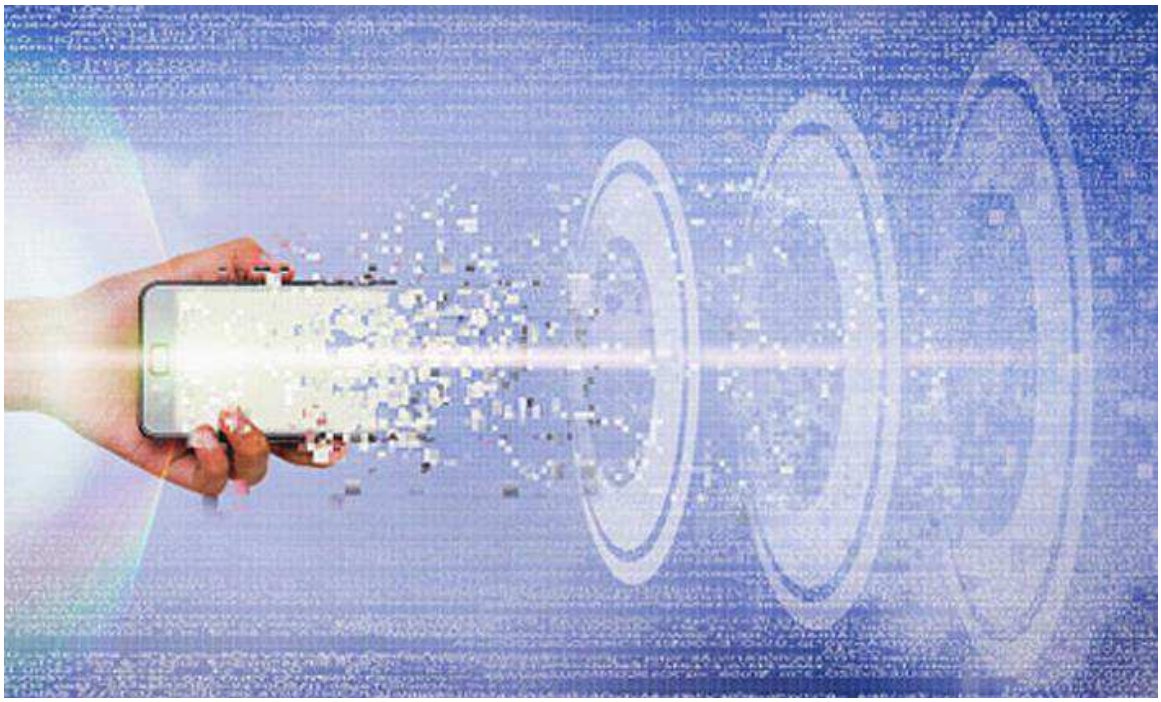
LUTH RESEARCH WHITEPAPER



UNLOCK YOUR IMAGINATION FOR DIGITAL DATA

By Becky Wu, Ph.D.
Sr. Executive Vice President
Luth Research

LUTH research

The definition of digital data can be elusive, depending on whom you talk to. Digital measurement and digital data become an increasingly relevant source of insights for gathering market intelligence and conducting market research. Its definition and scope warrant a closer examination. This whitepaper looks at the common types of digital data in current research practices and aims to delineate where they are different. Equally important, this discussion serves as an explication on the breadth and depth that can be afforded by digital data, much beyond the basic KPIs or URLs marketers and researchers have become accustomed to (and therefore are limited by).

Ways to Collect Digital Data

Before we examine the digital data and metrics, it is necessary to first take a look at the primary approaches in capturing digital data. Generally, when we speak about digital measurement, we refer to the ability to collect consumer activities while they are using the Internet via a computer or using a mobile device. There are other forms of digitized data collection more broadly for connected devices including TV, which are outside of the scope of this whitepaper.

Three approaches dominate today's digital tracking. They are: 1) cookies, 2) web beacons (also called pixel tagging, tracking bugs), 3) proprietary web and mobile tracking solutions based on underlying browser plug-ins, on-device metering, VPN, and/or proxy technologies.

Cookies have been widely deployed by websites and in the advertising world. They are small bits of text downloaded to the browser as the user surfs the Internet. Cookies are primarily used for understanding user behaviors and history, and as a result, also allowing for personalization of the web experience. There are a number of significant limitations and areas for controversy for cookies. For example, cookies expire, leaving holes in tracking the same individual. Mobile apps are not compatible with cookies and third party cookies are increasingly disallowed.

