# Phase III: Verification
# Database Assessment

REFERENCE DOCUMENT

by
Marhendra Lidiansa
l.marhendra@student.tudelft.nl


Web Information Systems Research Group
Faculty EECMS, Delft University of Technology
Delft, The Netherlands
www.ewi.tudelft.nl


Elsevier B.V
Strategy, Global Operations
Amsterdam, The Netherlands
www.elsevier.com

## Table of Contents

## List of Figures



## List of Tables

# Document Control

**Document Version : 1.0**

**Version history**

| Version | Final/Draft | Release date | Author | Comments |
|---|---|---|---|---|
| 0.1 | Draft | 1-Apr-2014 | Marhendra L | No previous document |
| 1.0 | Final | 21-May-2014 | Marhendra L | • This document is a reference document for the Main Report<br>• This Document describes Database Assessment in the phase III: Verifications |

# 1 DQ Metrics Requirements

## 1.1 Requirements for Data Quality Metrics

The database assessment activity is conducted to assess the requirements fulfillment of each specified data quality metrics. The requirements for data quality metrics that are affected by the database assessment are in Table 1.

Table 1 DQ Metrics Requirements

| No | Requirement | Code | Description |
|---|---|---|---|
| | Generic | | |
| 1 | Acceptability | DQ-R-03 | Base the determination of whether the quality of data meets business expectations on specified acceptability thresholds |
| | Methods related | | |
| 2 | Feasibility | DQ-R-07 | It is also required that the measurement procedure can be accomplished at a high level of automation. |
| 3 | Reproducible | DQ-R-08 | To produce consistent measurement results and to understand any factors that might introduce variability into the measurement |
| | Value and Methods | | |
| 4 | Aggregation | DQ-R-09 | The metrics must allow aggregation of values on a given level to the next higher level. |

Particularly for Acceptability (DQ-R-03), an important function of the database assessment is to determine the threshold values for the measurements by having an assessment against the sales data and Google Analytics data. This research determines the threshold value using 2 approaches:

i. Manual :
   a. Using a value on the basis of the importance of the field and data quality attributes in the application. This is for accuracy related measurements.
   b. Assessment result against the Sales and Google Analytics data

ii. Using assessment result that is considered as historical data for the measurements. An example is using the mean and standard deviation value from 3 or more Data Quality Assessment activities (Sebastian-Coleman [1]).

# 2 Database Structure Assessment

## 2.1 Book and Journal Entities

We assess the structure for Book and Journal in e-commerce using documents including CALoader.xml, pages in e-store.elsevier.com, e-store documentation, IFS documents, and the tables in ATG. The book and journal entities based on site's page and CALoader.xml are as depicted in Figure 1 and Figure 2. The list of entities and attributes are in Table 2

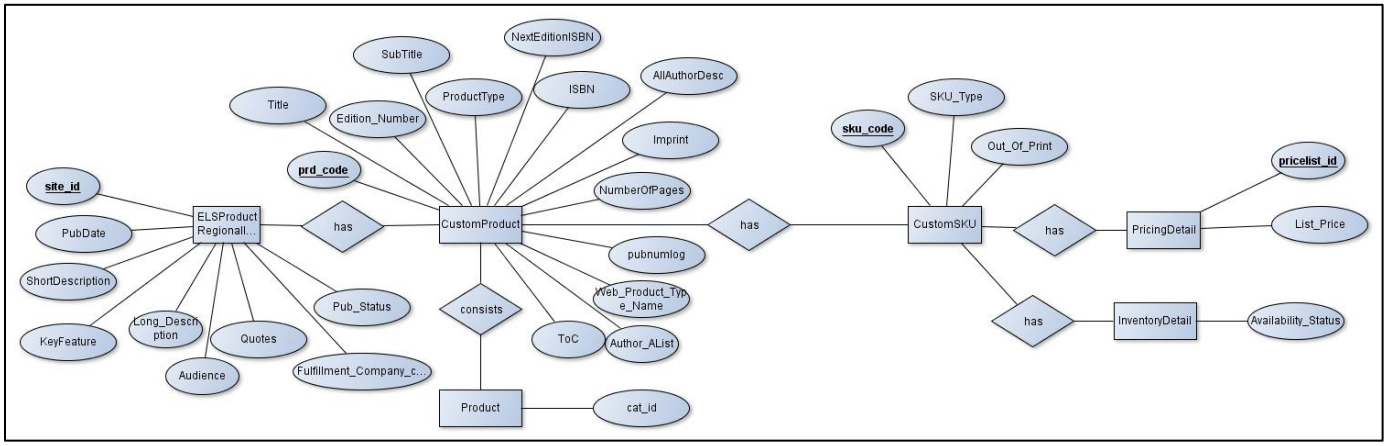


Figure 1 ER Book

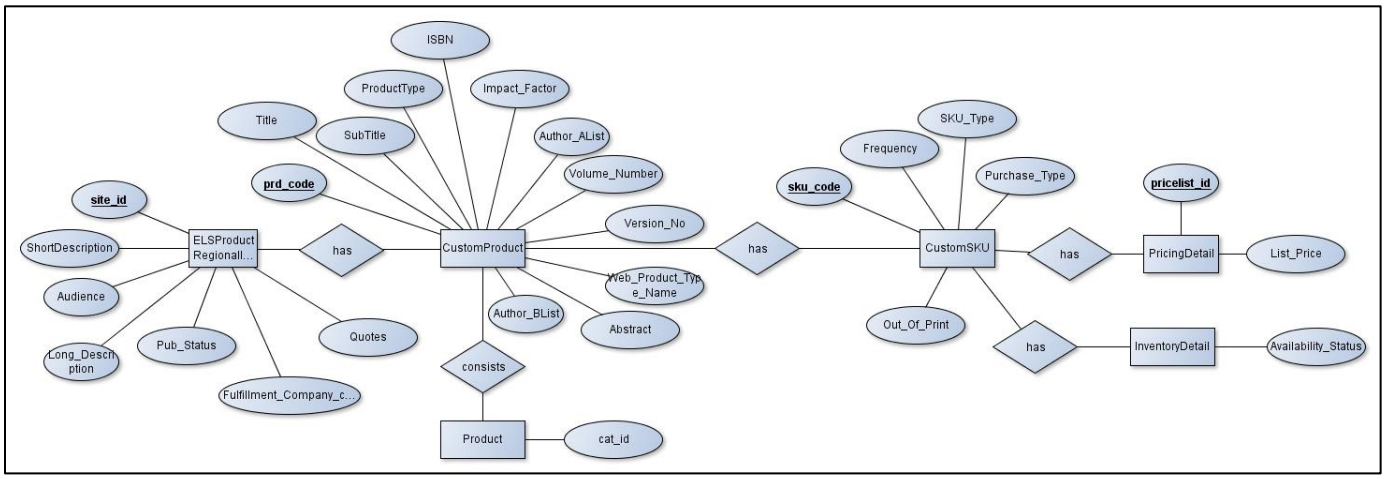


Figure 2 ER Journal

Table 2 Product e-store-CALoader

| Book (Web) | Journal (Web) | CALoader.XML |
|---|---|---|
| Title and Edition | Title | CustomProduct [Title]<br>CustomProduct[EditionNumber] |
| Sub Title | Sub Title | CustomProduct [SubTitle] |
| Product Type | Product Type [Journal] | CustomProduct[ProductType] |
| Author Name | | CustomProduct[AllAuthorDesc] |
| Publication Date | | ElsProductRegionalInfo[PubDate] |
| Imprint | | CustomProduct[IMPRINT] |
| Next Edition | | CustomProduct[NEXTEDITIONISBN] |
| Print Book ISBN | ISSN | CustomProduct[Isbn] |

