Digital Observatory Research Cluster

The 6ᵗʰ Wave: The New Role of Game Data Across Industry and Society

How game data evolved from a curiosity to underpinning the global game industry – and how they are now being used beyond games to shape the future of games research

Summary

Four billion people globally play games, and they generate petabytes of behavioural telemetry data every year. Over two decades, game data have gone from being a niche curiosity to forming a keystone in the $200bn+ global game industry and associated academic fields. The impact today on how games are developed, maintained and generate revenue, is enormous. But game data can do more, and we are just now seeing the emergence of a 6ᵗʰ wave of game data, which seeks to push the use of behavioural data beyond the game, employing data to support creators, inform businesses and assist societally beneficial research.

Key takeaways

- The field of game data science, or game analytics, has emerged over the past decade to support the use of game data across industry and academia – and increasingly for societally beneficial research.
- The history of game data can be divided into five waves, with a sixth wave currently emerging that seeks to use game data for a variety of purposes beyond game development and research, including supporting creators and audiences, transforming the business of games, and benefiting society.
- The 6th wave of game data offers unprecedented insights into how users play games, what works and what does not, and offers unprecedented ways of shaping the user experience around the individual and the context.

What are Game Data?

Game data is a somewhat nebulous concept but broadly refers to any data derived from game systems or the context in that games are played. This includes transactional data and data from our communities, used for example in sentiment analysis. We typically call this telemetry data. Additionally, we collect data from the agents of the game system, so we can monitor the behaviour and performance of e.g., AI bots.

Such data can be incredibly high-dimensional. A typical major commercial title generates thousands of behavioural features, across potentially dozens to hundreds of millions of users. Potentially every single action taken by a player can be recorded and analysed. If we additionally have data from the context of the play, e.g., multi-player interactions, this dimensionality increases.

Less massive in scale but highly important are production data, as well as data derived from the physical infrastructure of games. For example, server logs or infrastructure performance.

And finally, we can integrate contextual data from any other situation that might be useful for the purpose at hand, for example, user research data or reactions to game artwork – often we gain enormous value from combing behavioural telemetry with qualitative data from interviews and surveys.

This is largely how we have approached game data so far. But another perspective on game data is the following: games touch 4 billion lives and telemetry from games tell us not only how we play, when we play, where we play and what we do in games, but also indirectly how we deal with challenges, how we interact with others online, what makes us frustrated and how we deal with that frustration – for example exhibiting toxic behaviour.

We can even use game data to make at least educated guesses about addictive behaviour, wealth, intelligence and more. No other data source provides this kind of global view of humanity, and while there is a huge potential here, there are also ethical implications.

How Game Data is Used in the Industry

From the perspective of the industry, game data science can be fielded at different levels within an organization: Operational, Tactical and Strategic. The requirements and interests at each level are different:

The strategic level is the global view of how a game should evolve based on the analysis of user behaviour and the business model. For example, defining a monetization model, designing the player journey or scoping a new DLC pack.

The tactical level informs game design in the short term, for example, A/B testing a new game feature, building prediction models to understand how our player base might engage with the game in the future, or building profiles of our players to inform designers about how the game is being played.

The operational level is the analysis and evaluation of the immediate, and current, situation in the game. For example, removing a bug, adapting a game to a change in the game's meta, or reacting to prevent a new type of cheating or piracy.

While it is difficult to generalize, a mature implementation of game data science in a company would see data used at all these 3 levels, i.e., see data being infused in decision-making processes across the entire organization, and everyone in the organization having access to the data they need to help inform their decision-making processes.

It is very important to note that making games is a creative process, blending art, science and craft, and the use of game data should never conflict with the creative process, but instead seek to support it by informing decision-making.

Game Data Science

Game data are now used for a great variety of purposes within the games industry and academia. For example, we might run an A/B test to explore which of the two designs our audience finds most engaging. We could also analyse the way our users play our game to try and identify points where they have problems progressing. We can look at movement data to understand if all the environments we have built are being utilized. We could also build a prediction model which tells us something about which players that might stop playing.

