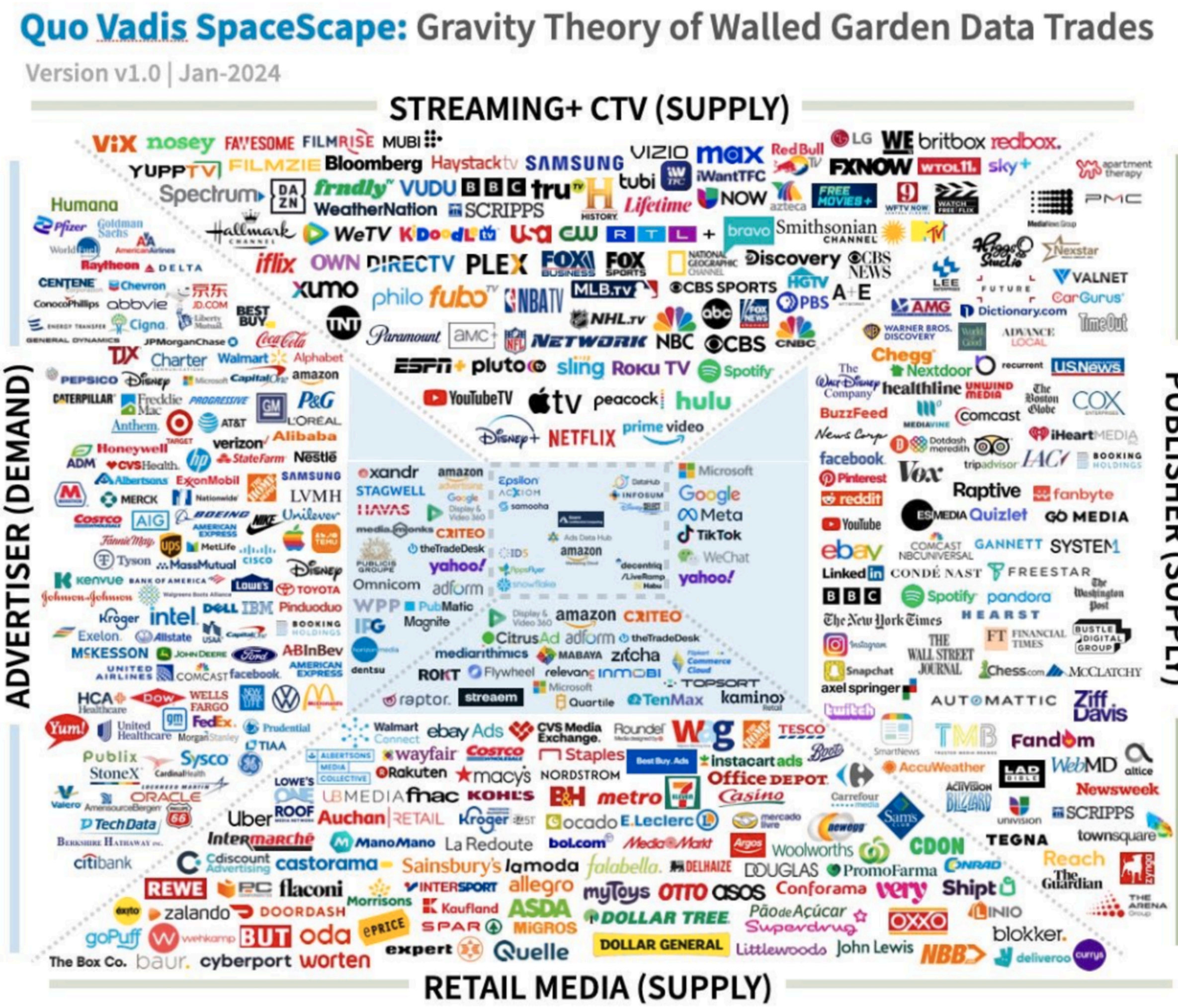

Fig. 2. Gravity Theory of Walled Garden Data Trades.

gardens. Enter the center players with clean rooms, such as Amazon Marketing Cloud and LiveRamp, where advertisers can integrate data sources.