Web beacons are tiny, clear objects being embedded into a web page or an ad, but not visible to the Internet users. When a web page loads with this image, it will make a call to a webserver for the image, which indicates a user has loaded the page. While not fulfilling a comprehensive role, web beacons are often used with or without cookies to track advertising exposure. It is worth noting that web beacons (also known as pixels) are not the same as the concept of beacon, which is for proximity based targeting in the physical world.

Then we have web and mobile tracking technologies that are developed with varying and often proprietary techniques. Luth Research's ZQ Intelligence is a pioneer and leading technology for passively monitoring consumer behaviors across computers, smartphones and tablets.

The Various Types of Digital Metrics

For the purpose of teasing out newer, uncommon yet valuable metrics from the rest, we categorize digital data points by using the criteria of 1) the degree to which the Internet user can directly observe the data transmitted and captured immediately following his or her actions, and 2) the degree to which the data point can be identified, measured and reported out

accurately without any pattern inferences. Why are these criteria relevant? It is because marketers and researchers who use digital metrics and insights often intuitively reference their own personal experience of using the Internet across the devices, when they try to understand and decipher the data points. However, what you see is sometimes not what you get in the world of digital metrics. And, what you can get is much more than what you see.

The first category is the ***foundational digital data metrics***. The leading ones are visit and timestamp for a specific website or app. With these data points, we can then determine further metrics such as frequency and time patterns of visitation. A word of caution is that, while conceptually simple, visit and timestamp (especially end timestamp) are derived constructs because the very act of "using" a website or an app is not a continuous action without interruption. Using any of the aforementioned data collection approaches, the collected digital data typically show a clear start of an action, but leave the end time open ended unless there is a specific end point behavior such as closing the browser. Both visit and timestamp are generally observable to the user as he/she performs a series of visits, despite that the end time sometimes need to be inferred.

Another common set of data points falling into this category are pageview URL and search term. Pageview URL is direct and visible to the Internet user most of the time in the browser's address bar. For example, when you go on Amazon.com and check out their furniture section, you will see a clear pageview URL https://www.amazon.com/Furniture/b/ref=sd_allcat_furn?ie=UTF8&node=1063306. Another example is a pageview URL resulting from playing a video on Youtube.com: https://www.youtube.com/watch?v=7WYmWo28svg. This URL does not provide as much understandable information as the Amazon URL does, but is still something the Internet user sees. Search terms are straightforward. The only complication at times is to examine the search term data while keeping in mind that the user may have chosen a suggested search term instead of typing in the full term.

Next, let's explore the second class of the metrics: the ***inferential metrics***. Duration and data volume are the main representatives for this type of metrics. The metrics are given this name because their measurement would not be accurate without applying a set of computing algorithm, data pattern rules and logics to infer the true value of each metric. One culprit for the necessity of inference is that digital data do not usually have a clean way of demarcating an end point of an action. The other factor is, in the digital world (mobile in particular), it is common for an app or a site to leverage other "helper" processes/apps as part of the mobile operating system, the carrier's infrastructure, the app/site's own technology partners to optimize user experience. The end goal can be a fast loading video, smooth music playing, or a secure transaction. Here is a simplified example. When the YouTube app is being used to watch a video, in parallel to the running of the YouTube app, there could be other "helper" apps running in the background to assist with traffic routing. Without properly accounting for the data volume and time attributed to these helper apps, it is not feasible to determine an accurate measurement of the actual duration and data volume for that particular YouTube app visit.

The above characteristics of digital data require 1) an educated judgement and decision on creating a set of business rules to define a plausible end point, and 2) a systematic, repeatable

and scalable approach in making high-confidence association between the app/site of interest and other supporting processes. These practices are critical in order to report on such metrics.

The last category is ***latent metrics***. What's included here is a rich set of data points, much beyond what we typically think we could generate from digital measurement. The metrics are diverse: the activities users do within a site or an app (e.g., watching a movie, customizing a product, downloading a whitepaper, applying for a credit card, looking at a review, pressing a Like or Share button), ads people are exposed to, purchases, the entire search engine result page after a search term is entered, and so on. We want to focus on this particular category of metrics for the remainder of the whitepaper. While these things are what Internet users know that they are doing, it is uncertain how the captured data are turning out to be, and to what degree each activity can be identified.

