Metadata Report

Lukasz Malucha

METADATA REPORT

INTRODUCTION

This report provides information obtained through the process of collection and analysis of Tech Comms documentation data from years 1955-2020. The report will mainly focus on metadata issues, inconsistencies observed and their impact on Fluid Topics catalogue and Knowledge Exchange project.

DATETIME FIELDS

Various DateTime Formats

Among all the metadata categories, there are **4** datetime columns: meta_lastEdition, meta_lastPublication, meta_revised_modified and meta_created_date. Across those fields, date can be found in at least 6 different formats:

- 07-17-2017 (mm-dd-yyyy)
- 08/31/2012 (mm/dd/yyyy)
- 19-09-2006 (dd-mm-yyyy)
- 1999-08-11 (yyyy-mm-dd)
- 2-5-2017 (d-m-yyyy)
- 2020-01-30T16:19:36.557600 (yyyy-mm-ddTHH:MM:SS)

Impact:

Variety of data formats makes data aggregation by date (i.e. monthly report) impossible without applying custom function. It aslo has an impact on searchability by date.

Suggestion:

Apply only one date format that is easily understandable across the globe:

- YYYY-MM-DD
- YY-MM-DD

METADATA REPORT

SERIES FIELD

Different Series Kept Together:

3 to 12.5 Ton, Coleman® Point™ Core, Core, Fraser-Johnston® Relia™ Core, Luxaire® Optimum™ Core, TempMaster® Omni™ Core, York® Sun™ Core

Product series are listed together. Series may belong to different brands (as in example above) or to unrelated product lines.

Impact:

Keeping product series together is a missed opportunity for creation of valuable search criteria in the future. Issue also closely related to merging various brands into one field.

Suggestion:

Create one record per product series. In a future it will allow to easier transform that metadata field into search criteria.

BRAND FIELD

Different Brands Kept Together:

Champion, Coleman, Fraser-Johnston, Johnson Controls, Luxaire, TempMaster, YORK

Various Product brands are listed together.

Impact:

Keeping different brands together adds difficulty when creating brand-specific page in Knowledge Exchange.

Suggestion:

Create one record per product brand. In a future it will allow to easier transform dataset into brand-specific pages in Knowledge Exchange.

METADATA REPORT

LINK/URL FIELDS

Easy-to-generate links kept in metadata:

Among the metadata fields, there are 4 url link fields: readerUrl, rightsApiEndpoint, topicsApiEndpoint and attachmentsApiEndpoint, that don't store any valuable info and can be easily recreated within html by adding /attachement or /rights to document id.

Impact:

This extra columns don't store any information while impacting database performance.

Suggestion:

Delete them and create those links via html page.

CATEGORY FIELD

String Errors:

- Various Product brands are listed together.
- " , and " instead of comma
- "EAS Tag / Label " and "EAS Tag/Label" present
- "(other choices available)" present

METADATA REPORT

DOCUMENT REVISION FIELD

String Errors:

- "(other choices available)" added to the field
- "So I can just" added to the revision field
- Three types of hyphen present in a data: - , —, –

BUSINESS FIELD

String Errors:

- (other choices available) added to the field

DOCUMENT TYPE FIELD

String Errors:

- (other choices available) added to the field

META_AUDIENCE_TYPE FIELD

String Errors:

- Different capitalisation for the same type (user-User, public-Public)

METADATA REPORT

PRODUCT FIELD

String Errors:

- "(other choices available)" added to the field
- " , and " instead of simple comma

Different Products kept together:

Product field lists multiple products together. Issue closely related to Brand and Series fields issue with keeping not related data together.

Impact:

Keeping different products together adds difficulty when creating product-specific sub-page in Knowledge Exchange.

Suggestion:

Create one record per product. In a future it will allow to create product-specific documentation sub-page.

METADATA REPORT

PRODUCT CODE FIELD

String Errors:

- (other choices available) added to the field
- " , and " instead of simple comma

Erratic Ways Of Multiple Code Listing:

There are variety of examples for erratic multiple code listing:

- ZDT03 to ZDTA6 – makes ZDT04 and ZDT05 products not searchable
- J07NL to J25NL/J10NM to J20NM – not clear, lot of products not searchable
- PC/PD/PE090 to PC/PD/PE240 – it's three product series merged together
- ZKTA3 to T12 – which should be ZK-A3, ZK-A4, ZK-A5, ZKT01 to ZKT12
- ZV -A3 to -12 – which represent ZV-A3, ZV-A4, ZV-A5, ZV-06 to ZV-12

Impact:

Makes large amount of documentation not searchable by specific product code.

User needs to know lower and upper bound of product series. Also it has a heavy impact on quality of filtering.

Suggestion:

Every product code should be listed separated by comma.