Which of these players have incentives to trade data? In contrast to physical assets, data has increasing marginal utility; i.e. its value increases when more bits of information are connected (Burnham 2007). Hence, we except to see more collaboration interest among business that offer complementary assets. In January, LiveRamp acquired Habu for $200 M to accelerate data collaboration with enhanced clean room technology (Howe 2024), describing the combination as

> *"A first-of-its-kind, unified view allows customers to measure campaigns across all programmatic channels and walled gardens, including Amazon, Google, and Facebook, all from one place. Simplify partnerships with all media platforms and access new partners across media networks and all major MVPDs, CTV platforms, and TV programmers, including household names such as the Albertsons Media Collective."*

Or as Rich Ashton (2024), Managing Partner at First Party Capital, put it:

> "This transaction, completed at a multiple of 11x forward revenues, highlights the immense strategic value that first party data and privacy-preserving tech solutions can bring to incumbent adtech players that need to innovate away from their reliance on cookies, as marketers lose more and more signal."

Beyond these unified data though, managers need to evaluate the usefulness of the generated insights for business. With the majority of current retail media spent on lower-funnel actions (the equivalent of retail display for consumers already going through the aisle), retail media growth is coming from upper-funnel actions to induce prospective customer awareness and consideration (Srinivasan et al. 2009). For over 122,000 brands and product category combinations, Qin et al (2024) leverage weekly data from Amazon Brand Index which automatically and regularly measures Amazon shoppers' brand awareness, consideration and purchases and test how they change with ad and retail actions. They find that new product launches and upper-funnel advertising products are particularly effective for brands of low level of consideration, while those brands have yet to fully take advantage of these opportunities. Medium and large brands benefit most from lower-funnel advertising. As to funnel stages, all three metrics benefit from number of new reviews, % discount, negative keyword and geo reach campaigns. Furthermore, they all improve with high-traffic shopping events such as Prime Day and Thanksgiving in the United States of America.

Business analytics on these integrated data allows us to capture channel overlap, and whether it is *wasteful duplication or helpful synergy*. Synergy (positive interaction sign) implies the customer should be reached in the same period through different channels, while a negative interaction sign implies cannibalization (one channel suffices) or even supersaturation, with customers becoming annoyed with ads. We observed such supersaturation for TV ads: after an extensive mass media campaign, consumers expressed anger when even their favorite content was 'hijacked' by elaborate product placement. In contrast, scheduling product placement first allowed a new product or brand to gain familiarity, after which an ad campaign allowed businesses to give more specific information to considering consumers. This example shows how important timing is in synergy: most attribution models only follow the consumer journey for a limited time (eg 2 weeks in Amazon ads, Qin and Pauwels 2023). Moreover, classic marketing mix models only consider same-period synergy, often by multiplying spending on the marketing actions in that period. In contrast, we have consistently demonstrated the prominence of dynamic synergy, in which exposure to an upper funnel ad makes consumers more likely to respond to a lower funnel ad.

3.3. Solution 3: Customer choice in privacy by design

As Rich Ashton indicated with 'privacy-preserving tech', these business solutions only work to the extent they obtain the (increasingly explicit) consent of (prospective) customers. Our proposed solution is customer choice in privacy by design, which means that businesses give consumers a real choice in how their data is being collected and used.

Research is needed to go beyond asking customers and show how they respond to tradeoffs in real-life situations. First, we see natural trade-offs between a customer's acceptance of personalization and the degree of their privacy concerns and sense of control over their data (White et al. 2008, Tucker 2014, Ghose 2017). The initial phase of omnichannel marketing concentrated on integrating fragmented consumer data (Neumann et al. 2019). The next phase of technological advancements in omnichannel marketing may shift towards empowering consumers with control over how their data is consolidated by firms.

The recent explosion of generative Artificial Intelligence (genAI) has brought privacy and control concerns to the top of the legislative and business agenda. Can employees use commercially popular platforms such as ChatGPT if it reveals sensitive information to the platform provider (OpenAI)? How do customer trade off the benefits with the privacy risks? To what extent does the genAI user feel ownership of the output, based on their prompts and question follow-up, but generated by an algorithm owned by another business?