| Book (Web) | Journal (Web) | CALoader.XML |
|---|---|---|
| eBook ISBN | ISSN | CustomProduct[Isbn] |
| Short Description | Short Description | ElsProductRegionalInfo[shortdescription] |
| Series | | NA |
| Number of Page | | CustomProduct[NUMBEROFPAGES] |
| Dimensions | | CustomProduct[pubnumlog] |
| | Impact Factor | customproduct[impact_factor] |
| | Impact Factor Year | NA |
| | Editor Information | CustomProduct[Author_AList] |
| | Journal Link | autogenerated |
| | Volume Number | CustomProduct[VOLUME_NUMBER] |
| | Issue Number | CustomProduct[Frequency] |
| | 5Yr Impact Factor | CustomProduct[VERSION_NO] |
| OVERVIEW | OVERVIEW | |
| Key Feature | | els_product_regional_info[KEYFEATURE] |
| Long Description | | els_product_regional_info [Prd_Long_Description] |
| Readership | Audience | els_product_regional_info[audience] |
| Breadcrumb | | product[cat_id] |
| | Aim and Scope | els_product_regional_info [Prd_Long_Description] |
| PRICE | PRICE | |
| Format | Format | CustomSku[SKU_TYPE], Custom_product[web_product_Type_name] CustomSku(Purchase_Type) |
| Price | Price | Pricing_Detail[list_price] |
| Stock | Stock | inv_detail[availability_status] customsku[out_of_print] |
| AUTHORS | | |
| Name | | CustomProduct[AUTHOR_ALIST] |
| Biography | | NA |
| Affiliations and Expertise | | NA |
| Recent Publications (4) | | Query Result |
| TOC | | |
| Title | | Product[Prd_Display_Name] |
| TOC | | CustomProduct[Toc] |
| EDITORIAL REVIEWS | | |
| Review | | els_product_regional_info[quotes] |
| | BIBLIOGRAPHIC INFORMATION | |
| | Jurnal Issue and Dispatch Date | |

| Book (Web) | Journal (Web) | CALoader.XML |
|---|---|---|
| | Abstracting and Indexing | CustomProduct[ABSTRACT] |
| | ISSUES | |
| | Free Sample Issue | NA |
| | Special Issue | Scrapping from Elsevier.com |
| | ARTICLES | |
| | Article | Scrapping from Elsevier.com |
| | EDITORIAL INFORMATION | |
| | Editorial Information | CustomProduct[Author_BList] |
| VIDEO | | |
| Media File | NA | NA |
| RESOURCE | | |
| Online Companion Materials link | NA | NA |
| Instructor Ancillary Support Materials link | NA | NA |

## 2.2 Tables of Book and Journal in ATG Database

Based on the class diagram of entities in CALoader (Figure 3) we could develop the class diagram of entities (Figure 4) and the related tables (Table 4) in the ATG database. The mapping for the class diagram in the CALoader and the database is as follows:

**Table 3 CALoader Entities-Database Mapping**

| CALoader.xml Entities | ATG Tables |
|---|---|
| CustomProduct and Product | DCSX_PRODUCT, ANG_PRODUCT, ELS_PRODUCT, ELS_PRODUCT_EN, DCS_PRD_PRNT_CAT, and DCS_CATEGORY |
| ProductRegionalInfo | ELS_PRODUCT_REGIONAL_INFO |
| CustomSKU and SKU | DCSX_SKU, ELS_SKU, and ELS_DCSX_SKU |
| PricingDetail | DCS_PRICE, and DCS_PRICE_LIST |
| InventoryDetail | NA |

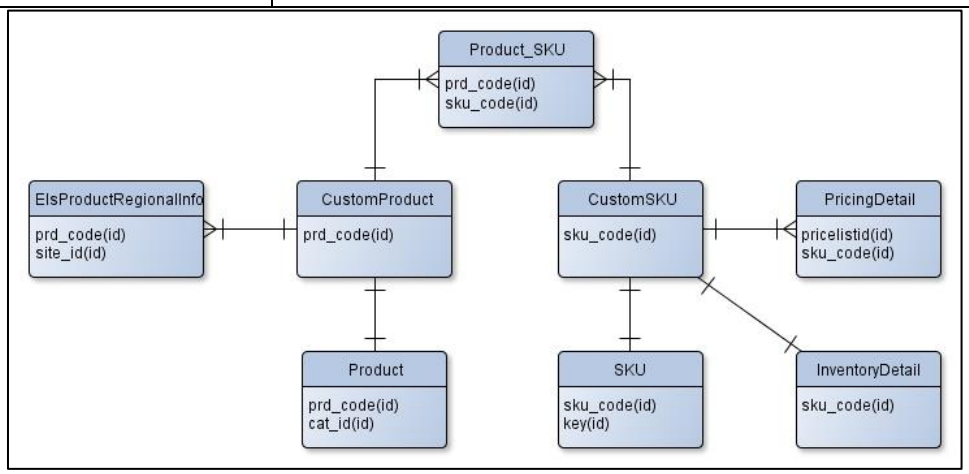


Figure 3 Product Class Diagram - CALoader

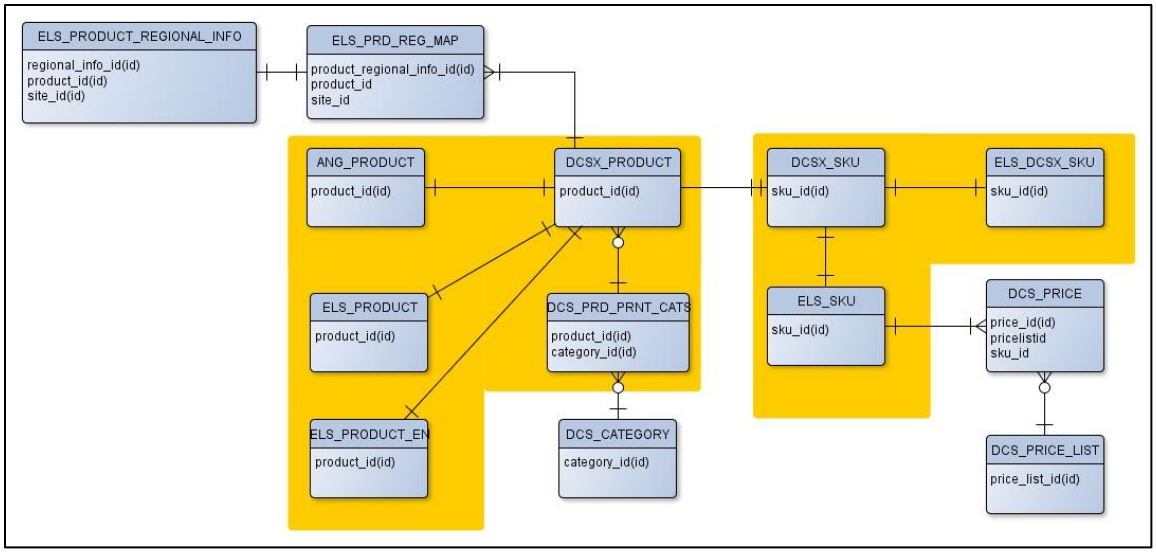


Figure 4 Product Class Diagram - ATG Database

**Table 4 Tables and Attributes of Product in Database**

| VIEW/ TABLE | COLUMN_NAME | DATA TYPE | NULL? | View | Book | Journal |
|---|---|---|---|---|---|---|
| ANG_PRODUCT | PRODUCT_ID | VARCHAR2 (40 Byte) | N | | | x |
| | IMPACT_FACTOR | NUMBER (10,3) | Y | x | | x |
| DCSX_PRODUCT | PRODUCT_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | TITLE | VARCHAR2 (256 Byte) | Y | x | x | x |
| | PRODUCT_TYPE | VARCHAR2 (40 Byte) | Y | x | x | x |
| | ISBN | VARCHAR2 (20 Byte) | Y | x | x | x |
| | ALL_AUTHOR | VARCHAR2 (4000 Byte) | Y | x | x | |
| | NUMBER_OF_PAGES | VARCHAR2 (50 Byte) | Y | x | x | |
| | PUB_NUM_LOG | VARCHAR2 (16 Byte) | Y | x | x | |
| | VERSION_NUMBER | VARCHAR2 (40 Byte) | Y | x | | x |
| | TABLE_OF_CONTENTS | CLOB | Y | x | x | |
| ELS_PRODUCT | PRODUCT_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | IMPRINT | VARCHAR2 (50 Byte) | Y | x | x | |
| | NEXT_EDITION_ISBN | VARCHAR2 (20 Byte) | Y | x | x | |
| | AUTHOR_ALIST | CLOB | Y | x | x | x |
| | AUTHOR_BLIST | CLOB | Y | x | | x |
| | VOLUME_NUMBER | VARCHAR2 (20 Byte) | Y | x | | x |
| | EDITION_NUMBER | VARCHAR2 (80 Byte) | Y | x | | |
| | FREQUENCY | VARCHAR2 (100 Byte) | Y | x | | x |
| ELS_PRODUCT_EN | PRODUCT_ID | VARCHAR2 (254 Byte) | N | | x | x |
| | SUB_TITLE | VARCHAR2 (255 Byte) | Y | x | x | x |
| | WEB_PRODUCT_TYPE_NAME | VARCHAR2 (100 Byte) | Y | x | x | x |
| | ABSTRACT | CLOB | Y | x | x | |
| DCSX_SKU | SKU_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | SKU_TYPE | VARCHAR2 (20 Byte) | Y | x | x | x |

