Consumer Knowledge Graphs for Media Companies

Marco Rota
Field CTO, Media and Entertainment
Microsoft

Every company wants to be more data-driven. This is especially true for media and entertainment companies, which have new requirements for customer data due to the industry's shift to direct-to-consumer. Today, businesses must not only understand their clients better, but they need data to decide what content to produce, what media to acquire for their streaming platforms, how to sell in non-linear ways, and more.

To survive in today's highly competitive and constantly evolving world, M&E companies need a way to get a 360-degree view of their consumers quickly and efficiently. What's the solution? A real-time, privacy compliant consumer knowledge graph that allows companies to easily share information with one another and get access to all kinds of timely and relevant data. With the consumer graph solution, companies can focus their efforts on their true competitive advantage: the data-interpreting models that allow them to turn all this information into meaningful business results.

For this paper, Microsoft and PwC spoke with senior executives, company directors and leading data experts across the M&E and consumer products sectors including gaming, television, streaming, data intelligence, quick service restaurants and more. They told us about their data-related challenges and what they need to remain competitive in the future.

Contents

The data challenge

While there is more than enough information available that companies can use to inform their strategic decision making, simply having data is not enough. All the companies we spoke with are no longer interested in amassing a vast library of consumer data – they would rather have relevant information that better applies to their business at that moment, while ensuring customers trust how their information is collected and used.

If companies had the data they needed, they could deliver better products and services than they do now. For instance, the networks and content creators we spoke to want to develop more accurate personalization and recommendation engines across their distribution channels. They want to use data to anticipate and greenlight successful projects, whether it be television shows, viral content or video games.

Advertisers need real-time data to help them determine what spots to purchase within a show or where to place products in steaming video. Live event promoters want to ensure that they're booking concerts into venues they know will sell out, while sports teams want to provide fan experiences that will maximize wallet share even after they've converted them to customers.

An executive from a quick service conglomerate we spoke to wants their customers to dine at more of its brands each week, rather than just once. Better data would allow the company to improve its messaging, create new services and present offers from its other restaurants.

Companies are having trouble getting to where they need to be in part because they're dealing with many data challenges today. Four main ones stand out: privacy, volume, granularity and quality.

Privacy

Marketers have always wanted to understand as much as possible about individual customers so they could hyper target them with ads or offers. It turns out, consumers don't like it when companies, and especially ones they don't have an affiliation with, send them highly personalized ads. "It's become creepy," was a common sentiment among the executives we interviewed. At the same time, Europe's General Data Protection Regulation (GDPR), and California's CCPA and others like it have made collecting and storing personally identifiable information (PII) more difficult. These rules were created to better protect consumers, in part because of how people are feeling about the way their data is being collected and used. They were also developed to protect consumers from data breaches, such as the 2020 attack on IT company SolarWinds[1], which saw a number Fortune 500 companies hacked.

As consumers become more knowledgeable about data, and especially with Google moving to eliminate cookies by 2023[2] and Apple making it easier for its users to stop their apps[3] from tracking their online behaviors, people are becoming increasingly hesitant to hand over highly personal information that could be used in ways that they don't agree with. A recent PwC survey found that 43% of U.S. consumers would not give companies permission to collect their personal data, such as location, age, lifestyle, preferences and purchase history[4], while Flurry Analytics reports that just 13% of iOS users have explicitly allowed their apps to track them[5]. As a result, M&E businesses need to collect data in an anonymized and compliant way.

Volume

Companies can track nearly everything, from when people go to sleep at night, to the route they take to get to a sporting event, to how many minutes they spend watching a particular program. As impressive as that may be, it has become wildly expensive for M&E companies to gather, store and analyze all this information. Many businesses are also collecting the same information from the same people, which means a lot of time and money is being invested into gathering data that others have and could share. If companies had a way to easily access this intelligence, they could spend more time using the data instead of collecting it.

Granularity

Given how much data is out there and the privacy concerns around what's getting collected, companies are questioning just how granular the information they gather must be. Rather than simply ingesting volumes of personally identifiable information (PII), many of the businesses we spoke with now want to define and target personas – groups of people with shared interests and characteristics. However, while businesses can broadly define personas – "video game user in their 40s," for instance – they need data with more dimensionality to effectively target more specific groups, such as "white females in their 20s who like golfing on weekends and enjoy watching sports movies." Defining personas at the right level of granularity is currently a challenge.

Quality

One of the biggest challenges around data is that much of it isn't usable. Every data provider – whether it's a company generating first-person intelligence or an outside business delivering third-party information – presents their findings in different formats and styles. Currently, there's no standard method of collecting or presenting data – which presents major challenges for companies, who spend too much time "cleaning" data to make it usable and not enough time analyzing it. "You have to spend time and resources to get the data into a proper state before you can even start doing anything with this stuff," said one executive we spoke with.

