

CANONICAL TAGS

A GUIDE TO CANONICALIZATION

ADELINA BOCSE
Paris, 13/02/2017



ARVAL
BNP PARIBAS GROUP

We care about cars. We care about you.

AGENDA

1

EXCEL FILE: CANONICALS SPREADSHEET

2

ETYMOLOGY & DEFINITION

3

SET UP CANONICAL URL_s

4

DUPLICATE CONTENT & PAGINATED CONTENT

5

DO'S & DON'TS



1

EXCEL FILE: CANONICALS SPREADSHEET



Excel file: canonical tags

With respect to the Excel file you received, take into consideration the information you'll find within the **spreadsheet** dedicated to **Canonical Tags**.

By having a look at the following screenshot, you'll see that the **yellow cells** within the « **Canonical matching Online URL** » column indicate « **No** » which reveals that the Canonical URL is different from the Online URL.

The « No » indicated does not necessarily mean that your Canonical tag is not correct.

Please correct the errors according to this guide.

Url	Canonical URL	Checked On	Canonical matching Online URL?
https://www.arval.be/nl	https://www.arval.be/nl/homepage	2017-01-09 22:30:2	No
https://www.arval.be/nl/countries-list	https://www.arval.be/nl/countries-list	2017-01-09 22:44:53	Yes
https://www.arval.be/nl/about-arval/ik-zoek-informatie-over-arval	https://www.arval.be/nl/about-arval/ik-zoek-informatie-over-arval	2017-01-09 22:44:57	Yes
https://www.arval.be/nl/renting	https://www.arval.be/nl/renting	2017-01-09 22:45:09	Yes
https://www.arval.be/nl/i-need-solution-my-fleet/homepage-manage	https://www.arval.be/nl/i-need-solution-my-fleet/homepage-manage-fleet-nl	2017-01-09 22:45:15	Yes
https://www.arval.be/nl/i-need-international-partner/ik-zoek-een-int	https://www.arval.be/nl/i-need-international-partner/ik-zoek-een-internationale-partner	2017-01-09 22:45:19	Yes
https://www.arval.be/nl/i-need-info-about-arval/informatie-over-uw-	https://www.arval.be/nl/i-need-info-about-arval/informatie-over-uw-leasewagen	2017-01-09 22:45:26	Yes
https://www.arval.be/nl/renting/offers	https://www.arval.be/nl/renting/offers	2017-01-09 22:46:38	Yes
https://www.arval.be/nl/renting/offers?manufacturer=70	https://www.arval.be/nl/renting/offers	2017-01-09 22:46:48	No
https://www.arval.be/nl/renting/toyota/c-hr	https://www.arval.be/nl/renting/toyota/c-hr	2017-01-09 22:47:07	Yes

1. It's up to you to check all the "No" cells and decide if the Canonical is correct or not.

2

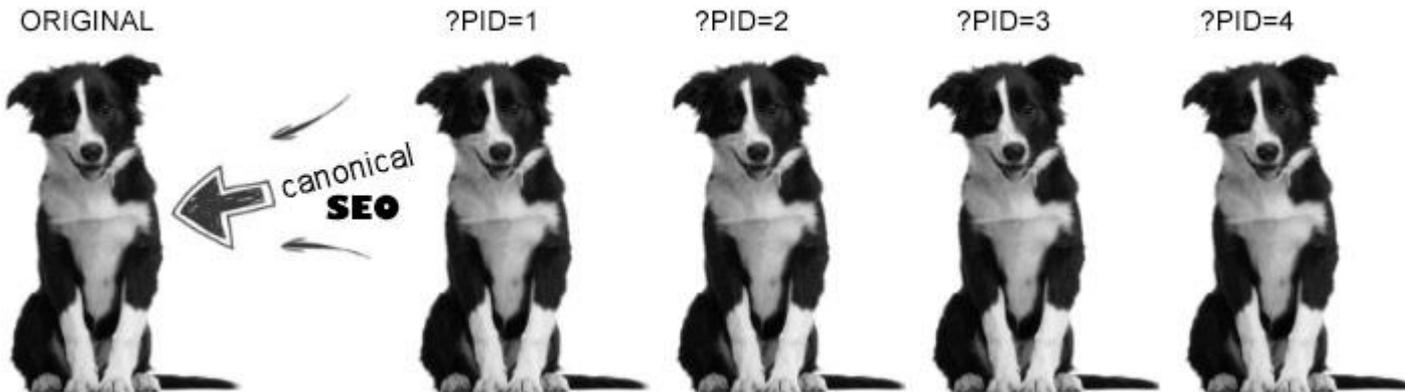
ETYMOLOGY & DEFINITION

Original meaning

Originally *canon* referred to biblical or secular rules and laws, a standard of judgement.