| VIEW/ TABLE | COLUMN_NAME | DATA TYPE | NULL? | View | Book | Journal |
|---|---|---|---|---|---|---|
| ELS_DCSX_SKU | SKU_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | FULFILLMENT_COMPANY_CODE | VARCHAR2 (40 Byte) | Y | | x | x |
| ELS_SKU | SKU_ID | VARCHAR2 (254 Byte) | N | | x | x |
| | OUT_OF_PRINT | NUMBER (1) | Y | | x | x |
| | PURCHASE_TYPE | VARCHAR2 (75 Byte) | Y | x | x | x |
| DCS_PRICE | PRICE_LIST_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | PRICE_LIST | VARCHAR2 (40 Byte) | N | | x | x |
| | LIST_PRICE | NUMBER (19,7) | Y | x | x | x |
| | SKU_ID | VARCHAR2 (40 Byte) | Y | | x | x |
| ELS_PRODUCT_ REGIONAL_INFO | REGIONAL_INFO_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | PUB_DATE | DATE | Y | x | x | |
| | PUB_STATUS | INTEGER | Y | | x | x |
| | SHORT_DESCRIPTION | CLOB | Y | x | x | x |
| | KEY_FEATURE | VARCHAR2 (4000 Byte) | Y | x | x | |
| | LONG_DESCRIPTION | CLOB | Y | x | x | |
| | AUDIENCE | CLOB | Y | x | x | x |
| | QUOTES | CLOB | Y | x | x | |
| | FULFILLMENT_COMPANY_CODE | VARCHAR2 (40 Byte) | Y | | x | x |
| | PRODUCT_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | SITE_ID | VARCHAR2 (40 Byte) | N | | x | x |
| DCS_PRD_PRNT_CATS | PRODUCT_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | CATEGORY_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | DISPLAY_NAME | VARCHAR2 (254 Byte) | Y | | x | x |
| ELS_PRD_REG_MAP | PRODUCT_REGIONAL_INFO_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | PRODUCT_ID | VARCHAR2 (40 Byte) | N | | x | x |
| | SITE_ID | VARCHAR2 (40 Byte) | N | | x | x |

The criteria used to filter the valid Journal and Book in the database (0) are :

a. Product
   - The product_id begins with 'EST_' and the main table is DCSX_PRODUCT
   - The product_type is in printbook, ebook, or journal
   - It has at least 1 active site_id in product_regional_info

b. SKU
   - The sku_id begins with 'EST_' and the main table is DCSX_SKU
   - It should be related to a product
   - It has at least 1 price

c. Product-SKU relation
   - A product_id is related with SKUs that have the same string if "SKU" is replaced with "PK" and the sku_type is removed, e.g. a product with id EST_GLB_BS-PK-9780080582535 is related with the sku with these ids: EST_GLB_BS-SKU-9780080582535_GL, EST_GLB_BS-SKU-9780080582535_VST.

# 3 Data Quality Assessment

The data quality assessment is conducted by implementing the data quality metrics using the SQL query and procedures in the database. The list of procedures, views, and temporary table that are implemented for the assessment is in Appendix 2.

The data quality metrics to assess the data quality related to business problems and preventive/ reactive measures in Appendix 4 are implemented as described in following sections.

## 3.1 Business Problems - Poor Data

### 3.1.1 Completeness

- Related Business Problems : i, iii, and iv
- Related Preventive and Corrective measures : 1-3 (CPR-01, CPR-02, CPR-03)
- Method : Total = (70% x CPR-01) + (15% x CPR-02) + (15% x CPR-03)

1) Total
   a) Journal = (70% x CPR-01) + (15% x CPR-02) + (15% x CPR-03) = 0.8364
   b) Book = (70% x CPR-01) + (15% x CPR-02) + (15% x CPR-03) = 0.9265

2) CPR-01 : Ratio of Record with non-blank or non-null field in Product Repository
   a) Journal:
      - Records : 2,129
      - Method : (20/27 x Journal_Product) + (7/27 x Avg_Journal_SKU)
      - Assessment using PL/SQL Stored Procedure:
        - execute PROC_DQ_COMP_JOURNAL_INFO;
        - execute PROC_DQ_COMP_JOURNAL_REG;
        - execute PROC_DQ_COMP_JOURNAL_SKU;
        - execute PROC_DQ_COMP_JOURNAL_PRICE;
        - execute PROC_DQ_COMP_JOURNAL;
        - select avg(v_res), min(v_res), max(v_res) from dq_comp_journal;
      - Completeness : Avg = 0.8951 , Min = 0.6296 Max = 0.9630

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 0 |
| 50 | 70 | 118 |
| 70 | 80 | 114 |
| 80 | 90 | 621 |
| 90 | 100 | 1,276 |
| Total Records | | 2,129 |

      i) Journal Product
         - Journal product is composed of Journal Info and Journal Regional Info
         - Records : 2,129

- Method : (15/20 x Journal_Info) + (5/20 x Avg_Journal_Regional_Info)
- Assessment using PL/SQL Stored Procedure:
  - execute PROC_DQ_COMP_JOURNAL_INFO;

**Table 5 Journal Info CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| PRODUCT_ID | Not Blank and Not Null | 2,129 | - |
| TITLE | Not Blank and Not Null | 2,129 | - |
| PRODUCT_TYPE | Not Blank and Not Null | 2,129 | - |
| ISBN | Not Blank and Not Null | 2,129 | - |
| VERSION_NUMBER | Not Blank and Not Null | 2,129 | - |
| AUTHOR_ALIST | Not Blank and Not Null | 1,944 | 185 |
| AUTHOR_BLIST | Not Blank and Not Null | 1,903 | 226 |
| VOLUME_NUMBER | Not Blank and Not Null | 1,903 | 226 |
| EDITOR | Not Blank and Not Null | 1,931 | 198 |
| FREQUENCY | Not Blank and Not Null | 2,086 | 43 |
| ABSTRACT | Not Blank and Not Null | 1,690 | 439 |
| SUB_TITLE | Not Blank and Not Null | 1,028 | 1,101 |
| WEB_PRODUCT_TYPE_NAME | Not Blank and Not Null | 2,125 | 4 |
| CATEGORY_ID | Not Blank and Not Null | 2,129 | - |
| IMPACT_FACTOR | Not Blank and Not Null | 1,566 | 563 |

  - execute PROC_DQ_COMP_JOURNAL_REG;

**Table 6. Journal Regional Info CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| SITE_ID | Not Blank and Not Null | 14,870 | - |
| PUB_STATUS | Not Blank and Not Null | 14,870 | - |
| SHORT_DESCRIPTION | Not Blank and Not Null | - | 14,870 |
| AUDIENCE | Not Blank and Not Null | 8,523 | 6,347 |
| FULFILLMENT_COMPANY_CODE | Not Blank and Not Null | 14,870 | - |

  - execute PROC_DQ_COMP_JOURNAL;
  - select avg(v_res), min(v_res), max(v_res) from dq_comp_prd_journal;
- Completeness : Avg = 0.8585 , Min = 0.5, Max = 0.95

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 4 |
| 50 | 70 | 228 |
| 70 | 80 | 272 |
| 80 | 90 | 1,000 |
| 90 | 100 | 625 |
| Total Records | | 2,129 |