Data unification

Organizational and data siloes are another big issue. Many companies are collecting vital information across their various lines of business, but they're not sharing that data with other divisions. These siloed data stores are preventing companies from unifying their data collection efforts and from getting the most out of their information.
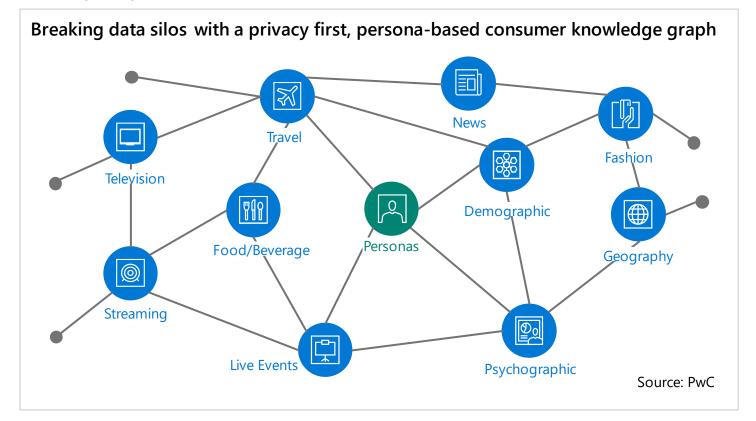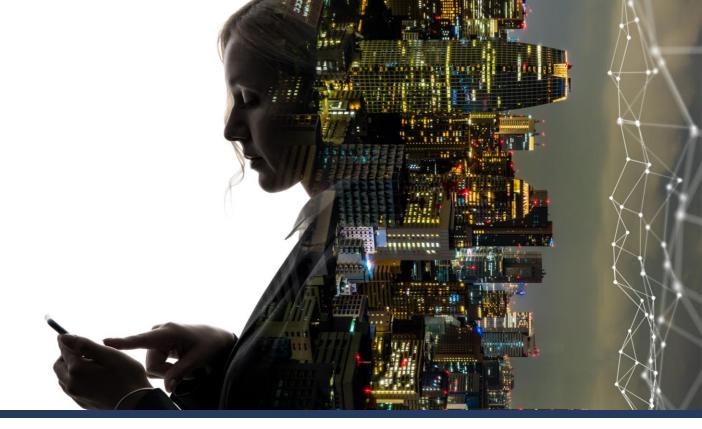
Understanding the consumer knowledge graph

Fortunately, the executives we interviewed are optimistic about the future. They recognize that big data is still in its early days, with one person saying that the data industry is only in the second of nine innings. The rest of the data game may advance sooner than people think. With more people interested in persona-driven content, anonymized and secure information, real-time intelligence and more, companies need access to a consumer knowledge graph.

A consumer knowledge graph would let M&E companies access valuable information quickly and efficiently from a vast array of sources. It would also allow them to publish their data to a graph in real time with the option of sharing and subscribing to other data sets in a secure and privacy compliant manner. It would be a marketplace that connects disparate and diverse consumer data sets in a single platform that's unified and standardized.

Companies would have access to all kinds of anonymized first- and third-party data – from what consumers stream at night to how often they go to concerts to what fast-food restaurants they enjoy and so much more – which they could use whenever and however they wanted. With this kind of graph, which differs from other kinds of graphs in that most existing ones are specific to a company's own data, businesses can get just enough information without having to collect and analyze everything. This allows them to focus more on developing the right models on top of these data sets.