Later, it was used to point to the works of a writer that had been accepted as original (authentic).

It is this last meaning which is related to the internet, SEO and search engines.



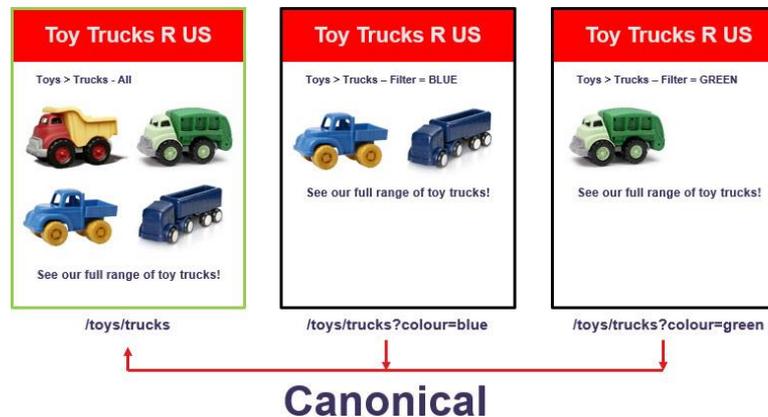
What is a canonical tag?

“A **canonical URL or tag** represents the **preferred version** of a set of pages with **similar content** which can cause **duplicate content** issues.”

Canonical Tags are a powerful way to tell Google and other search engines which URLs you want them to index.

As search engines do not know which version to include or exclude from their indices when the category URL systems create multiple versions of the same page, you should tell them which version is the favorite one by adding it in Drupal or within the header section of the HTML file:

✓ `<link rel="canonical" href="https://www.thepreferredversion.xx"/>`



Let's imagine that...

Google finds on the web three pages with similar content but different page layout/graphic charter.

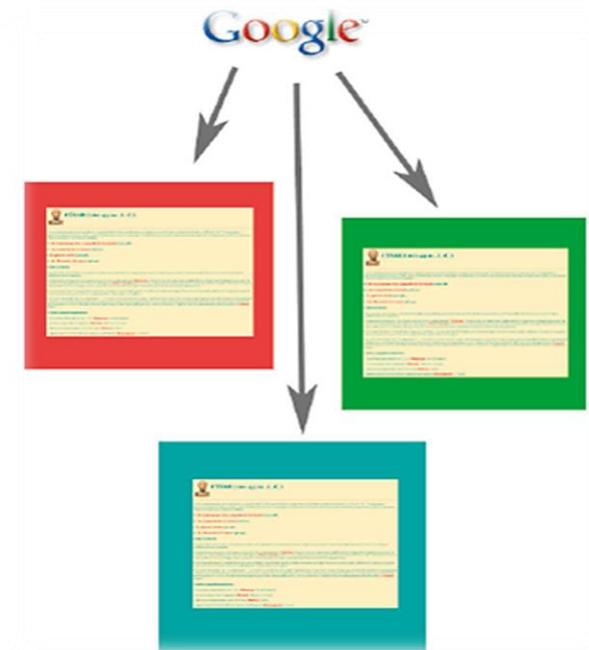


Image Source: Réussir son référencement web, Olivier Andrieu

What is Google doing?

Google will not list all the URLs found but it will choose the original one - the canonical.