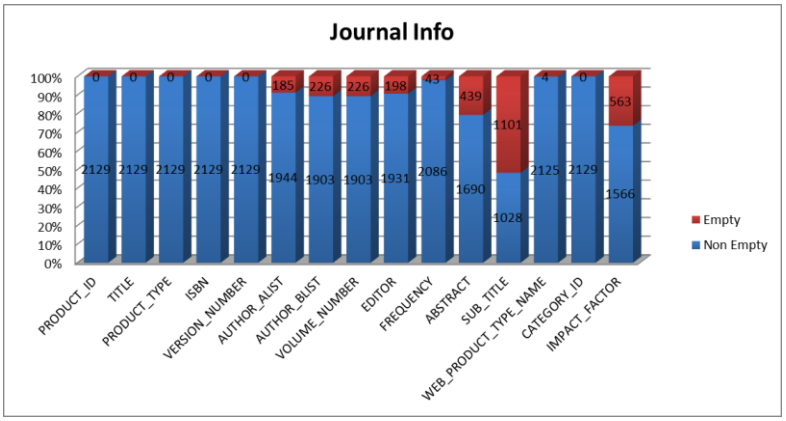


**Figure 5 Journal Info CPR-01 Assessment**

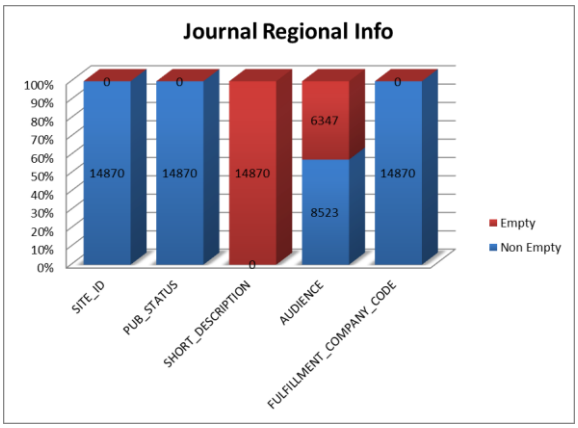


**Figure 6 Journal Reg Info CPR-01 Assessment**

ii) Journal SKU

- Journal SKU is composed of SKU and Price
- Records : 5,633
- Method : (5/7 x Journal_SKU) + (2/7 x Avg_Journal_Price)
- Assessment using PL/SQL Stored Procedure:
  - execute PROC_DQ_COMP_JOURNAL_SKU;

**Table 7 Journal SKU CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| SKU_ID | Not Blank and Not Null | 5,634 | 0 |
| SKU_TYPE | Not Blank and Not Null | 5,634 | 0 |
| FULFILLMENT_COMPANY_CODE | Not Blank and Not Null | 5,634 | 0 |
| OUT_OF_PRINT | Not Blank and Not Null | 5,634 | 0 |
| PURCHASE_TYPE | Not Blank and Not Null | 5,634 | 0 |

  - execute PROC_DQ_COMP_JOURNAL_PRICE;

**Table 8 Journal Price CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| SKU_ID | Not Blank and Not Null | 20,139 | 0 |
| PRICE_LIST | Not Blank and Not Null | 20,139 | 0 |
| LIST_PRICE | Not Blank and Not Null | 20,139 | 0 |

  - execute PROC_DQ_COMP_JOURNAL;
  - select avg(v_res), min(v_res), max(v_res) from dq_comp_sku_journal;

- Completeness : Avg = 0.9999 , Min = 0.8571, Max = 1.00

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 0 |
| 50 | 70 | 0 |
| 70 | 80 | 0 |
| 80 | 90 | 1 |

| | | |
|---|---|---|
| 90 | 100 | 5,632 |
| Total Records | | 5,633 |

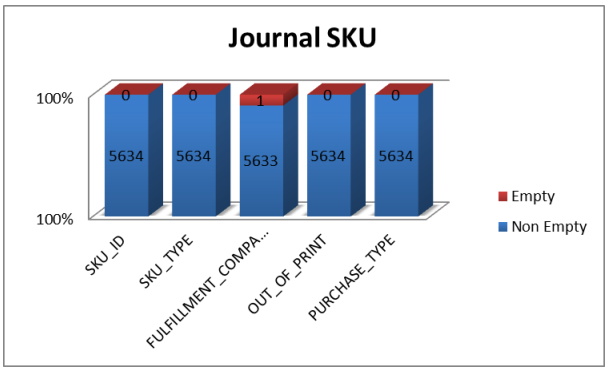


**Figure 7 Journal SKU CPR-01 Assessment**


**Figure 8 Journal Price CPR-01 Assessment**

b) Book

- Records : 43,463
- Method : (23/30 x Book_Product) + (7/30 x Avg_Book_SKU)
- Assessment using PL/SQL Stored Procedure:
  - execute PROC_DQ_COMP_BOOK_INFO;
  - execute PROC_DQ_COMP_BOOK_REG;
  - execute PROC_DQ_COMP_BOOK_SKU;
  - execute PROC_DQ_COMP_BOOK_PRICE;
  - execute PROC_DQ_COMP_BOOK;
  - select avg(v_res), min(v_res), max(v_res) from dq_comp_book;
- Completeness : Avg = 0.7920 , Min = 0.5667, Max = 0.9667

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 0 |
| 50 | 70 | 9,941 |
| 70 | 80 | 15,116 |
| 80 | 90 | 16,820 |
| 90 | 100 | 1,586 |
| Total Records | | 43,463 |

i) Book Product

- Book Product is composed of book info and regional info
- Records : 43,464
- Method : (14/23 x Book_Info) + (9/23 x Avg_Book_Regional_Info)
- Assessment using PL/SQL Stored Procedure:
  - execute PROC_DQ_COMP_BOOK_INFO;

**Table 9. Book Info CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|

| | | | |
|---|---|---|---|
| PRODUCT_ID | Not Blank and Not Null | 43,464 | - |
| TITLE | Not Blank and Not Null | 43,464 | - |
| PRODUCT_TYPE | Not Blank and Not Null | 43,464 | - |
| ISBN | Not Blank and Not Null | 43,464 | - |
| ALL_AUTHOR | Not Blank and Not Null | 43,108 | 356 |
| NUMBER_OF_PAGES | Not Blank and Not Null | 40,016 | 3,448 |
| PUB_NUM_LOG | Not Blank and Not Null | 13,327 | 30,137 |
| TABLE_OF_CONTENTS | Not Blank and Not Null | 29,718 | 13,746 |
| IMPRINT | Not Blank and Not Null | 43,464 | - |
| NEXT_EDITION_ISBN | Not Blank and Not Null | 2,263 | 41,201 |
| AUTHOR_ALIST | Not Blank and Not Null | 43,413 | 51 |
| SUB_TITLE | Not Blank and Not Null | 14,061 | 29,403 |
| WEB_PRODUCT_TYPE_NAME | Not Blank and Not Null | 41,518 | 1,946 |
| CATEGORY_ID | Not Blank and Not Null | 43,464 | - |

- execute PROC_DQ_COMP_BOOK_REG;

**Table 10Book Regional Info CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| SITE_ID | Not Blank and Not Null | 292,647 | - |
| PUB_DATE | Not Blank and Not Null | 201,488 | 91,159 |
| PUB_STATUS | Not Blank and Not Null | 292,647 | - |
| SHORT_DESCRIPTION | Not Blank and Not Null | 149,420 | 143,227 |
| KEY_FEATURE | Not Blank and Not Null | 158,548 | 134,099 |
| LONG_DESCRIPTION | Not Blank and Not Null | 169,122 | 123,525 |
| AUDIENCE | Not Blank and Not Null | 129,000 | 163,647 |
| QUOTES | Not Blank and Not Null | 65,021 | 227,626 |
| FULFILLMENT_COMPANY_CODE | Not Blank and Not Null | 292,647 | - |

- execute PROC_DQ_COMP_BOOK;
- select avg(v_res), min(v_res), max(v_res) from dq_comp_prd_book;

- Completeness : Avg = 0.7488, Min = 0.4782, Max = 1.00

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 12 |
| 50 | 70 | 16,533 |
| 70 | 80 | 10,843 |
| 80 | 90 | 11,943 |
| 90 | 100 | 4,133 |
| Total Records | | 43,464 |