Breaking data silos with a privacy first, persona-based consumer knowledge graph

Source: PwC

To put it another way: Imagine going to a mall that had every brand you could possibly think of and then creating a wardrobe from the stores that fit your style and personality most. A consumer knowledge graph is no different – every piece of data would be available from whomever gathered it, whether that be a third-party company or a competitor, and you can pick and choose whatever information makes the most sense with your strategic objectives.
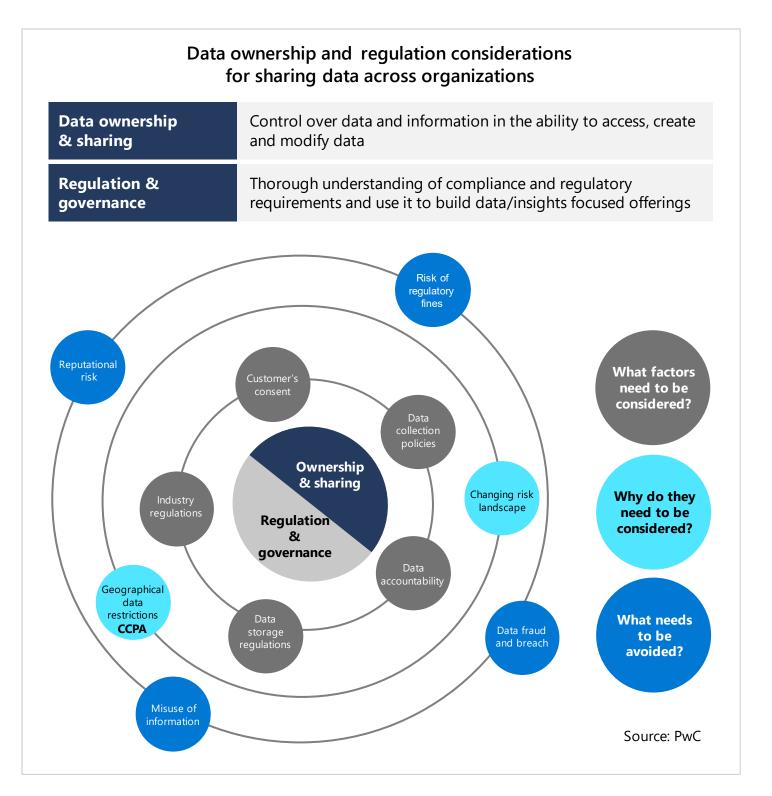
How it works

A consumer knowledge graph should be easy to use:

- Companies would collect their own data and then publish their information to the graph in real time using an API. These APIs would replicate data structures that represent common industry platforms, such as social media APIs, Google, Adobe and Bing analytics, and other well-adopted M&E customer data platforms.

- This streamlines the need for complex data engineering to publish and to subscribe to the graph and allows artificial intelligence to be automated and built on top of the graph.

- Businesses could then share that data with other companies – through a simple contract interface – and purchase intelligence that's uploaded to the graph.

- All the data sets uploaded to the graph – demographic, psychographic, technographic, behavioral and geographic consumer-related information – would be in a common and anonymized language.

- Privacy and security would also be built into the graph, while an audit trail will ensure custody of data.

- Companies would access information deemed relevant from the graph and plug it into their data models to ultimately drive meaningful business value.

- The data would be stored in the public cloud rather than on a company's own on-prem servers. This would reduce significant data storage costs.

Of course, building a knowledge graph is easier said than done. To get there, companies must share their data and embrace data monetization, both of which businesses may not be used to doing. However, many of the executives we spoke to said they were interested in both sharing and selling data, as long as their information could be anonymized and not used to target specific individuals.

Data ownership and regulation considerations for sharing data across organizations

Data ownership & sharing	Control over data and information in the ability to access, create and modify data
Regulation & governance	Thorough understanding of compliance and regulatory requirements and use it to build data/insights focused offerings

Source: PwC

For the graph to work successfully, companies would also have to trust that the true value lies in the analytics and models that sit on top of the data, rather than the data itself. According to the people we spoke with, data is important, but businesses must also know how to properly analyze and action upon the information to realize value. This is key, especially for those concerned with how a graph could impact a business' competitive advantage.

That may always be a worry but given that companies will have their own data models, and their own goals and purposes, companies will use the information from the graph in different ways. The competitive advantage would come down to what kind of data a business thinks is valuable and how they ultimately use that information.

How a consumer knowledge graph can help

Despite some logistical and cultural challenges, many agreed that a knowledge graph is the future. Why? Because they recognize that a graph would solve key pain points that currently plague the M&E industry today.

Keep data anonymous and build trust

One of the biggest benefits of a knowledge graph is anonymity, as every piece of data that gets uploaded would be free of PII. This would help companies remain compliant with regulations, while customers can feel confident knowing that their data is being used safely and securely.

Also, the graph could have policies and approval gates built in that mine for PII or any PII risks. This kind of anonymous and privacy-conscious platform would also help increase trust between brands and consumer. Since companies would only add relevant bits of data to the graph, businesses could ensure privacy and data security.

Create personas

A graph would allow companies to define personas more easily, as they'll have access to all kinds of demographic, psychographic, technographic, behavioral, and geographic data. They can then use that information to target more specific groups, without needing to collect PII.

A more unified language

Another advantage to a graph is that it would force industries to create a unified and connected data language. Information would have to be presented in a standardized and usable format, so that everyone can benefit. This would help to eliminate data cleansing and allow businesses to spend much more time analyzing information.