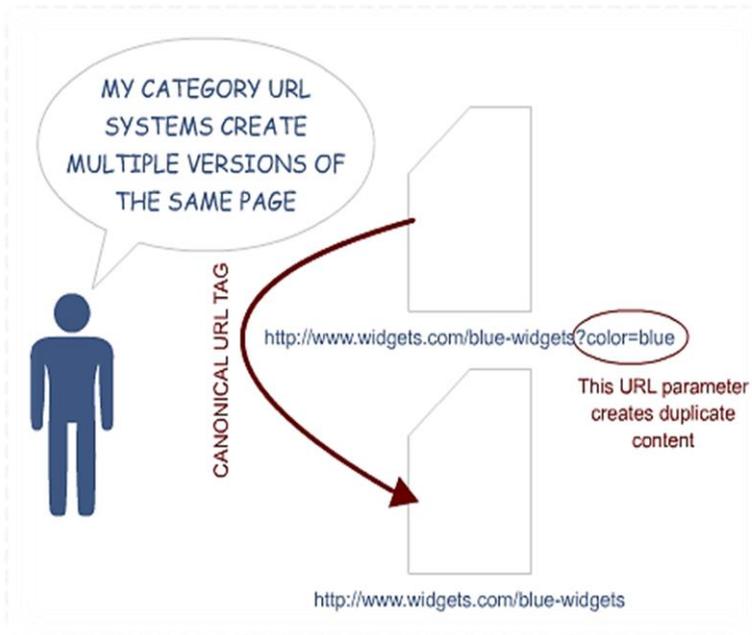


Image Source: MOZ

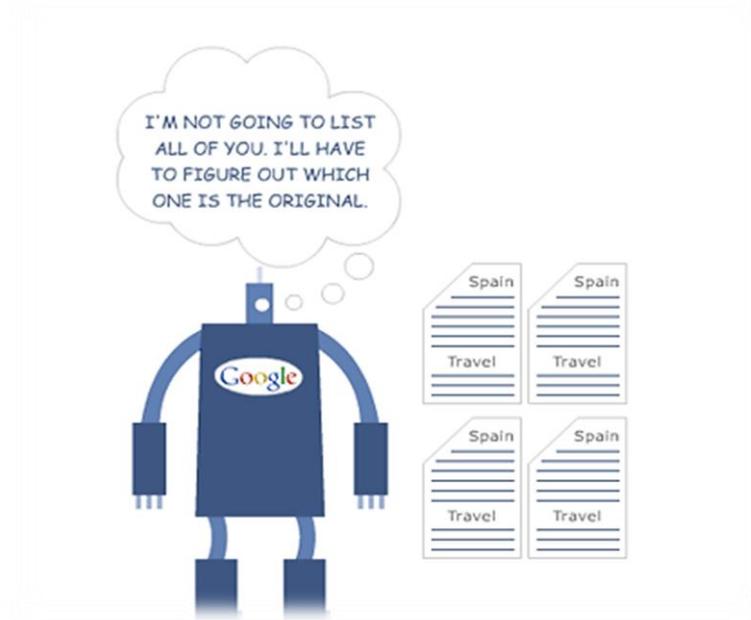


Image Source: MOZ

3

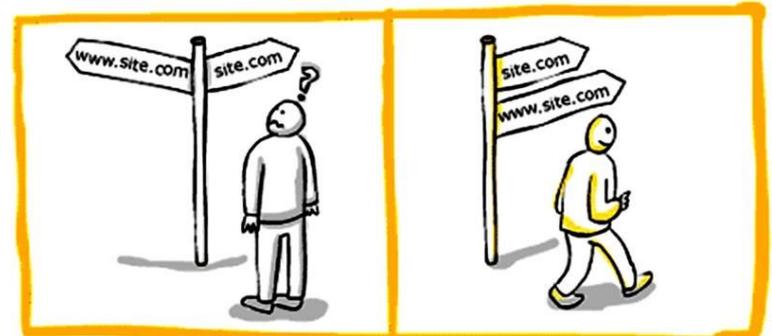
SET UP CANONICAL URLs



How to correctly apply a canonical tag?

Suppose you want the URL below to be the preferred version (even though a variety of URLs can access your content):

- ✓ <https://www.arval.com/en/profile/press-room/element-arval-global-alliance-expands-services-50-countries>



You can indicate this to search engines in the CMS or in the HTML code as follows:

- ✓ <https://www.arval.com/en/profile/press-room/element-arval-global-alliance-expands-services-50-countries>

Absolute vs relative path

“A **path** is a sequence of **directory names**, usually leading to a **file** or a **directory**.”