In a cross-continental project, we leverage conjoint analysis to investigate how people trade off privacy for convenience under different privacy regulation scenarios. It is crucial to understand this complex relationship to help regulators develop effective regulations that balance privacy and convenience while facilitating technological innovation. The tradeoffs typically involve "privacy calculus", where people evaluate both benefits and risks associated with sharing their data with companies, along with ethical considerations that may affect the perceived benefits and risks under different privacy regulation scenarios.

Last but not least, we highlight that these seemingly separate challenges and solutions are interrelated. For one, overcoming organizational silos (challenge 1) is necessary to obtain a 360-degree view of user activity (challenge 2). Likewise, this 360-degree view has limited value if consumers continue to oppose tracking their data with high precision (challenge 3). Thus, while the three solutions have independent value in themselves, their impact increases even further upon solving the other (and related) problems.

4. Research questions

As the above challenges and proposed solutions demonstrate, there is so much we don't know yet – which ensures big data analytics in omnichannel remains an exciting research topic! We offer the following research questions to get started:

1. Impact of Organizational Silos on Data, Analytics and Insights
- How do organizational silos affect the ability of businesses to leverage big data analytics for omnichannel marketing?
- What strategies can be implemented to break down these silos and enhance data-driven decision-making across the business?

2. Usefulness of Clean Rooms and Retail Media:
- How should clean rooms be designed and managed?
- Which market players should be incentivized to share data in clean rooms?
- How does retail media impact customer behavior across different stages of the customer journey?
- What are the comparative effects of upper-funnel, mid-funnel, and bottom-funnel retail media strategies on brand awareness, consideration, and purchase decisions?

3. Data Integration and Consumer Privacy:
- What are the most effective methods for integrating consumer data from multiple channels while ensuring privacy and compliance with regulations?
- How do consumers perceive the trade-off between personalized marketing and data privacy in an omnichannel environment?
- Does the negative customer attitude towards decisions by predictive algorithms (Hermann and Puntoni 2024) also apply to generative algorithms?

4. Attribution: multitouch (MTA), experiments and marketing mix modeling (MMM)
- What is the best purpose for multi-touch attribution (MTA) models versus MMM and field experiments in the shape of randomized controlled trials (RCT)?
- How should the attribution results of different models be triangulated?
- What are the challenges and limitations of attribution in a mixed media environment (online and offline channels)?

5. Consumer Tracking and Privacy Concerns:
- What are the primary consumer concerns regarding data tracking across different digital and non-digital touchpoints?
- Which consumer segments in privacy calculus are substantial enough to justify privacy-by-design choices business can offer?
- How do consumer preferences for privacy and personalization vary across different demographic groups and regions?
- How do privacy regulations (e.g., GDPR, CCPA) impact the ability of business to track and analyze consumer behavior in an omnichannel context?

5. The future of big data analytics

Looking beyond these research questions, we see several exciting avenues for the future of big data analytics.

First, the incorporation of both offline and online touchpoints allows the business a more holistic picture of its (prospective) customers. Different channels are converging because they are expanding to include other channel's traditional strengths. While social media is adding easy purchase options, retail media is moving offline and leveraging influencer marketing. These developments provide additional data and require analytics to examine synergies (e.g. prospects are more likely to buy when reached in different channels) versus annoyances (e.g. prospects do not like to be followed around the internet let alone also in offline retail).

Second, triangulation of methods is coming within reach as the holy grail of attribution. Businesses used to rely on multi-touch attribution, but the limitations of this method have become apparent in recent years, with both the accuracy and the legitimacy of third party cookies declining. Instead, marketing-mix modeling (MMM) is enjoying a revival as a privacy-safe and full funnel solution to quantifying which channels and campaigns drove sales, and recommending price, marketing budget size and allocation. Importantly, MMM uses time as the counterfactual, calculating sales lift as incremental sales above the baseline of times when a marketing action did not occur (controlling for all other variables in the model). In contrast, randomized control trials (RCTs) are experiments that use consumers as the counterfactual, calculating sales lift as the purchase difference between consumers exposed vs not exposed to the ad. More recently, geo-experiments have become popular, where all consumers in a delineated geography (e.g. a ZIP-code) are exposed while those in similar geographies are not. This approach combines the privacy benefits of MMM with the cross-sectional counterfactual of experiments, and thus holds plenty of promise for triangulating the – often very different- results of the different attribution methods.