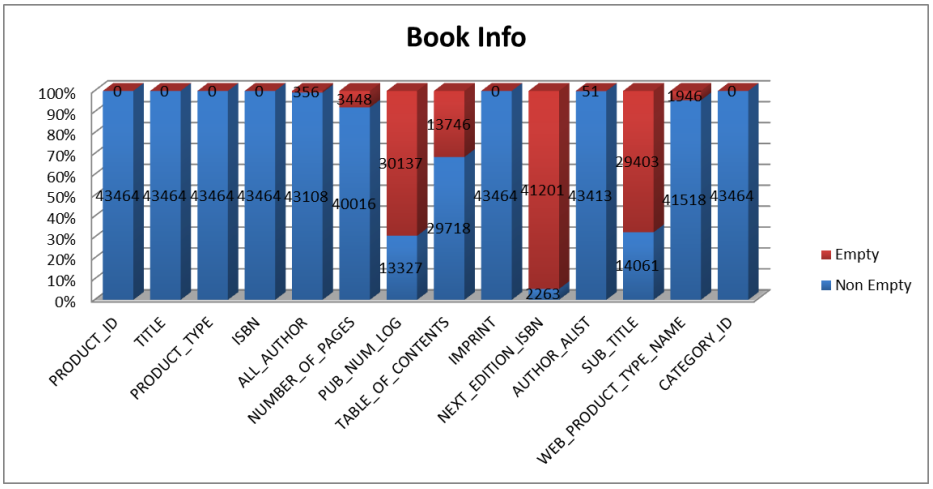


**Figure 9 Book Info CPR-01 Assessment**

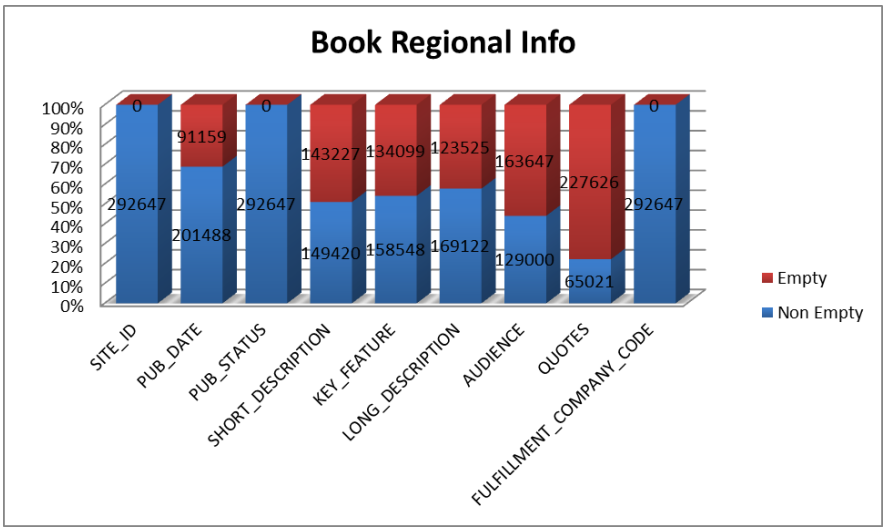


**Figure 10 Book Regional Info CPR-01 Assessment**

ii) Book SKU

- Book SKU is composed of SKU and Price
- Records : 103,414
- Method : (5/7 x Book_SKU) + (2/7 x Avg_Book_Price)
- Assessment using PL/SQL Stored Procedure:
  - execute PROC_DQ_COMP_BOOK_SKU;

**Table 11 Book SKU CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| SKU_ID | Not Blank and Not Null | 103,449 | - |
| SKU_TYPE | Not Blank and Not Null | 103,400 | 49 |
| FULFILLMENT_COMPANY_CODE | Not Blank and Not Null | 77,813 | 25,636 |
| OUT_OF_PRINT | Not Blank and Not Null | 102,479 | 970 |
| PURCHASE_TYPE | Not Blank and Not Null | 102,192 | 1,257 |

- execute PROC_DQ_COMP_BOOK_PRICE;

**Table 12 Book Price CPR-01 Assessment**

| Column Name | Criteria | OK | NOT OK |
|---|---|---|---|
| PRICE_LIST | Not Blank and Not Null | 321,417.00 | - |
| LIST_PRICE | Not Blank and Not Null | 321,319.00 | 98.00 |

- execute PROC_DQ_COMP_BOOK;
- select avg(v_res), min(v_res), max(v_res) from dq_comp_sku_book;

▪ Completeness : Avg = 0.9614, Min = 0.4286, Max = 1.00

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 49 |
| 50 | 70 | 287 |
| 70 | 80 | 715 |
| 80 | 90 | 25,427 |
| 90 | 100 | 76,936 |
| Total Records | | 103,414 |

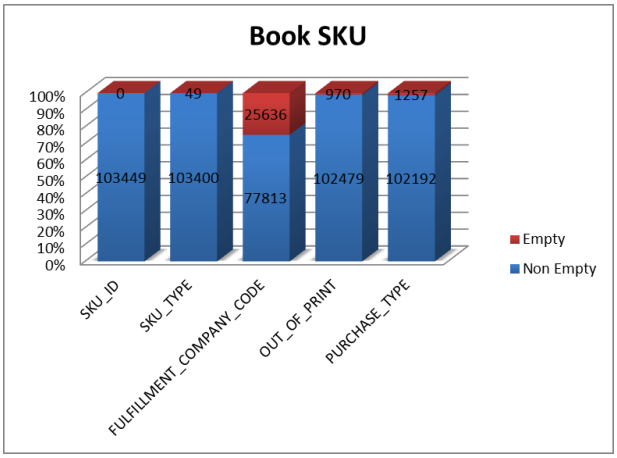


**Figure 11 Book SKU CPR-01 Assessment**

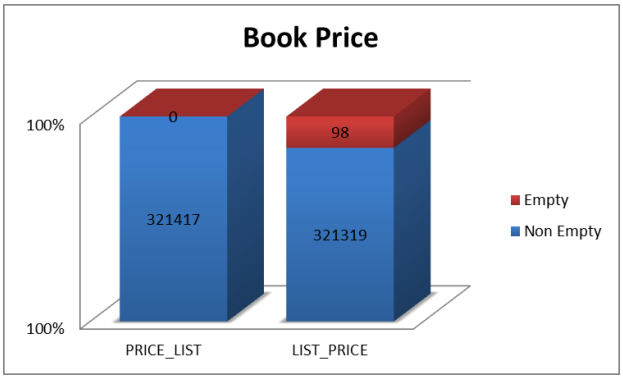


**Figure 12 Book Price CPR-01 Assessment**

3) CPR-02 and CPR-03

a) Table Size Summary

i) PRODUCT

| | Table | Book | Journal |
|---|---|---|---|
| a | DCSX_PRODUCT | 52,777 | 2,273 |
| b | ANG_PRODUCT | NA | 2,273 |
| c | ELS_PRODUCT | 52,777 | 2,273 |
| d | ELS_PRODUCT_EN | 50,169 | 2,268 |
| e | DCS_PRD_PRNT_CATS | 42,401 | 2,251 |
| f | ELS_PRODUCT_REGIONAL_INFO | 43,464 | 2,129 |
| g | V_[PRODUCT]_INFO (a-e) | 43,464 | 2,129 |

| | | | |
|---|---|---|---|
| h | PRODUCT-REGIONAL_INFO | 43,464 | 2,129 |

ii) SKU

| | Table | Book | Journal |
|---|---|---|---|
| b | DCSX_SKU | 103,449 | 5,634 |
| c | ELS_DCSX_SKU | 103,400 | 5,634 |
| d | ELS_SKU | 102,479 | 5,634 |
| e | DCS_PRICE | 103,414 | 5,634 |
| f | V_[PRODUCT]_SKU (a-d) | 103,449 | 5,634 |
| g | SKU-PRICE | 103,414 | 5,633 |

iii) PRODUCT_SKU

| | Table | Product | SKU |
|---|---|---|---|
| a | Product(h) - SKU(g) [BOOK] | 43,463 | 91,005 |
| b | Product(h) - SKU(g) [JOURNAL] | 2,129 | 5,633 |

b) CPR-02 : Ratio of NON-Parentless Record in Product Repository (e.g. SKU is referenced by a Product)
   i) Book = iii.a-SKU / ii.g = 91,005/ 103,414 = 0.8800
   ii) Journal = iii.a-SKU / ii.g = 5,633 / 5,634 = 0.9998

c) CPR-03 : CPR-03 : Ratio of NON-Childless Record in Product Repository (e.g. Product has SKU)
   i) Book = iii.a-PRODUCT / i.h = 43,463/ 43,464 = 0.9999
   ii) Journal = iii.a-PRODUCT / i.h = 2,129/ 2,129 = 1.00

### 3.1.2 Consistency (Title-Category)

- Related Business Problems : ii, vi
- Related Preventive and Corrective measures : 6 (SC-02) ,8 (AOC-01), 9 (AOC-02)
- Method : Total = (15% x SC-02) + (15% x AOC-01) + (70% x AOC-02)