A full website address includes <https://www>.

An absolute path *a*lways includes the full path to the file.

✓ <https://www.arval.com/en/profile/our-clients> (Absolute Path)

A relative path is only a portion of the full path and it points to a file or a file path.

✓ </en/profile/our-clients> (Relative Path)

It is highly recommended to use absolute paths than relative paths.

➤ *Use this structure:*

✓ <https://www.arval.com/en/profile/press-room/element-arval-global-alliance-expands-services-50-countries> (Absolute Path)

➤ *Not this:*

✓ [en/profile/press-room/element-arval-global-alliance-expands-services-50-countries](/en/profile/press-room/element-arval-global-alliance-expands-services-50-countries) (Relative Path)

4

DUPLICATE CONTENT & PAGINATED CONTENT



What is duplicate content?

As duplicate content is strongly related to canonical tags, it seems necessary to define it.

*“**Duplicate Content** occurs when there are multiple pieces of identical or similar content.”*

In such a case, it is difficult to decide which version is more relevant. So, in order to provide better user experiences, search engines choose the version which is most likely to be the original (canonical) or the best.



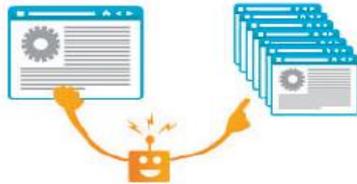
How is duplicate content determined by search engines?

Here is an illustration of how search engines identify duplicate content.

How a Search Engine Determines Duplicate Content

1 Discovers

When content is discovered by a search engine bot, it is compared to everything else that was previously found to determine if it is duplicate content.



2 Discards

First, it discards any page that comes from link farms, MFA sites or blacklisted IPs.



3 Dissects

Next, it dissects each page looking at inbound links, link juice and the quality of the sites from which each link originates.



4 Determines

Lastly, by reviewing the time of discovery and topical links, it determines which page it considers to be the originator of the content.



Image Source:
Réussir son référencement web, Olivier Andrieu

When Google detects duplicate content, its algorithm groups the duplicate URLs into one cluster and selects what the algorithm thinks is the best URL to represent the cluster in search results (for example, Google might select the URL with the most content).

Google then tries to consolidate what we know about the URLs in the cluster, such as link popularity, to the one representative URL to ultimately improve the accuracy of its page ranking and results in Google Search.

Duplicate content: examples

Websites can have duplicate content issues for a variety of reasons:

- **URL Parameters or Multiple URLs** (particularly on e-commerce websites where URLs are created through filter options for price, color, rating, etc.)

An URL Parameter is something which appears after a question mark :

<http://www.widgets.com/blue-widgets?color=blue>

- **Session IDs**

Appending a session ID to the URL creates duplicate content.

<http://www.widgets.com/blue-widgets?SESSID=123>

- **HTTP, HTTPS & WWW**

Search engines see <http://www.mydomain.com>, <http://mydomain.com> and <https://www.mydomain.com> as distinct pages, and will crawl (and possibly index) them as such.

- **Mobile URL**

Using a special URL (typically m.mydomain.com) for the mobile version of your website.

- **Printer-Friendly**

Printer-friendly versions of content can cause duplicate content issues when multiple versions of the pages get indexed.

<http://www.widgets.com/print/blue-widgets>

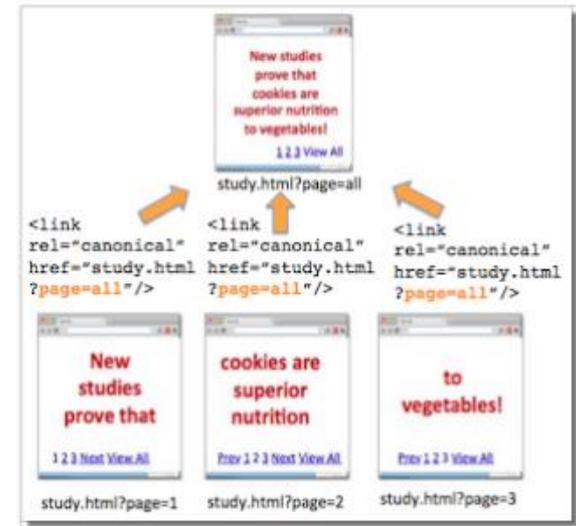
Putting this in the « print » subcategory creates duplicate content.



