# Ethical Behavior: Is your data strategy harmful or helpful?

By Matt Green,
VP of Customer Operations
May 2018

inter|ana

Stop Guessing | **Start Knowing**

*"Much like the Hippocratic Oath defines Do No Harm for the medical profession, the data science community must have a set of principles to guide and hold each other accountable as data science professionals. To collectively understand the difference between helpful and harmful. To guide and push each other in putting responsible behaviors into practice. And to help empower the masses rather than to disenfranchise them. Data is such an incredible lever arm for change, we need to make sure that the change that is coming, is the one we all want to see."*

DJ PATIL, FORMER U.S. CHIEF DATA SCIENTIST, RECENTLY WROTE IN A CODE OF ETHICS FOR DATA SCIENCE

Part of this change is a shift in who controls the data. In today's world of large-scale sophisticated data mining, governments, businesses, and individuals are starting to ask whether we're all benefiting equally from the data. The EU has enacted the General Data Protection Regulation (GDPR) as an attempt to answer that question by giving their citizens direct control over where and how their information is used online.

**Individuals are shifting their behavior to reduce the information they share with online services—**how many of us have installed ad blockers, used private browser windows, or reduced or even stopped our use of social media? Either by choice or legal mandate, organizations are giving their users more control over their own activity online and how that activity is later used. The overarching themes in this shift are transparency and control— transparency in what an organization is doing, and control for the user over how their activities are used.

Behavioral analytics is similar to running a focus group or participating in a research panel. The difference is that, as an individual, I know when I walk into a room to participate in a focus group, but when I interact with a service, I don't necessarily realize that I'm a subject in a research panel. I have the option of walking into that focus group room and consciously consenting to being a part of the group. When I use a service, I'm starting already inside that room and may not make the same conscious connection. Users should have the ability to leave the room and say they don't want to be part of your focus group.
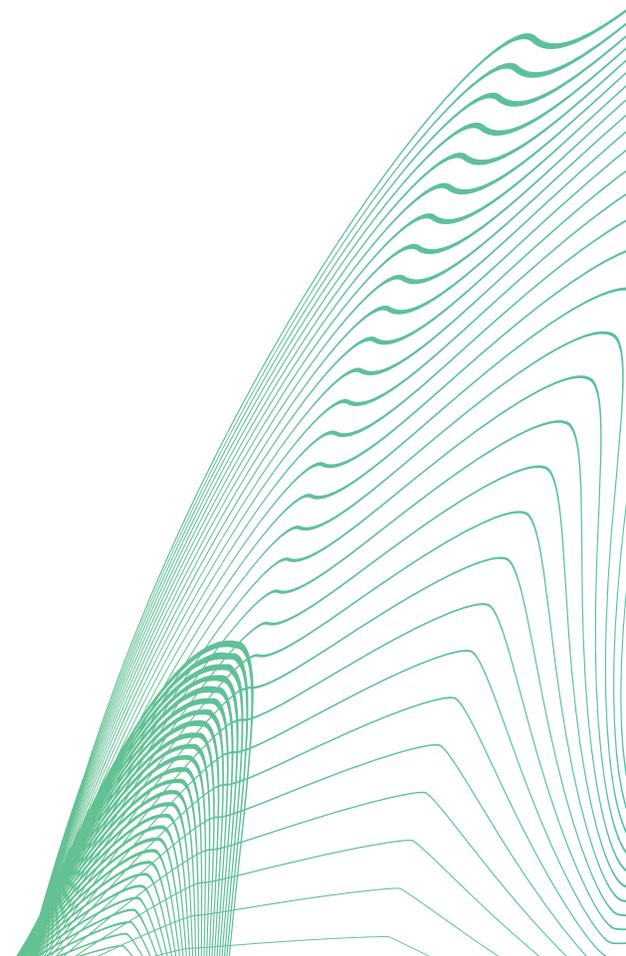
*Interana believes that the users are not secondary to the behavioral analysis conducted on them. Recognizing them as individuals is at the core of who we are. Making their experience with your service the best it can be is why we exist. Understanding their experience with your service is why you use Interana.*

However, not all of your users may feel comfortable with you doing that, so Interana gives you tools to abstract the individual out of the analysis or even forget them altogether. We call this Ethical Behavior Analysis.