Third, the relevance and rigor of business analytics has been expanded to include responsibility (Chintagunta 2024). As the panelists of the 2024 Marketing Analytics Symposium Sydney (MASS) discussed, business leaders went from constantly checking their stock price (and thus monetary compensation) to examining interactions with all key stakeholders. For one, while Environmental and Social investments on average decrease business evaluation by institutional investors, they increase business evaluation by customers (Malshe et al 2023). For another, increasing ethnic diversity in business communication only increases business prospects in certain conditions (Overgoor et al. 2023). While original 'responsible business research' looked for and identified win–win scenarios (doing well by doing right), current findings are more nuanced, helping managers understand the tradeoffs of their actions on behalf of the business.

Last but certainly not least, automation of big data analytics has entered a new, AI-led phase. One startup, MMM labs, offers an auto-MMM, which automatically verifies data quality, runs a range of reasonable models, provides the decision maker with scenarios and offers genAI-generated business implementation. For instance, if the model shows a higher effectiveness for an underused channel (e.g. retail media's upper funnel products), it also suggests budget allocation, specific messages and as execution for that channel. The human decision maker can and should still verify these suggestions, but can truly take a management-by-exception approach.

In conclusion, let's move big data analytics beyond silos. These silos relate to channels (e.g. social media and retail media), business disciplines and organizational departments, and competitive interests of walled gardens. Moreover, customer privacy is essential in big data analytics. We propose to (1) democratize digital transformation in your business, (2) organize data i clean rooms to collaborate with competing walled gardens, and (3) design customer choice in privacy. Doing so, managers can improve insights actionability within robust frameworks of within-business leadership, cross-business data trade and customer privacy.

CRediT authorship contribution statement

Koen Pauwels: Writing – original draft, Conceptualization. **Zeynep Aksehirli:** Writing – review & editing, Writing – original draft, Visualization.

Data availability

No data was used for the research described in the article.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology, 30*(4), 736–758.

Aksehirli, Z., Bart, Y., Chan, K., & Pauwels, K. (2022). *Break the wall: Why and how to democratize digital in your business*. Emerald Publishing Limited.

Ashton, Rich (2024), https://www.linkedin.com/posts/rich-ashton-fpc_data-platform-liveramp-buys-marketing-startup-activity-7153669320634437632-Zx_v/?utm_source=share&utm_medium=member_desktop.

Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing, 37*(3), 466–480.

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing, 93*(1), 79–95.

Burnham, Brad (2007) https://www.usv.com/writing/2007/12/googles-data-asset/.

Cadden, T., Weerawardena, J., Cao, G., Duan, Y., & McIvor, R. (2023). Examining the role of big data and marketing analytics in SMEs innovation and competitive advantage: A knowledge integration perspective. *Journal of Business Research, 168*, Article 114225.

Chintagunta, Pradeep (2024) Keynote on the Future of Marketing Analytics, Marketing Analytics Symposium Sydney.

Colicev, A., Malshe, A., Pauwels, K., & O'Connor, P. (2018). Improving consumer mindset metrics and shareholder value through social media: The different roles of owned and earned media. *Journal of Marketing, 82*(1), 37–56.

Cui, T. H., Ghose, A., Halaburda, H., Iyengar, R., Koen Pauwels, S., Sriram, C. T., & Venkataraman, S. (2021). Informational challenges in omnichannel marketing: Remedies and future research. *Journal of marketing, 85*(1), 103–120.

Dekimpe, M. G., & Hanssens, D. M. (1999). Sustained spending and persistent response: A new look at long-term marketing profitability. *Journal of Marketing Research, 36*(4), 397–412.

De Haan, E., Wiesel, T., & Pauwels, K. (2016). The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing, 33*(3), 491–507.

Dwivedi, Y. K., Rana, N. P., Jeyaraj, A., Clement, M., & Williams, M. D. (2019). Re-examining the unified theory of acceptance and use of technology (UTAUT): Towards a revised theoretical model. *Information systems frontiers, 21*, 719–734.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 23–43.

Eggers, F., Beke, F. T., Verhoef, P. C., & Wieringa, J. E. (2023). The market for privacy: Understanding how consumers trade off privacy practices. *Journal of Interactive Marketing, 58*(4), 341–360.