1) Total
   a) Journal = 1.00
   b) Book = 0.7908

2) AOC-01 : Ratio of reasonable fields (subject related)in Product Repository
   a) Journal
      - Records : 2,129
      - Field that has the same top-5 values based on its distribution compared with previous data
      - Assessment using SQL execution :
        - select * from DQ_LOV_JOURNAL_CATEGORY_ID where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| EST_SA-9-91-206 | 50 | 2,129 | 2.3485 |

| | | | |
|---|---|---|---|
| EST_SA-6-42 | 47 | 2,129 | 2.2076 |
| EST_SA-8-82-141 | 44 | 2,129 | 2.0667 |
| EST_SA-9-91-205 | 39 | 2,129 | 1.8318 |
| EST_SA-8-73 | 38 | 2,129 | 1.7849 |

- Consistency = 1.00 (1st iteration)

b) Book

- Records : 43,464
- Field that has the same top-5 values based on its distribution compared with previous data
- Assessment using SQL execution :
  - select * from DQ_LOV_BOOK_CATEGORY_ID where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| EST_SA-8-85 | 1,587 | 43,464 | 3.6513 |
| EST_IMP-23 | 1,098 | 43,464 | 2.5262 |
| EST_SA-7-54 | 949 | 43,464 | 2.1834 |
| EST_SA-8-75 | 935 | 43,464 | 2.1512 |
| EST_SA-8-83 | 846 | 43,464 | 1.9464 |

- Consistency = 1.00 (1st iteration)

3) AOC-02 : Ratio of record which adhere business rule in Product Repository

a) Journal

- Records : 2,129
- Rule : Products with the same title should have the same category
- Assessment using SQL execution :
  - select count(*) from dq_cons_journal_title1_cat;
    return no rows
- Consistency = (2,129-0)/ 2,129 = 1.00

b) Book

- Records : 43,464
- Field that has the same top-5 values based on its distribution compared with previous data
- Assessment using SQL execution :
  - select count(*) from dq_cons_book_title1_cat;
    return 12,992 rows
- Consistency = (43,464-12,992)/ 43,464 = 0.7011

4) SC-02 : Ratio of record which has non deviated value in Product Repository

a) Journal

- Records : 2,129
- Rule : Products with category in LoV (Appendix 3)
- Assessment using SQL execution :

- select count(*) from v2_journal_info where category_id not in (select category_id from dq_lov_journal_category_id);
  return no rows

- Consistency= (2,129-0)/ 2,129 = 1.00 (1st iteration)

b) Book

- Records : 43,464
- Rule : Products with category in LoV (Appendix 3)
- Assessment using SQL execution :
  - select count(*) from v2_book_info where category_id not in (select category_id from dq_lov_book_category_id);
    return no rows
- Consistency = (43,464-0)/ 43,464 = 1.00 (1st iteration)

### 3.1.3 Accuracy (Location-Price)

- Related Business Problems : v(a)
- Related Preventive and Corrective measures : 1 (CPR-01), 6 (SC-02) , 9 (AOC-02)
- Method : Total = (15% x SC-02) + (15% x CPR-01) + (70% x AOC-02)

1) Total
   a) Journal = 1.00
   b) Book = 0.9989

2) AOC-02 : Ratio of record which adhere business rule
   a) Journal
      - Records : 2,129
      - Rule : The location and price should be within these combinations:

| Site | Currency | PriceListID |
|---|---|---|
| EST_UK_BS | GBP/ EUR | plist8380012/ plist2310003 |
| EST_AU_BS | USD | plist2310002 |
| EST_ASIA_BS | USD | plist2310002 |
| EST_US_BS | USD | plist2310002 |
| EST_JP_BS | JPY/ USD | plist2310004/ plist2310002 |
| EST_MEA_BS | USD | plist2310002 |
| EST_EU_BS | EUR | plist2310003 |

      - Assessment using SQL execution :
        - select count(*) from DQ_CONS_SITE_PRICE_JOURNAL where cnt_site!=cnt_price;
          return 0 rows
      - Consistency = 1.00

   b) Book

- Records : 43,464
- Rule : The location and price should be within these combinations:

| Site | Currency | PriceListID |
|---|---|---|
| EST_UK_BS | GBP/ EUR | plist8380012/ plist2310003 |
| EST_AU_BS | AUD/ USD | plist2310005/ plist2310002 |
| EST_ASIA_BS | USD | plist2310002 |
| EST_US_BS | USD | plist2310002 |
| EST_JP_BS | JPY/ USD | plist2310004/ plist2310002 |
| EST_MEA_BS | USD | plist2310002 |
| EST_EU_BS | EUR | plist2310003 |

- Assessment using SQL execution :
  - select count(*) from DQ_CONS_SITE_PRICE_BOOK where cnt_site!=cnt_price;
    return 64 rows
- Consistency = (43,464-64)/ 43,464= 0.9985

3) SC-02 : Ratio of record which has non deviated value in Product Repository
   a) Journal
      - Records : 14,870 (site) and 20,139 (price)
      - Rule : Products with site and price in LoV (Appendix 3)
      - Assessment using SQL execution :
        - Site: select count(*) from t_journal_els_prd_reg_info where site_id not in (select site_id from dq_lov_journal_site_id);
          return no rows
        - Price : select count(*) from v2_journal_dcs_price where price_list not in (select price_list from dq_lov_journal_price_list);
          return no rows
      - Correctness = avg((14,870-0)/ 14,870; (20,139-0)/ 20,139) = 1.00 (1st iteration)

   b) Book
      - Records 292,647 (site) and 321,417 (price)
      - Rule : Products with site and price in LoV (Appendix 3)
      - Assessment using SQL execution :
        - Site : select count(*) from t_book_els_prd_reg_info where site_id not in (select site_id from dq_lov_book_site_id);
          return no rows
        - Price : select count(*) from v2_book_dcs_price where price_list not in (select price_list from dq_lov_book_price_list);
          return no rows
      - Correctness = avg((292,647 -0)/ 292,647; (321,417-0)/ 321,417) = 1.00 (1st iteration)

4) CPR-01 : Ratio of Record with non-blank or non-null price_list and site_id field in Product Repository
   a) Journal

- Records : 14,870 (site) and 20,139 (price)
- Assessment using SQL execution :
- Site : select count (product_id) from dq_comp_journal_reg where v_site_id =1; return 14,870 rows
- Price : select count(sku_id) from dq_comp_journal_price where v_price_list =1; return 20,139 rows
- Completeness = avg(14,870/ 14,870; 20,139/ 20,139) = 1.00

b) Book
- Records 292,647 (site) and 321,417 (price)
- Rule : Products with site and price in LoV (Appendix 3)
- Assessment using SQL execution :
  - Site: select count (product_id) from dq_comp_book_reg where v_site_id =1; return 292,647 rows
  - Price : select count(sku_id) from dq_comp_book_price where v_price_list =1; return 321,417 rows
- Completeness = avg(292,647 / 292,647; 321,417/ 321,417) = 1.00

### 3.1.4 Accuracy (Location/ Format (Type)-Fulfillment System)

- Related Business Problems : v(a)
- Related Preventive and Corrective measures : 1 (CPR-01), 6 (SC-02) , 9 (AOC-02)
- Method : Total = (15% x SC-02) + (15% x CPR-01) + (70% x AOC-02)

1) Total
   a) Journal = 0.9261
   b) Book = 0.9552

2) AOC-02 : Ratio of record which adhere business rule
   a) Journal
      - Records : 5,419
      - Rule : The location, format, and fulfillment company code should be within these combinations:

| Site | Print Journal | | eJournal |
|---|---|---|---|
| | PJROMIS | PJARGI | EJSD |
| EST_AU_BS | DELTA | ARGI | CRM |
| EST_EU_BS | DELTA | - | CRM |
| EST_MEA_BS | DELTA | ARGI | CRM |
| EST_UK_BS | DELTA | - | CRM |
| EST_JP_BS | DELTA | - | CRM |
| EST_US_BS | DELTA | ARGI | CRM |
| EST_ASIA_BS | DELTA | ARGI | CRM |

- Assessment using SQL execution :
  - select count(*) from DQ_CONS_SITE_FULFILL1_JOURNAL where cnt_site!=cnt_fulfill; return 572 rows
- AOC-02 = (5,419-572)/ 5,419 = 0.8944

b) Book
- Records : 91,008
- Rule : The location, format, and fulfillment company code should be within these combinations:

| Site | Print Book | eBook | |
|---|---|---|---|
| | Physical | EBSD | Others |
| EST_AU_BS | BOOKMASTER | CRM | DELTA |
| EST_EU_BS | DELTA | CRM | DELTA |
| EST_MEA_BS | DELTA | CRM | DELTA |
| EST_UK_BS | DELTA | CRM | DELTA |
| EST_JP_BS | COPS | CRM | DELTA |
| EST_US_BS | COPS | CRM | DELTA |
| EST_ASIA_BS | COPS | CRM | DELTA |