## What is personally identifiable information?

Personally Identifiable Information (PII) can mean different things to different people. When GDPR goes into force in May 25, 2018, the EU will impose a rather broad definition of PII upon any business that has EU citizens in its user population, regardless of the borders the business operates within. Part of the EU's (and Interana's) definition of PII is that behavior patterns are just as effective at identifying someone's uniqueness as following the UID assigned to them during registration for a service.

*In The New York Times article A Face Is Exposed for AOL Searcher No. 4417749, the authors describe how, even with a redacted username in a dump of AOL searches, they could identify a user with the ID 4417749 as Thelma Arnold of Lilburn, GA, through her online searches:*

*• "Landscapers in Lilburn, GA" = user resides in community and likely owns their home*

*• Dog pees on everything = user has a dog*

*• 60 single men = user is age 60 or older, likely single, and likely female*

*• Searches for people with the surname Arnold = likely searching for relatives*

*How many single people with the last name Arnold are 60 or older, have a dog, and own a home in the Shadow Lake subdivision of Lilburn, GA? Each of these searches on their own is not terribly effective at identifying a specific individual, but these patterns of search behavior allowed the journalists to quickly unmask the 62-year old dog-loving widow from Georgia.*

# How Interana supports
# ethical behavioral analysis

Many companies in the behavioral analytics space are telling customers to tackle the issue of privacy by omitting, redacting, or anonymizing subject identifiers from event data. Employing one of these techniques often comes with expensive organizational overhead, including updates to upstream event producers, change controls to ensure new sources of PII are not introduced, and constant monitoring to catch when PII falls through the cracks.

Even with all of that, simply dropping the subject identifier in event data does not account for the fact that patterns of behavior are also PII. Further, these methods can help to obfuscate PII, but they do not give users the control they want over their data. This means their guidance is flawed and you will be out of compliance.

With Interana, we are constantly releasing new feature to help support you in respecting your customers' wishes while also lowering new technical overhead.

## **Selective Delete**

Interana's Selective Delete feature allows an administrator to define and schedule granular data retention rules (e.g. keep this PII data source for only as long as the rules allow) or remove events that match a specific filter criteria (e.g. delete any event where username = foo).

However, this feature is not just about rules supporting the ongoing curation of your event data. We recognize that unintended bits of information sometimes find their way into the data stream (see Mixpanel's accidental recording of client's passwords via their SDK), and you can't defend against all of the ways an accident can happen. Selective Delete provides a backstop, so if the event data changes underneath you, you can clean up the change without having to delete all of your data and start over.

## Privacy Purge

Personal information is not confined to the event data. The PII of your customers could find its way into other parts of Interana, such as a global filter, named expression, or dashboard panel. Your users could include some of those personal identifiers in the work they do in Interana. As a user of the system, the queries they run, the times when they access the system, and the dashboards they view are all elements of the user's behavioral activity and subject to the same PII rules.

Privacy Purge leverages the Selective Delete feature to scan all tables across all time for the subject's supplied privacy identifiers. In addition to scanning event data, Privacy Purge also scans the metadata (filters, saved queries, named expressions, structured logging) in order to remove instances of the privacy identifiers. Multiple subject identifiers of different types (user IDs, IP addresses, cookies, etc.) can be passed in a single purge request for efficiency and ease of administration. If there is a need to remove PII for a customer and an Interana user, they can be passed in the same request — the purge makes no distinction between the two. Tell us what needs to go, and it's gone!

## Ethical behavioral analysis prioritizes the user

With today's emphasis on data transparency and control, Ethical Behavioral Analysis gives users a say in how you use their behavior. Interana supports ethical behavioral analysis by accounting for the role of behavior patterns in PII while reducing expensive technical overhead.