Ghose, A. (2017). *TAP: Unlocking the Mobile Economy*. MIT Press.

Ghose, A., & Todri, V. (2016). Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly, 40*(4), 889–910.

Hanssens, D. M., Pauwels, K. H., Srinivasan, S., Vanhuele, M., & Yildirim, G. (2014). Consumer attitude metrics for guiding marketing mix decisions. *Marketing Science, 33*(4), 534–550.

Hanssens, D. M., & Pauwels, K. (2016). Demonstrating the value of marketing. *Journal of Marketing, 80*(6), 173–190.

Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research, 180*, Article 114720.

Howe, Scott (2024), https://liveramp.com/blog/liveramp-habu-accelerate-data-collaboration/.

Hoyne, N. (2022). Converted: The data-driven way to win customers' hearts. *Portfolio*.

Iyer, P., & Bright, L. F. (2024). Navigating a paradigm shift: Technology and user acceptance of big data and artificial intelligence among advertising and marketing practitioners. *Journal of Business Research, 180*, Article 114699.

Johnson, G. A., Shriver, S. K., & Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science, 39*(1), 33–51.

Lin, T. (2020). Valuing intrinsic and instrumental preferences for privacy. *Marketing Science, 41*(4), 663–681.

Malshe, Ashwin, Yi Yin, Anatoli Colicev, Yakov Bart and Koen Pauwels (2023). How ESG Reduces Risk: The Role of Consumers and Institutional Investors. *Northeastern U. D'Amore-McKim School of Business Research Paper* No. 4410032, April 4, Available at SSRN: https://ssrn.com/abstract=4410032.

Mucci, Tim and Cole Stryker (2024), https://www.ibm.com/topics/big-data-analytics.

Munro, Brock (2024), https://www.publift.com/blog/walled-gardens#:~:text=A%20walled%20garden%20is%20a,goal%20of%20creating%20a%20monopoly.

Naik, P., & Raman, K. (2003). Understanding the impact of synergy in multimedia communications. *Journal of Marketing Research, 40*, 375–388.

Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling? Evidence from field studies. *Marketing Science, 38*(6), 918–926.

Neff, Jack (2024), https://www.adexchanger.com/ad-exchange-news/pgs-pritchard-dont-want-waste-time-money-crappy-media-supply-chain/.

Oesterreich, T. D., Anton, E., Teuteberg, F., & Dwivedi, Y. K. (2022). The role of the social and technical factors in creating business value from big data analytics: A meta-analysis. *Journal of Business Research, 153*, 128–149.

Overgoor, Gijs, Gokhan Yildirim, Yakov Bart and Koen Pauwels (2023), From Representation to Reception: Evaluating the Impact of Diversity in TV Advertising on Consumer Purchase Intention, *Northeastern U. D'Amore-McKim School of Business Research Paper* No. 4471248, Available at SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4471248.

Pauwels, K. (2014). It's not the size of the data–it's how you use it: smarter marketing with analytics and dashboards. *Amacom.*

Pauwels, K., Demirci, C., Yildirim, G., & Srinivasan, S. (2016). The impact of brand familiarity on online and offline media synergy. *International Journal of Research in Marketing, 33*(4), 739–753.

Pauwels, K., & Van Ewijk, B. (2020). Enduring attitudes and contextual interest: When and why attitude surveys still matter in the online consumer decision journey. *Journal of Interactive Marketing, 52*(1), 20–34.

Qin, V., & Pauwels, K. (2023). Upper funnel ad effectiveness and seasonality in consumer durable goods. *Applied Marketing Analytics, 8*(4), 345–352.

Qin, V., Pauwels, K., & Zhou, B. (2024). Data-driven budget allocation of retail media by ad product, funnel metric, and brand size. *Journal of Marketing Analytics*, 1–15.

Quovadis, https://triscari.substack.com/p/69-quo-vadis-spacescape-gravity-theory. Retrieved Sep. 2024.

Seggie, S. H., Soyer, E., & Pauwels, K. H. (2017). Combining big data and lean startup methods for business model evolution. *AMS review, 7*, 154–169.

Sluis, Sarah (2017), https://www.adexchanger.com/ad-exchange-news/pgs-pritchard-dont-want-waste-time-money-crappy-media-supply-chain/.