The field of game data science, or game analytics, has emerged over the past decade or so to support the use of game data across industry and academia. There is, however, no commonly accepted definition of what game data science is, but basically, game data science combines multiple disciplines towards extracting meaningful insights from game data.

Because the User Experience (UX) is vital for the success of games, game data science goes hand in hand with user research and user testing. Data science can tell us what players do, and while we can draw inferences as to why we see specific behaviours, we need user research to tell us why.

While we are glossing over a considerable amount of complexity, it is safe to say that data are now foundational not just for the creation of games but also for how we manage them and make a living from them. Importantly, in games, data is also increasingly something that is shared with our players, through APIs or data services, so they can utilize the data themselves, for example, to improve their performance in esports. The games industry is unique in this data sharing.

The vast majority of game data research happens in the industry, not at universities. Academia is out-competed by a factor of as much as 1:500 in terms of the number of data scientists. That being said, academics serve a vital support function and pursues strategic research goals that can be difficult to justify in a company. Furthermore, in academia there is a growing interdisciplinary body of work that also looks at using game data as a basis for supporting research across numerous disciplines, for example studying human behaviour, exploring new algorithms, understanding teamwork and much more.

The History of Game Data

If we should try to summarize the roughly 20+ years of history in the use of game data - and game data science – we can identify 5 waves, with a 6th currently emerging.

1. **Data for User Research:** The first wave stretches from perhaps the late 90s to the mid-2000s, which was a period when the use of telemetry data was relatively rare. Microsoft took a big leap during this period when the company began using telemetry to support user testing. At the same time, on the academic side, it was realized that MMOs have substantial economies, which led to a proliferation of research.

2. **Rise of Social Media:** The second wave arrived with the rise of social media, smartphones and the introduction of the Free-to-Play business strategy. These new formats required constant awareness of the users and started integrating basic Business Intelligence functionality before and notably after the game launch. Informing LiveOps teams became important.

3. **Enter the Algorithm:** The third wave saw the introduction of machine learning. Anders/I was heavily involved in this wave. This was a renaissance period where data collection exploded, and the introduction of machine learning meant we could derive insights from data that were not possible before, such as building detailed behavioural profiles of our players. At the same time, game data service providers emerged at a rapid pace.

4. **Looking for Why:** The fourth wave saw game data science working closely in hand with game user research to broaden the scope of what we were doing, integrating psychological methods and cognitive theories to move beyond just looking at what players do, but also to draw inferences about why. This is also when the first ultra-large-scale, cross-game studies were done in industry and academia, looking for fundamental laws and principles of game behaviour, e.g., the *Playtime Principle* which provided the first mathematical model of player engagement.

5. **Maturity:** The fifth wave is the one we are currently exiting. Covering the past 5-7 years, this wave saw the introduction of advanced machine learning and AI in the context of industrial game data science, and sophisticated tool generation. We can today adapt games to individual players in real time. Interestingly, the use of game data to build experiences around games also proliferated, notably in esports where data are used by virtually everyone in the community. This has encouraged even more data sharing by publishers. People also started thinking a lot about the ethical implications of game data in general.

6. **Beyond the Game:** The sixth wave is emerging. For the 6th wave of game data use, we are seeing new approaches towards transforming the business of games, supporting creators and audiences, and even using game data to understand human behaviour and benefit society. This wave centres on the idea that the space around games is a new frontier for experiences that occur around games and that data will be needed across a diverse set of experiences and educational opportunities in this space.

The 6th Wave of Game Data

Starting with business, the 6th wave of game data is expanding analytics outside the confines of specific games and recognizing that experiences built around games are equally important to the user experience and the financial success of games. Esports is leading the push on this. Furthermore, to map lessons learned from one game to other games in the portfolio. There is a big push happening toward building future virtual economies and figuring out what kinds of tools we need in a situation where game economies become interlaced and take on the complexity of real-world economic systems.

The 6th wave is also seeing much more detailed views on players and creators emerging, with traditional models and expectations about who players are and what they do giving way to much more sophisticated profiling and real-time adaptation.

