

Product Data Index Quality

Status

ON-HOLD

In recent months, we have experienced enormous growth in the number of product and price changes that we process on a daily basis. To give you an idea, we currently process > 7.5 MIO products > 70 MIO prices per day between 23:00 and 08:00 and this number is still growing. We have therefore decided to re-develop the Product Exchange service towards the Product Data Store (import) so that we become more scalable to handle the future workload. There is also a user story in the product backlog to adjust the brands and product group matching so that they can be used in the suspecting algorithms you devised. Our development resources are limited and we expect to be working on this for the rest of the year and would like to work with you again in 2022.

Introduction

The current way of indexing the products has its limitations., sometimes wrong products are linked to a Product Data Index product, resulting in an inability to find the right products.

Secondly, the classifications and branding of the products can be improved.

Some ideas are to develop logic to identify whether the products belonging to a Product Data Index product are really the same.

Another idea is to use the database with 2.5 MIO products with the manual knowledge of classification inside, to extract that knowledge with machine learning so not the user has to decide on a classification mapping, but the machine does.

The idea

The idea, in general, is to identify suspects by comparing GTIN's, brands and manufacturer part numbers from the linked supplier items with each other. The suspects will be displayed in the portal to the suppliers for their own products, the portal will also show the products that failed processing and not available in the Product Data Store. They can either correct their internal product data, so with the next processing, it will be correct, or mark the product to be correct. In that case, another supplier has probably to correct the product data. Also, for the buyers, a portal will be available to display the suspects. And on top of that, the suspect flag will be available in the product data output and in the real-time Product Now.

Please find below a table for illustration with the processing errors and suspects, next to that is the user stories describing the suspect identification, supplier and buyer portal.

Sample processing errors and suspects

SCENARIO	SELLER	SKU	GTIN/EAN	MFPR	BRAND	PRICE	REQUIRED OUTPUT
Valid product	A	A	A	A	A	€ 100,00	Valid product
	B	B	A	A	A	€ 101,00	Valid product
Duplicate SKU	A	A	A	A	A	€ 100,00	
	A	A	B	B	B	€ 200,00	Report duplicate SKU to the supplier, product not available
Duplicate MFPR	A	A	A	A	A	€ 100,00	
	A	B	B	A	A	€ 200,00	Report duplicate MFPR to the supplier, product not available
Duplicate GTIN	A	A	A	A	A	€ 100,00	
	A	B	A	B	B	€ 200,00	Report duplicate GTIN to the supplier, product not available
Duplicate GTIN/MFPR	A	A	A	A	A	€ 100,00	
	A	B	A	A	A	€ 200,00	Report duplicate GTIN/MFPR to the supplier, product not available
Brand difference	A	A	A	A	A	€ 100,00	Report suspect on brand, if correct, then report to other suppliers
	B	B	A	A	B	€ 101,00	Required in output with suspect flag
	C	C	A	A	B	€ 102,00	
GTIN difference	A	A	A	A	A	€ 100,00	Report suspect on GTIN to PDM, check and add to PDI or reject
	B	B	B	A	A	€ 101,00	

MFPR difference	A	A	A	A	A	€ 100,00	Report suspect on MFPR, required in output with suspect flag
	B	B	A	B	B	€ 101,00	

User stories

Suspect identification

As a Product Exchange user

I want matched products to be compared and suspects are identified

So that the buyers don't get different products in their search results

Acceptance criteria:

- Compare matched products which each other and flag as suspect:
 - when the GTIN differs between the sellers
 - Multiple GTIN's on PDI level can exist (different story)
 - when the brands between the sellers differ
 - when the MFPR between the sellers differs
 - Check on GTIN
 - Check on brand
 - Check on MFPR

Background:

Products are matched on GTIN or MFPR and brand. When the GTIN or MFPR is invalid, this can result in different products for one GTIN or MFPR. These can be identified as suspects by comparing prices and brands.

Supplier dashboard

As a Product Exchange seller

I would like to, in case of suspects and errors as described in user story 1 and user story 3, be notified by mail and in the portal

So that we can monitor our product data quality and integrity to improve the searchability of our products

Acceptance criteria:

- Only the own products of the seller should be visible
- Email:
 - How many errors/suspects per flag-type
 - Date - time stamp of oldest error and flag
 - Number of new flagged or error products
 - Send once a day
 - With a link to the dashboard
- Dashboard:
 - Filter on headers
 - Sort on headers
 - Show all items with error or suspect-type
 - Show the number of days/date
 - Headers of the columns: SKU/MFPR/GTIN/Brand/Upload date/reason of error or suspicion/Flagged by (system, user)
 - Export to Excel with the current selection (filtered output)
 - Unflag products because the product data is right, multi-select if the filter is used
 - Additional user role required
- Product exchange processing:
 - Flag's remain, unless the suspicion has been resolved in either the data or unflagged in the portal
 - The flag will not return if the product is unflagged and the data is unchanged
- Unflagging:
 - Multiple brands can be the same
 - Products can have multiple GTIN's
 - Other
 - If more sellers are unflagging a suspicious product, then Onetrail should be involved
 - Optionally specify textual reason when unflagging
- History for unflagged products (nice2have in the dashboard, mandatory for backend)
 - Keep track of the history of the unflagging
- In case of duplicate error (single seller, different SKU's), the seller can select the single product to process
- Check master-file against cust specific file. if not in master and in customer file then show error

Buyer dashboard

As a Product Exchange buyer

I would like to, in case of suspects and errors as described in story 1 and 2, be notified by mail and in the portal

So that we can monitor the data quality of our suppliers and Onetrail

Acceptance criteria

- Email:

- Per seller the number of errors and suspects
 - Send once a day
 - Overview of conflicts which Onetrail is checking
- Dashboard:
 - Filter on headers
 - Sort on headers
 - Headers of the columns: Seller Name/SKU/MFPR/GTIN/Brand/Upload date/reason of error or suspicion/Flag-type (system, user)
 - Only show flag type system + buyer itself (not other buyers)
 - Filter in search bar to hide suspects or errors;
 - Export to Excel with the current selection (filtered output)
- Product Viewer
 - Show flag and reason in Product Viewer
 - Able to set a flag on a product in the Product Viewer with suspect-type and reason (text), only available for limited users (user role)
 - Flag-type= User, only visible for the buyer who has set the flag and for the flagged sellers;
- Product Now:
 - Add flag, reason and flag-type (user/system) to Product Now and Product data batch API
- Product Exchange output:
 - Filter on the suspicious flag (extension to Product Filter)