Smolinsky, Denny (2024), https://www.bebolddigital.com/blog/amazon-marketing-cloud#:~:text=AMC's%20clean%20room%20solution%20enables,and%20improved%20decision%2Dmaking%20processes.

Soyer, E., Pauwels, K., & Seggie, S. H. (2019). Big and lean is beautiful: A conceptual framework for data-based learning in marketing management. In *Marketing in a Digital World* (pp. 63–83). Emerald Publishing Limited.

Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). Mind-set metrics in market response models: An integrative approach. *Journal of Marketing Research, 47*(4), 672–684.

Srinivasan, S., Pauwels, K., Silva-Risso, J., & Hanssens, D. M. (2009). Product innovations, advertising, and stock returns. *Journal of Marketing, 73*(1), 24–43.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic management journal, 28*(13), 1319–1350.

Tomaino, G., Wertenbroch, K., & Walters, D. J. (2023). Intransitivity of consumer preferences for privacy. *Journal of Marketing Research, 60*(3), 489–507.

Tucker, C. E. (2014). Social networks, personalized advertising, and privacy controls. *Journal of marketing research, 51*(5), 546–562.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omnichannel retailing: Introduction to the special issue on multi-channel retailing. *Journal of Retailing, 91*(2), 174–181.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of business research, 70*, 356–365.

White, T., Zahay, D., Thorbjornsen, H., & Shavitt, S. (2008). Getting too personal: Reactance to highly personalized email solicitations. *Marketing Letters, 19*(1), 39–50.

Zhang, Y., Zhang, M., Li, J., Liu, G., Yang, M. M., & Liu, S. (2021). A bibliometric review of a decade of research: Big data in business research–Setting a research agenda. *Journal of Business Research, 131*, 374–390.

Koen Pauwels is the Associate Dean of Research and Distinguished Professor at Northeastern University and founding General Director of its Digital, Analytics, Technology and Automation (DATA) Initiative. Koen's European positions include Adjunct Professor at BI Norwegian Business School, and Visiting Professor at Vrije Universiteit Amsterdam. For 4.5 years, he was Principal Research Scientist at Amazon Ads, with brand building and budget allocation recommendations reaching hundreds of thousands of advertisers. Koen received his Ph.D. from UCLA, where he was chosen Top 100 Inspirational Alumnus. After getting tenure at the Tuck School of Business at Dartmouth, he helped build the startup Ozyegin University in Istanbul. Named a worldwide top 2% scientist, and 'The Best Marketing Academic on the Planet', Koen published over 100 articles on marketing effectiveness. This research was awarded by managers (Google, WPP and Syntec), academics (O'Dell, Davidson, and Varadarajan) and both (Gary Lilien ISMS-MSI-EMAC Practice Prize). Koen is the incoming editor-in-chief of the International Journal of Research in Marketing, and Associate Editor for the Journal of Marketing, the Journal of Consumer Research, and the Journal of Interactive Marketing. His books include 'Modeling Markets' and 'Advanced Methods for Modeling Markets' for analysts, and 'Break the Wall: Why and How to Democratize Digital in Your Business', and 'It's Not the Size of the Data – It's How You Use It: Smarter Marketing with Analytics and Dashboards' for managers.

Zeynep Aksehirli is a distinguished interdisciplinary educator, consultant, and author, dedicated to enhancing the human dimensions of organizations. Her expertise spans organizational change, corporate culture, market strategy, and remote work. Zeynep has an portfolio of publications, including articles and books on corporate culture, organizational network structures, leadership competencies, and intranet design. Her recent book, Break the Wall: Why and How to Democratize Digital in Your Business, delves into digital transformation, featuring case studies from various companies. Currently, Prof. Aksehirli is working on a new book titled Culture of Curiosity, which explores corporate culture within high-tech firms. Her consulting experience is extensive, collaborating with organizations such as the Committee for Public Counsel, Kenvue/J&J, Goldman Sachs, and Amgen. She is also a prolific judge and mentor within entrepreneurial ecosystems such as the MassChallenge Community. Before her tenure at Northeastern University, Zeynep was a founding faculty member and Dean of Students at Ozyegin University School of Business, and a faculty member at Tuck School of Business at Dartmouth.