For creators and players, game data are beginning to be used in combination with machine learning and creative AI to provide new insights into gameplay, new ways to interact with games and new ways to build experiences - and make a living - in the space inside and around games. The continued release of game data, combined with the emergence of walled gardens like Roblox, is giving creators opportunities that did not exist a few years ago.

The 6th wave is differentiated from prior game data science in four ways:

1. **Scale –** we are working with massive, highly-granular datasets that span thousands of titles and tens of millions of users.

2. **Focus –** not just on games, but on their context to the experiences built around games which is a rapidly expanding market space. This extends to the real world, correlating game data with societal and geographic data.

3. **Social responsibility:** 6th wave game analytics does not just seek to help generate business impact in terms of revenue but attempts to create socially positive results for broad groups of stakeholders, ranging from small creators to individual players and communities, thereby strengthening brands, engagement, and positive wellbeing impacts.

The focus on new experiences is not only focused on building in the space around games but importantly also on informing game design. Telemetry data are a rich source of information about which designs that work, and which do not, across the complexity of humanity.

Beyond the analysis of data from current games to inform the design of future games, we predict the emergence of "data artists" who use telemetry to build visualizations and artwork that seeks to engage audiences, and on the developer side also create bonds between the art and science of creating games.

There is also an immense educational potential, exemplified by esports which sees a community hundreds of millions strong engaging with data to improve their own performance and a plethora of services around data, from AI-driven coaches to tools for evaluating the performance of your team. This is an enormous contribution to the generation of data literacy skills worldwide.

On the research side, we are seeing the emergence of the combination of game data with data from real-world contexts to generate socially positive research. For example, understanding human behaviour at huge scales, exploring economies and spending, connecting games and well-being, or studying human psychology.

Conclusions

With the realization that games are an integral part of society and that the experiences around games can be just as important as games themselves, comes a new frontier that has not been much explored, but holds substantial potential.

The data generated from games inform us not only about how we play games but also increasingly about the world beyond games. We have barely started exploring how we can use game data to understand social interaction, decision-making, wealth, emotional well-being and more. Arguably, this early 6th-wave work is mainly using game data to *learn* about society, rather than more directly impacting society, but the direction is clear. Exceptions include informing policymakers about the impact of different monetisation strategies and thereby promoting balanced and positive legislation in the Creative Industries.

For the industry and its creators, the 6th wave of game data offers unprecedented insights into how our users play our games. What works and what does not? Game data can inform our design and offers unprecedented ways of shaping the user experience around the individual and the context.

Despite its history of two decades, game data science has barely scraped the tip of the iceberg, and the 6th wave in many ways feels more like the beginning of a new paradigm for the use of game data than merely a new chapter.

Further Information

David Zendle, Catherine Flick, Darel Halgarth, Nick Ballou, Simon Demediuk and Anders Drachen: *Cross-cultural patterns in mobile playtime: an analysis of 118 billion hours of human data*. Sci Rep 13, 386 (2023). https://doi.org/10.1038/s41598-022-26730-w

David Zendle, Catherine Flick, Sebastian Deterding, Joe Cutting, Elena Gordon-Petrovskaya, and Anders Drachen: *The Many Faces of Monetisation: Understanding the Diversity and Extremity of Player Spending in Mobile Games via Massive-scale Transactional Analysis*. ACM Games 1, 1, Article 4 (2023). https://doi.org/10.1145/3582927

To quote

Drachen, A., Zendle, D.: The 6th Wave: The New Role of Game Data Across Industry and Society, Digital Observatory (York, UK), April 21st, 2023. Available from: https://thedigitalobservatory.files.wordpress.com/2023/06/drachenzendle_6thwave_whitepaper.pdf

This white paper has been released by the Digital Observatory Research Cluster, a non-profit, cross-institutional academic research group focusing on delivering data-driven insights and observations about our digital lives to inform people, society, industry and policy. Find out more on: digitalobservatory.wordpress.com

Contact: Dr David Zendle, Director, david.zendle@york.ac.uk