Paginated content

When it comes to paginated content, you have three options:

- **Specify a *View All page*.** As searchers usually prefer to view a whole article or category on a single page, you can add a **rel="canonical"** link to the component pages to tell Google that the View All page is the version you want to appear in search results.
- **Use *rel="next"* and *rel="prev"* links** to indicate the relationship between component URLs. Thus, Google treats these pages as a logical sequence by consolidating their linking properties and usually sending searchers to the first page.
- **Do nothing.** As paginated content is common and Google returns the most relevant results, you are supposed to do nothing and let Google do the work. Paginated content is common and Google tries to return the most relevant results to users, regardless of whether content is divided into multiple pages.



For more information on paginated content, check the following link:

<https://support.google.com/webmasters/answer/1663744?hl=en>

5

DO'S & DON'TS



Canonical URL for the home page

DO  home page URL: <https://www.arval.be/nl>
canonical URL: <https://www.arval.be/nl>

DON'T  home page URL: <https://www.arval.be/nl>
canonical URL: <https://www.arval.be/nl/homepage>

NOTE:

- Verify the home page Canonical URL in Drupal to see if it is correctly configured.
- Please have a look at the next slides to see how to configure the Canonical tag for the home page in the back office.



```
1 <!DOCTYPE html>
2 <!--[if IEMobile 7]><html class="iem7" lang="nl" dir="ltr"><![endif-->
3 <!--[if lte IE 6]><html class="lt-ie9 lt-ie8 lt-ie7" lang="nl" dir="ltr"><![endif
4 <!--[if (IE 7)&!IEMobile]><html class="lt-ie9 lt-ie8" lang="nl" dir="ltr"><![en
5 <!--[if IE 8]><html class="lt-ie9" lang="nl" dir="ltr"><![endif-->
6 <!--[if (gte IE 9)|(gt IEMobile 7)]><!--><html lang="nl" dir="ltr" prefix="og: ht
7 http://ogp.me/ns/product# content: http://purl.org/rss/1.0/modules/content/ dc: ht
8 http://www.w3.org/2004/02/skos/core# xsd: http://www.w3.org/2001/XMLSchema#"><<
9 <head>
10 <meta charset="utf-8" />
11 <link href="https://www.arval.be/fr" rel="alternate" hreflang="fr" />
12 <link href="https://www.arval.be/nl" rel="alternate" hreflang="nl" />
13 <link href="https://www.arval.be/sites/all/themes/arval_be/fav" rel="shortcut icon" href="https://www.arval.be/sites/all/themes/arval_be/fav" />
14 <link href="https://www.arval.be/en" rel="alternate" hreflang="en" />
15 <meta http-equiv="X-UA-Compatible" content="IE=edge" />
16 <meta name="generator" content="Drupal 7 (http://drupal.org)" />
17 <link rel="canonical" href="https://www.arval.be/nl/homepage" />
```

1. We suggest using Canonical Tags within every page, setting them as auto-referring tags. Anyway, Drupal should already have this option enabled by default.
2. The Canonical URL for the home page should always match the Online URL.
3. It is usually a major error if the home page of your website contains a different Canonical URL in respect of the Online one.