- Assessment using SQL execution :
  - select count(*) from DQ_CONS_SITE_FULFILL1_BOOK where cnt_site!=cnt_fulfill; return 5,822 rows
- AOC-02 = (91,008-5,822)/ 91,008 = 0.9360

3) SC-02 : Ratio of record which has non deviated value in Product Repository

a) Journal
- Records : 2,129 (product) and 14,870 (site)
- Rule : Products with site and fulfillment code in LoV (Appendix 3)
- Assessment using SQL execution :
  - select count(*) from t_journal_els_prd_reg_info where site_id not in (select site_id from dq_lov_journal_site_id) or fulfillment_company_Code not in (select fulfillment_company_Code from dq_lov_journal_fulfill_cc); return no rows
  - select count(*) from v2_journal_info where product_type not in (select product_type from dq_lov_journal_product_type) return no rows
- Correctness = avg( (14,870-0)/ 14,870; (2,2129-0)/ 2,2129) = 1.00

b) Book
- Records : 43,464 (product) and 292,647 (site)

- Rule : Products with site and price in LoV (Appendix 3)
- Assessment using SQL execution :
  - select count(*) from t_book_els_prd_reg_info where site_id not in (select site_id from dq_lov_book_site_id) or fulfillment_company_Code not in (select fulfillment_company_Code from dq_lov_book_fulfill_cc);
    return 1 rows
  - select count(*) from v2_book_info where product_type not in (select product_type from dq_lov_book_product_type)
    return no rows
- Correctness = avg((292,647 -1)/ 292,647; (43,464 -0)/ 43,464 ) = 0.9999

4) CPR-01 : Ratio of Record with non-blank or non-null site_id and fulfillment_company_code field in Product Repository
   a) Journal
      - Records : 2,129 (product) and 14,870 (site)
      - Assessment using SQL execution :
        - select count (product_id) from dq_comp_journal_reg where v_site_id =1 and v_fulfillment_company_code=1;
          return 14,870 rows
        - select count(product_id) from dq_comp_journal_info where v_product_type =1;
          return 2,129 rows
      - Completeness = avg(14,870/ 14,870; 2,129/ 2,129 = 1.00

   b) Book
      - Records 43,464 (product) and 292,647 (site)
      - Rule : Products with site and price in LoV (Appendix 3)
      - Assessment using SQL execution :
        - select count (product_id) from dq_comp_journal_reg where v_site_id =1 and v_fulfillment_company_code=1;
          return 292,647 rows
        - select count(product_id) from dq_comp_journal_info where v_product_type =1;
          return 43,464 rows
      - Consistency = avg(292,647 / 292,647; 43,464/ 43,464) = 1.00

### 3.1.5 Accuracy (ISN Journal)

- Related Business Problems : v(c), v(d)
- Related Preventive and Corrective measures : 14 (ACR-01)

1) Total
   a) Journal = 0.8750
2) ACR-01 : Ratio of Record with non-blank or non-null field in Product Repository
   a) Journal:

- Records : 2,129
- Assessment using SQL :
  - V2_JOURNAL_INFO : 2,129 records
  - O_JOURNAL_ID_PRICE_2013 : 2,212 records
  - select b.product_id from (select product_id fromv2_journal_info) a, O_JOURNAL_ID_PRICE_2013 b where A.PRODUCT_ID = b.product_id; result = 2,030 records
- ACR-01 and ACR-03 = (2,030/2,129) x (2,030 x 2,212) = 0.8750

## 3.2 Preventive and Reactive Measures

### 3.2.1 Completeness per row

1) Total
   a) Journal = 0.8364 (See 3.1.1)
   b) Book = 0.9265 (See 3.1.1)

### 3.2.2 Syntactical Correctness

1) Total
   a) Journal = avg (SC-01, SC-02) = 0.9989
   b) Book = avg (SC-01, SC-02) = 0.9754
2) SC-01
   i. The data type in ATG is specified within the database whether it is a numeric, text, or date
   ii. All fields are nullable except for id field, e.g., product_id, site_id, sku_id, and price_list_id
   iii. The value is bounded to data type specification in Oracle and NULL for nullable fields,e.g. varchar2(4000) means any text with a length maximum of 4000 characters or NULL.
   iv. The lists of values in Appendix 3 are based on the data imported from the data sources (PPM, PROMIS, and CRM)
   v. The non-standard values are :
      - NULL/ EMPTY, which is assessed by CPR-01
      - Value not in List of Values, which is assessed by SC-02
      - **Thus, the SC-01 only assess NON LoV fields that are considered incorrect** :
        a. Text : "UNKNOWN", "EMPTY", "BLANK", "NULL"
        b. Numeric : <=0
        c. Date : 1 Jan 1900 or 1 Jan 1970

   a) Journal
      - Total = avg(SC-01 per field) = 0.9979
      - Method : use SQL

| Field | SQL | Incorrect | Total | Correct Ratio |
|---|---|---|---|---|
| Info | | | | |
| TITLE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where | - | 2,129 | 1.0000 |

| Field | SQL | Incorrect | Total | Correct Ratio |
|---|---|---|---|---|
| | lower(nvl(trim(TITLE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | | | |
| ISBN | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(ISBN),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| VERSION_ NUMBER | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where TO_NUMBER(VERSION_NUMBER)<=0) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| AUTHOR_ ALIST | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(AUTHOR_ALIST),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| AUTHOR_ BLIST | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(AUTHOR_BLIST),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| VOLUME_ NUMBER | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(VOLUME_NUMBER),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| EDITOR | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(EDITOR),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| FREQUENCY | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(FREQUENCY),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| ABSTRACT | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(ABSTRACT),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| SUB_TITLE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from | - | 2,129 | 1.0000 |

| Field | SQL | Incorrect | Total | Correct Ratio |
|---|---|---|---|---|
| | (select count(product_id) cnt from v2_journal_info where lower(nvl(trim(SUB_TITLE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_journal_info) b; | | | |
| IMPACT_ FACTOR | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_journal_info where IMPACT_FACTOR<=0) a, (select count(product_id) tot_row from v2_journal_info) b; | - | 2,129 | 1.0000 |
| Regional Info | | | | |
| SHORT_ DESCRIPTION | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_journal_els_prd_reg_info where lower(nvl(trim(SHORT_DESCRIPTION),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_journal_els_prd_reg_info) b; | - | 14,870 | 1.0000 |
| AUDIENCE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_journal_els_prd_reg_info where lower(nvl(trim(AUDIENCE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_journal_els_prd_reg_info) b; | - | 14,870 | 1.0000 |
| Price | | | | |
| LIST_PRICE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(sku_id) cnt from v2_journal_dcs_price where LIST_PRICE<=0) a, (select count(sku_id) tot_row from v2_journal_dcs_price) b; | 125 | 20,139 | 0.9938 |

b) Book

- Total = avg(SC-01 per field) = 0.9990
- Method : use SQL

| Field | SQL | Incorrect | Total | Correct Ratio |
|---|---|---|---|---|
| Info | | | | |
| TITLE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(TITLE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| ISBN | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(ISBN),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| ALL_AUTHOR | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where | - | 43,464 | 1.0000 |

| Field | SQL | Incorrect | Total | Correct Ratio |
|---|---|---|---|---|
| | lower(nvl(trim(ALL_AUTHOR),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | | | |
| NUMBER_OF_ PAGES | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(NUMBER_OF_PAGES),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| PUB_NUM_ LOG | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(PUB_NUM_LOG),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| TABLE_OF_ CONTENTS | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(TABLE_OF_CONTENTS),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| IMPRINT | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(IMPRINT),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| NEXT_ EDITION_ ISBN | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(NEXT_EDITION_ISBN),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| AUTHOR_ ALIST | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(AUTHOR_ALIST),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| SUB_TITLE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from v2_book_info where lower(nvl(trim(SUB_TITLE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from v2_book_info) b; | - | 43,464 | 1.0000 |
| Regional Info | | | | |
| PUB_DATE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where (extract(year from PUB_DATE)=1970 AND extract(month from PUB_DATE)=1 AND extract(day from PUB_DATE)=1) OR | 5,438 | 292,647 | 0.9814 |