In the same way that Patil describes data science as empowering the masses, we at Interana believe in empowering our users - and we hope you'll join us on this journey.
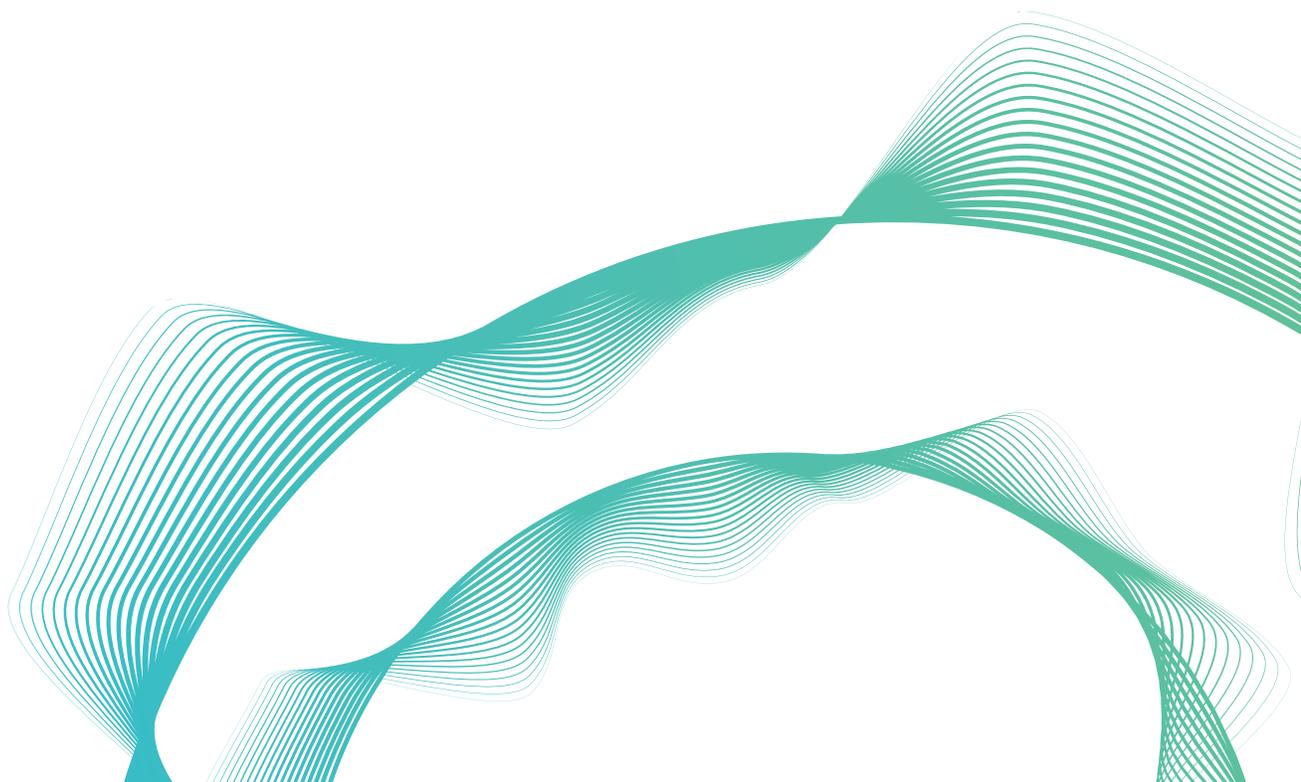
Interana ingests event data produced by all of your applications, services and data pipelines and makes it available for ad hoc exploration via an intuitive, visual interface. They make it fast and easy for product managers, data teams, customer success, and other groups to ask imperfect questions and get answers in seconds—the way humans really think about data.

Define and iterate on your behavioral objects on the fly; no more waiting on your data team.

**Contact Interana today to learn more.**

# inter|ana

Stop Guessing | **Start Knowing**

www.interana.com | 🐦@interanacorp | info@interana.com