

# Wurth Data Scrubber

## Logon details

- Customer Number: 281910
- Partner Number: 2106677
- Password: Arsenal

## Url

The URL's have a simple fixed pattern, with the SKU being required as part of it.

<https://eshop.wuerth.de/0/<SKU>.sku/en/US/EUR/>

Replace the above red text with the local SKU, after removing the first 0.

## Validate

Ensure the page is a valid product before continuing.

Look for `#body > #main-wrapper.product/ProductNotAvailable`. If it is present, product is not valid.

Or confirm that the following is present: `#body > #main-wrapper.product/ProductPage`.

## Scrubbing

### Category

`#body > .main > .container-fluid > .breadcrumb > .breadcrumb-item:nth(2) > a > text`

`#body > .main > .container-fluid > .breadcrumb > .breadcrumb-item:nth(3) > a > text`

Get the second and third breadcrumb item. Connect them into a single Category Path such as  
/Technical chemicals/Lubricants and rust removers

### Image

`#body #detailImage img.img-fluid[src]`

Get the src value from the detail image.

### Specification (Full Description)

`#body #blockProductInfo #productDetailCenterDescription`

Get the full description from the product detail center description div. Contains html content

### Application (undetermined)

`#body #blockProductInfo #productDetailCenterApplicationDescription`

Get the application from the product detail center application description. Unsure where to store.  
Not always present.

### Technical Information (attributes)

`#body #blockProductInfo #productDetailCenterTechnicalInformations table tbody`

Each tr contains a single attribute.

Attribute set will be "Technical Information".

Attribute is first column.

Value is second column.

## Documents

Doesn't seem to be a simple way to do this. Inconsistent.

Check for text "Safety data sheet", get all pdfs linked to that text.

Check for <a> with text "Technical data sheet". Get each second one?