How to set your homepage canonical URL in Drupal

- *In order to modify the canonical URL of your home page, go in the back office and click on « **Edit** » (be sure to have chosen the correct page; for instance here is the home page you have to go on to see or change its canonical URL).*

The screenshot shows the ARVAL back office interface. At the top, there is a navigation menu with items: Manage Content, Add Brand, Manage Structure, Offers Reviews, Configuration, and Shortcuts. Below the menu is the ARVAL logo and the tagline "We care about cars. We care about you." The main content area features a search bar with the text "I need" and a dropdown menu showing "to see car lease deals". Below the search bar, there are four buttons: View, Edit, Replicate, and Translate. The "Edit" button is highlighted with a red box. A green arrow points from the word "Select" above the buttons to the "Edit" button. Below the buttons, there is a "FIND YOUR CAR" section with dropdown menus for TYPE (set to CAR), MAKE (set to Please select), and MODEL. To the right of this section is a "SPECIAL OFFER" for a VAUXHALL INSIGNIA Diesel Hatch 1.6 CDTi ecoFLEX Tech Line 5dr. The offer price is £151 + VAT PCM, with terms: 36 Months | 10,000 miles p.a. | 6 payments upfront. A small image of a blue car is shown below the offer details.

Canonical URL: « Edit Homepage Homepage »

- By clicking on Meta Tags, you have access to « Basic tags, Advanced tags or Twitter card ».
- To see how the preferred URL of the home page is configured, choose Advanced tags.

The screenshot shows the Drupal administration interface for editing the homepage. The breadcrumb trail is: Home > Manage Content > Add Brand > Manage Structure > Offers Reviews > Configuration > Shortcuts > Edit Homepage Homepage. The page title is "Edit Homepage Homepage". The navigation menu includes View, Edit, Replicate, and Translate. The left sidebar contains several sections: "Homepage *", "Menu settings", "XML sitemap", "URL path settings", "Meta tags", and "Publishing options". The "Meta tags" section is highlighted with a green box and a green arrow pointing to it from the text "Select this". The "Meta tags" section contains the following information: "Page title: Car and Vehicle Leasing | Arval UK", "Description: A part of BNP Paribas Gro...", "Canonical URL: [site:url]", and "Description: Arval". The right sidebar shows the "Meta tags" configuration area. It includes a "Basic tags" section with a "Page title" field containing "Car and Vehicle Leasing | Arval UK" and a "Description" field containing "A part of BNP Paribas Group we provide unrivalled customer service in the car and vehicle leasing sector for customers of all sizes. | Arval UK". Below these are "Keywords" and "Advanced tags" sections. The "Advanced tags" section is highlighted with a green box and a green arrow pointing to it from the text "Click on « Advanced tags » to display all the options".

Select this

Click on « Advanced tags » to display all the options

How to indicate the Canonical URL: [site:url]

- To see if the canonical URL for the home page is correctly set, look at the field corresponding to it and be sure that you have [site:url] indicated in it. If not, add it to the field and then click on the save button.

▼ Advanced tags

Browse available tokens.

Robots

- Allow search engines to index this page (assumed).
- Allow search engines to follow links on this page (assumed).
- Prevents search engines from indexing this page.
- Prevents search engines from following links on this page.
- Prevents cached copies of this page from appearing in search results.
- Prevents descriptions from appearing in search results, and prevents page caching.
- Blocks the [Open Directory Project](#) description from appearing in search results.
- Prevents Yahoo! from listing this page in the [Yahoo! Directory](#).
- Prevent search engines from indexing images on this page.
- Prevent search engines from offering to translate this page in search results.

Provides search engines with specific directions for what to do when this page is indexed.

Google News Keywords

A comma-separated list of keywords about the page. This meta tag is used as an indicator in [Google News](#).

Canonical URL

Preferred page location or URL to help eliminate duplicate content for search engines.

« [site:url] » 
to add if necessary

Canonical URL on pages with similar content

DO



- page URL: <https://www.arval.be/en/mid-term-rental?data-duration=12>
- canonical URL: <https://www.arval.be/en/mid-term-rental>

DON'T



- page URL: <https://www.arval.be/en/mid-term-rental?data-duration=12>
- canonical URL: No canonical set in the back office

NOTE:

- This is a good illustration to take as an example when you set Canonical tags.
- There is no modification to do in the back office for the above example.

<https://www.arval.be/en/mid-term-rental?data-duration=12>

<https://www.arval.be/en/mid-term-rental>

1. When you have two or more pages with similar content as in this case, please use a Canonical tag.
2. The Canonical tag you choose should be the preferred version/page you want Google to index it.