Type	Foundational Metrics	Inferential Metrics	Latent Metrics
Metric Examples	Visit	Duration	On-site activities
	Visit timestamps	Data volume	In-app activities
	Pageview & pageview URL		Advertising exposure
	Search terms		Click
			Purchase
			Search engine result page (SERP)
Observability	Directly observable by device users	Often require inferences based on data patterns and parallel processes	Data corresponding to the user behavior is largely invisible from device users
Availability in Digital Insight Products & Services on the Market	Common	Common (but accuracy varies)	Limited to Non-existent

Gateway to Better Digital Metrics

The real test for gaining better digital insights lies in whether or not we can develop the means to harness the latent metrics.

The data that can be used to derive the latent metrics are being collected differently, depending on if you are using cookies or the proprietary digital tracking technologies like ZQ Intelligence™. Using cookies, the data point for the specific user activity (e.g., clicking on a link on a website to complete a registration form) need to be set up beforehand specifically for that user activity in order to enable tracking for the user activity. This requires planning and knowing which activities you want to track, and hence the number of activities is finite. Google Analytics is a prime example of leveraging this approach, allowing companies to have a little bit better

visibility into how certain select features on their websites are performing. However, unless a single type of cookies or coding is applied to all sites for all user activities, it remains not feasible to measure common user activities with adequate breadth and consistency. Despite being the most prevalent cookie based tracking method, Google Analytics is far from being able to produce substantive metrics for in-depth user behaviors. Furthermore, mobile apps are cookies' Achilles heel. Cookies do not work in that environment. When close to 90% of consumers' usage on a smartphone take place on mobile apps, we can no longer ignore the increasingly glaring void that cookies fail to cover.

Alternatively, when using ZQ Intelligence, the collected data include the behind the scene data traffic in the form of http(s) (both secure and non-secure) requests and responses, reflecting the interactions and activities on site and on app, more than what is observable to the Internet user. These http(s) requests and responses are considered the continuous, documented "conversations" between the device and servers. The device (user) loads a webpage or an app, asking the servers for what content to display, what ads to show, mostly depending on the user's activities. The servers push out the content accordingly. If we make a purchase within an app, a distinct https request and response pattern appears. The same goes with other in-app or on-site activities. [1]

The totality of http(s) requests and responses is the source of promise for richer digital metrics. They are the language of the Internet. By collecting and deciphering the communication, we can then develop systematic approaches to measure and report on the target metrics. The potential of mining this data cannot be underestimated.

However, unlocking this promise is not without challenges. The prominent one is the lack of consistent patterns for any given user activity across different apps or sites. This challenge is easy to understand but extremely under-recognized. The purchase patterns for Amazon, Walmart, Bestbuy, Macy's, and other retailers are all unique and prolific. It is necessary to identify and parse out the "code" for the purchase behavior for each individual app and site. In some cases, different devices and carriers also are a reason for pattern variation for the same user behavior. Adding to the complexity, the pattern(s) from different apps and sites can change over time. There needs to be a parallel process to perform regular maintenance and updates for the pattern libraries. These inevitably make building pattern libraries a non-trivial endeavor.

The solution for better harnessing the latent metrics lies in a two-prong approach, which combines intensive pattern testing, often done by human analysts, and rigorous data engineering. Both streams of efforts are particularly effective when coupled with a history of pattern validation for data covering 12 months or longer. Luth's ZQ Intelligence™ provides pattern engines that have been developed based on digital data from more than 24 months. We have seen significant success in extracting pattern libraries for an increasing number of verticals including but not limited to: online TV/video viewing behaviors, ecommerce, social activities, and gaming. These libraries become the foundation of repeatable syndication of deeper digital

[1] There are exceptions where a user activity does not result in a detectable or clear enough data pattern due to reasons such as using a non http Internet protocol and non-standard programming

intelligence. In addition, this above approach provides the method to get at highly custom metrics that brands and companies are seeking for their specific business needs.

Final Thoughts

Digital data collected through passive tracking are supposed to be vast and complex. They are made up by data forms resulting from diverging computing languages, Internet protocols, and communication conventions. These are currently unfamiliar to marketers and researchers. As discussed in this whitepaper, pulling back the curtain on the behind the scene data and operation helps open up our vision for the potential of yielding broader and more versatile insights.

We have long been “conditioned” to associate digital metrics with simpler data points like visit and time spent. Technologies for the Internet of things as well as advertising continue to advance at a rapid rate. It is time for us to develop the understanding so we can go much further in the face of a bit more intricacies and ambiguities. Opportunities abound.