

Data Dictionary V2: Guide for Private Lease Website Tagging

Introduction

The purpose of this guide is to elaborate on and guide users through the Data Dictionary V2 document. The Data Dictionary has been created to give a clear overview of the data being collected with Google Tag Manager (GTM) for the Dutch private lease websites of the brands: Peugeot, Citroen, Opel and Alfa Romeo (more will be added). Previously, when working with data collected in Google Analytics 4 (GA4), a lot of guesswork and human effort was needed to correctly use it, due to a lack of knowledge on how the data collection was implemented (within and between brands). This Data Dictionary creates an overview of the most important events in GA4 and how they are measured, resulting in less guesswork for all people involved with this data.

Background information

Our efforts to document the data collection procedures, started out as a document that mapped all GTM containers and GA4 properties connected to the privatelease.peugeot.nl website. In 2023, data analytics was transferred from Universal Analytics (UA) to GA4. After the transfer to GA4, data driven dashboards started functioning incorrectly. This led to the question where all new data is coming from, and how to properly implement it for high quality insights as before.

During the creation of the first Data Dictionary, multiple problems arose in other projects, as data being collected differed between brands. E.g., when creating KPI performance dashboards, this meant manually figuring out how to use collected events per brand, leading to a lot of man hours spent.

Due to the lack of overview on how data was collected by GTM across brands, the urge arose to create a clear document where all important events, triggers and parameters are described in detail. Resulting in a document to fall back on for any questions about the collected data.

Structure explanation

The Data Dictionary is structured per brand. As is visible in the first column, every brand is given a distinct color to easily be easily identifiable. In the second column, the corresponding site is indicated.

For every brand, the 12 most important events (considering KPIs) are listed and described in detail. These are named as they appear in GA4.

After the event description, the event parameters of the corresponding event are listed (column E). This is the extra information gathered on the website that is sent with the event when it is recorded. Most events use the same set of event parameters which is a list called: *Basic Configuration Event Settings*. When a different set of event parameters is used, these are listed in the next column (column F).

Then, the name of how the event is collected in GTM (called a Tag) is listed (column G). The following column (H) lists all the Triggers names of these Tags. A Trigger is a logical condition that needs to be met before a Tag (and eventually an event in GA4) is recorded. Tags are able to utilize multiple Triggers. The type of Trigger is defined as well (column I). In column J, the name of the event in the dataLayer of the website is listed.

The most important column of the entire Data Dictionary is column K. This column describes the rules (in logic) when a Trigger is fired. Every Trigger can have multiple (lines of) rules that need to be met before the Trigger is fired. The information of these rules that need to be met is saved in the dataLayer of the

website, and corresponds to the previously mentioned event parameters that are included in all recorded events.

Example:

The event *form_start* (in GA4) is used to indicate whether a website visitor has clicked on a button and visited the page where it is possible to fill in a form. This Tag in GTM is called *ANA - Event - GA4 - form_start*. The Trigger related to this Tag is named: *ANA - Form start*. This tag fires if a *form_start* event is received from the website and the following conditions are being met::

1. *url matches RegEx (ignore case).** (URL is not empty)
2. *cj_cookie_checker_ana contains true* (Cookies have been accepted)
3. *dl_formsName does not match RegEx (ignore case) undefined|empty* (The name of the form on the page is not categorized as undefined or empty)
4. *dl_mainStepIndicator equals 1* (The step of the form visited is indicated to be the first step)
5. *dl_formsName matches RegEx (ignore case).+* (The form name should be *.+*, which essentially means 'anything' in RegEx language)

Finally, in column L, it is mentioned from which other brands the corresponding trigger differs. This creates a clear overview of the differences between brands.

Brand Comparison

As previously mentioned, the main goal of creating this Data Dictionary is to accentuate the differences of GTM implementation between brands of the most important events. Keeping the *Different* column (L) in mind, it is clearly visible that the events and Triggers of Alfa Romeo are almost completely different from the other brands.

Example:

When comparing the trigger rules from the event *form_complete* between Peugeot and Alfa Romeo, the differences become apparent. Note: keep in mind that this is only an example, not all triggers from Alfa Romeo are necessarily less detailed. For Alfa Romeo, in this example, a *form_complete* event triggers the *ANA - Form complete* which is tested against the regular expressions below:

Peugeot	Alfa Romeo
<pre>"url matches RegEx (ignore case) .* cj_cookie_checker_ana contains true dl_formsName does not match RegEx (ignore case) undefined empty dl_mainStepName matches RegEx (ignore case) confirmation dl_formsName matches RegEx (ignore case) .+"</pre>	<pre>"url matches RegEx (ignore case) .* cj_cookie_checker_ana contains true"</pre>