Break down silos

A graph will also help connect currently siloed data. If companies upload their information to the graph, then walled gardens will crumble and a new data-rich world will open up. Data that was once impossible to access will become available to all.

The knowledge graph in action

Once the graph gets made, getting data into the graph is simple. Here's how it works using content streaming as an example.

1. Customer engagement
A consumer creates an account on their streaming platform of choice. Various data points are collected, such as age, gender, payment information and phone number.

2. Customer experience
The consumer then watches a few shows. That generates more data, such as titles watched, genres and actors they like, time spent streaming and more.

3. Data capture
The streaming company receives several data signals that fit into several personas. Those signals might include segments, context, PII, purchase and mobility patterns.

4. Cleanse and publish
That data is then anonymized and cleaned up in a standard data format that others can use. It then gets uploaded to that company's part of the consumer knowledge graph.

5. Utilize marketplace
The streaming company then pays for access to another business' anonymized data. They can also sell their own data into the marketplace if they so choose.

6. Derive insights
The company, now armed with its own data and information taken from the graph, can run their own models on top of that data to generate insights.

7. Delight customers
With better insights, the streaming company can provide value-added services to its customers and drive continuous engagement. It might, for instance, recommend more romantic comedy movies to users who fit a certain persona, or greenlight different kinds of programs for specific groups.

Use what matters

In a marketplace environment, companies won't have volume problems anymore. Rather than having to collect every piece of information, you would focus on gathering data that others would find useful – and then pay for. Businesses would only buy the intelligence they need. This intelligence would be stored in the cloud, rather than by each company.

Ethics

A graph could also make data collection more ethical. Because companies would be able to choose from a variety of anonymized data sources, they can go beyond the typical user and generate insights on a more diverse population. A graph would open new markets to previously untargeted consumers, allow for a shift away from

demographics that create bias and move towards a way that enables more ethical marketing.

Easier to manage

A graph will dramatically reduce the number of third-party companies and data-collecting technology platforms businesses will need to use. While some of these tools may still be enlisted to generate the data uploaded to the graph, none are required beyond that. The graph itself will become the central platform.

More holistic view

Ultimately, a graph will provide companies with a far more holistic view of their consumers. They will learn about customers in ways they never could before, ultimately allowing them to deliver superior services and experiences.

More data uses

Every executive who was interviewed for this piece wants one thing: to better understand their consumers. The more they know about what people like, the better products and services they can create for their customers.

As this industry evolves, use cases for data, and the consumer knowledge graph, will expand exponentially. At some point networks might create online content for highly specific personas; advertisers could serve up ads in different ways for different people; retailers and restaurants could create offerings more tailored to the personas they've defined and much more.

Whether it's marketing, advertising, production or distribution, every company needs data to make more-informed business decisions. For the Media

and Entertainment industry to reach the ninth inning, it will need a knowledge graph that everyone can benefit from.

Microsoft empowers media and entertainment organizations to achieve more with our trusted and secure platform supported by a comprehensive partner ecosystem with industry-leading solutions for creativity, collaboration, content management, audience insights, and personalized customer experiences.

Learn more about intelligent media and entertainment to see how the capabilities of our solutions have optimized workflows, streamlined content delivery, enabled informed decision-making, and deepened audience engagement for the likes of Digital Domain, WPP, the NBA, the NFL, LaLiga, Gruner + Jahr, waave, and the BBC.

References

1. Isabella Jibilian, Katie Canales, The US is readying sanctions against Russia over the SolarWinds cyber attack. Here's a simple explanation of how the massive hack happened and why it's such a big deal, Business Insider, April 15, 2021, https://www.businessinsider.com/solarwinds-hack-explained-government-agencies-cyber-security-2020-12

2. Abner Li, Google now plans to phase out third-party cookies in Chrome by 2023, 9To5Google, June 24, 2021, https://9to5google.com/2021/06/24/google-chrome-cookies-plan/

3. Apple advances its privacy leadership with iOS 15, iPadOS 15, macOS Monterey, and watchOS 8, Apple, June 7, 2021, https://www.apple.com/newsroom/2021/06/apple-advances-its-privacy-leadership-with-ios-15-ipados-15-macos-monterey-and-watchos-8/

4. Tom Puthiyamad, Jose Reyes, Experience is everything: Here's how to get it right, PwC, https://www.pwc.com/us/en/zz-test/assets/pwc-consumer-intelligence-series-customer-experience.pdf

5. Mayank Sharma, Barely any iPhone users have agreed to app tracking in iOS 14.5, May 10,2021, TechRadar, https://www.techradar.com/news/barely-any-iphone-users-have-agreed-to-app-tacking-in-ios-145