Wrong webstore canonical URL linking to the home page

DO



- page URL: <https://www.arval.fr/pro/professionnels-et-entrepreneurs>
- canonical URL: <https://www.arval.fr/pro/professionnels-et-entrepreneurs>

DON'T



- page URL: <https://www.arval.fr/pro/professionnels-et-entrepreneurs>
- canonical URL: <https://www.arval.fr>

NOTE:

- Here are the countries concerned by this issue : FR/IT/NL/UK.
- This example concerns only the Canonical URLs set within the webstore.
- The preferred version of the “professionnels et entrepreneurs” cannot be the home page of Arval France website.
- Revise all your pages to see if any modification is needed.
- If needed, please contact Ohini, the person in charge of the webstore, and communicate him what to do.

Page URL

```
1 <!DOCTYPE html> <html class="aui ltr" dir="ltr" lang="fr-FR"> <head> <title>Location lc
equiv="content-type" /> <meta content="index, follow" lang="fr-FR" name="robots" /> <me
name="description" /> <meta content="lld, location longue durée, financement, voitures
content="Location longue durée, LLD de voiture et leasing auto - Arval"/> <meta propert
entrepreneurs." /> <meta name="twitter:title" content="Location longue durée, LLD de vc
Arval pour les professionnels et les entrepreneurs." /> <meta property="og:image" conte
rel="canonical" href="https://www.arval.fr" /> <meta property="og:url" content="https:/
France" /> <meta property="og:type" content="website" /> <meta name="twitter:card" cont
name="WT.z_clienttype2" content="null"> <meta name="WT.z_clienttype3" content="null"> <
content="null"> <meta name="WT.site" content="null"> <meta name="WT.z loggedin" content
```

Canonical URL

1. All the canonical URLs do not have to point to your homepage otherwise you risk having none of the pages crawled or indexed by search engines.
2. It is important to add the Canonical tag on both pages: on the preferred one as well as on the non-canonical one.
3. Do not forget that you have to use a Canonical tag only if your page is similar to another one or if your page is accessible through multiple URLs.

Canonical URL: no question mark at the end

DO  page URL: <https://www.arval.fr/location/BMW-BRA010>
canonical URL: <https://www.arval.fr/location/BMW-BRA010>

DON'T  page URL: <https://www.arval.fr/location/BMW-BRA010>
canonical URL: <https://www.arval.fr/location/BMW-BRA010?>

```
narque\x253Fbrand\x253DBMW-BRA010";</script> <script src="/html/js/barebone.jsp?
browserId=other&themeId=Arval_Webstore_WAR_Webstoretheme&colorSchemeId=01&minifierTy
type="text/javascript">Liferay.Portlet.list=["arvalwsearchbasic_WAR_arvalwspportlet_INSTAN
browserId=other&minifierType=js&languageId=fr_FR&b=6210&t=148607298100" type="t
property="og:title"/> <meta content="Choisissez parmi tous les modèles BMW votre véhicule en l
name="twitter:title"/> <meta content="Choisissez parmi tous les modèles BMW votre véhicule en l
name="twitter:image"/> <link href="https://www.arval.fr/location/BMW-BRA010?" rel="canonical"/>
class="lfr-css-file" href="https://www.arval.fr/location/BMW-BRA010?theme&css&main&css&browserId&other&themeId&Arval_Webstore_WAR
type="text/css">tabs-container a{color:#999!important}.header_logo{margin-top:17px}.sub-header-
slider-container{width:430px!important}.phone_input.is_mandatory.is_integer{height:20px!importar
select{width:47px!important}div#voiture1.comparer_left.col-comp-left{height:480px}div#voiture2.<
type="text/css">#p_id_arvalwsearchbasic_WAR_arvalwspportlet_INSTANCE_fRoS5aJlSAJP_ .port.
<!--[if IE 9]> <link rel="stylesheet" href="https://www.arval.fr/Webstore-theme/css/webstore/ie
```

Canonical URL

Not ok to have a question mark at the end of the canonical tag

1. The Canonical URL you choose to be the preferred version should be the same as the online one without any parameter added to it.
2. The Canonical URL should perfectly match the URL you want to indicate as the favorite one.
3. No question mark/parameter added at the end of your URL.

MERCI

E-mail: digitalteam@arval.com
Digital Team