| Field | SQL | Incorrect | Total | Correct Ratio |
|---|---|---|---|---|
| | (extract(year from PUB_DATE)=1900 AND extract(month from PUB_DATE)=1 AND extract(day from PUB_DATE)=1)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | | | |
| SHORT_ DESCRIPTION | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where lower(nvl(trim(SHORT_DESCRIPTION),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | - | 292,647 | 1.0000 |
| KEY_ FEATURE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where lower(nvl(trim(KEY_FEATURE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | - | 292,647 | 1.0000 |
| LONG_ DESCRIPTION | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where lower(nvl(trim(LONG_DESCRIPTION),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | - | 292,647 | 1.0000 |
| AUDIENCE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where lower(nvl(trim(AUDIENCE),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | - | 292,647 | 1.0000 |
| QUOTES | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where lower(nvl(trim(QUOTES),'-')) in (select values_name from t_blank_values)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | - | 292,647 | 1.0000 |
| Price | | | | |
| LIST_PRICE | select cnt, tot_row, (tot_row-cnt), (tot_row-cnt)/ tot_row from (select count(sku_id) cnt from v2_book_dcs_price where LIST_PRICE<=0) a, (select count(sku_id) tot_row from v2_book_dcs_price) b; | - | 321,417 | 1.0000 |

3) SC-02

a) Journal

- Total = avg(SC-02 per field) = 0.9998
- Method : use SQL

| Field | (SQL | Result | Total Record | Ratio |
|---|---|---|---|---|
| Info | | | | |
| PRODUCT_ TYPE | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_journal_info where PRODUCT_TYPE in (select | 2,129 | 2,129 | 1.0000 |

| Field | (SQL | Result | Total Record | Ratio |
|---|---|---|---|---|
| | PRODUCT_TYPE from DQ_LOV_JOURNAL_PRODUCT_TYPE)) a, (select count(product_id) tot_row from v2_journal_info) b; | | | |
| WEB_ PRODUCT_ TYPE_NAME | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_journal_info where WEB_PRODUCT_TYPE_NAME in (select WEB_PRODUCT_TYPE_NAME from DQ_LOV_JOURNAL_WEB_PT_NAME)) a, (select count(product_id) tot_row from v2_journal_info) b; | 2,125 | 2,129 | 0.9981 |
| CATEGORY_ID | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_journal_info where CATEGORY_ID in (select CATEGORY_ID from DQ_LOV_JOURNAL_CATEGORY_ID)) a, (select count(product_id) tot_row from v2_journal_info) b; | 2,129 | 2,129 | 1.0000 |
| Regional Info | | | | |
| PUB_STATUS | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_journal_els_prd_reg_info where PUB_STATUS in (select PUB_STATUS from DQ_LOV_JOURNAL_PUB_STATUS)) a, (select count(product_id) tot_row from t_journal_els_prd_reg_info) b; | 14,870 | 14,870 | 1.0000 |
| FULFILLMENT _COMPANY _CODE | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_journal_els_prd_reg_info where FULFILLMENT_COMPANY_CODE in (select FULFILLMENT_COMPANY_CODE from DQ_LOV_JOURNAL_FULFILL_CC)) a, (select count(product_id) tot_row from t_journal_els_prd_reg_info) b; | 14,870 | 14,870 | 1.0000 |
| SITE_ID | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_journal_els_prd_reg_info where SITE_ID in (select SITE_ID from DQ_LOV_JOURNAL_SITE_ID)) a, (select count(product_id) tot_row from t_journal_els_prd_reg_info) b; | 14,870 | 14,870 | 1.0000 |
| SKU | | | | |
| SKU_TYPE | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_journal_sku where SKU_TYPE in (select SKU_TYPE from DQ_LOV_JOURNAL_SKU_TYPE)) a, (select count(sku_id) tot_row from v2_journal_sku) b; | 5,634 | 5,634 | 1.0000 |
| FULFILLMENT _COMPANY _CODE | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_journal_sku where FULFILLMENT_COMPANY_CODE in (select FULFILLMENT_COMPANY_CODE from DQ_LOV_JOURNAL_FULFILL_CC1)) a, (select count(sku_id) tot_row from v2_journal_sku) b; | 5,633 | 5,634 | 0.9998 |
| OUT_OF_ PRINT | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_journal_sku where OUT_OF_PRINT in (select OUT_OF_PRINT from DQ_LOV_JOURNAL_OUT_OF_PRINT)) a, (select count(sku_id) tot_row from v2_journal_sku) b; | 5,634 | 5,634 | 1.0000 |
| PURCHASE_ TYPE | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_journal_sku where PURCHASE_TYPE in (select PURCHASE_TYPE from DQ_LOV_JOURNAL_PURCHASE_TYPE)) a, | 5,634 | 5,634 | 1.0000 |

| Field | (SQL | Result | Total Record | Ratio |
|---|---|---|---|---|
| | (select count(sku_id) tot_row from v2_journal_sku) b; | | | |
| Price | | | | |
| PRICE_LIST | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_journal_dcs_price where PRICE_LIST in (select PRICE_LIST from DQ_LOV_JOURNAL_PRICE_LIST)) a, (select count(sku_id) tot_row from v2_journal_dcs_price) b; | 20,139 | 20,139 | 1.0000 |

b) Book

- Total = avg(SC-02 per field) = 0.9518
- Method : use SQL

| Field | (SQL | Result | Total Record | Ratio |
|---|---|---|---|---|
| Info | | | | |
| PRODUCT_ TYPE | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_book_info where PRODUCT_TYPE in (select PRODUCT_TYPE from DQ_LOV_BOOK_PRODUCT_TYPE)) a, (select count(product_id) tot_row from v2_book_info) b; | 43,464 | 43,464 | 1.0000 |
| IMPRINT | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_book_info where IMPRINT in (select IMPRINT from DQ_LOV_BOOK_IMPRINT)) a, (select count(product_id) tot_row from v2_book_info) b; | 43,464 | 43,464 | 1.0000 |
| WEB_PRODUCT _TYPE_NAME | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_book_info where WEB_PRODUCT_TYPE_NAME in (select WEB_PRODUCT_TYPE_NAME from DQ_LOV_BOOK_WEB_PT_NAME)) a, (select count(product_id) tot_row from v2_book_info) b; | 41,518 | 43,464 | 0.9552 |
| CATEGORY_ID | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from v2_book_info where CATEGORY_ID in (select CATEGORY_ID from DQ_LOV_BOOK_CATEGORY_ID)) a, (select count(product_id) tot_row from v2_book_info) b; | 43,464 | 43,464 | 1.0000 |
| Regional Info | | | | |
| PUB_DATE | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where extract(year from PUB_DATE ) in (select PUB_YEAR from DQ_LOV_BOOK_PUB_DATE)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | 201,488 | 292,647 | 0.6885 |
| PUB_STATUS | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where PUB_STATUS in (select PUB_STATUS from DQ_LOV_BOOK_PUB_STATUS)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | 292,647 | 292,647 | 1.0000 |
| FULFILLMENT _COMPANY_ CODE | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where FULFILLMENT_COMPANY_CODE in (select FULFILLMENT_COMPANY_CODE from DQ_LOV_BOOK_FULFILL_CC)) | 292,647 | 292,647 | 1.0000 |

| Field | (SQL | Result | Total Record | Ratio |
|---|---|---|---|---|
| | a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | | | |
| SITE_ID | select cnt, tot_row, cnt/ tot_row from (select count(product_id) cnt from t_book_els_prd_reg_info where SITE_ID in (select SITE_ID from DQ_LOV_BOOK_SITE_ID)) a, (select count(product_id) tot_row from t_book_els_prd_reg_info) b; | 292,647 | 292,647 | 1.0000 |
| SKU | | | | |
| SKU_TYPE | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_book_sku where SKU_TYPE in (select SKU_TYPE from DQ_LOV_BOOK_SKU_TYPE)) a, (select count(sku_id) tot_row from v2_book_sku) b; | 103,400 | 103,449 | 0.9995 |
| FULFILLMENT _COMPANY_ CODE | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_book_sku where FULFILLMENT_COMPANY_CODE in (select FULFILLMENT_COMPANY_CODE from DQ_LOV_BOOK_FULFILL_CC1)) a, (select count(sku_id) tot_row from v2_book_sku) b; | 77,813 | 103,449 | 0.7522 |
| OUT_OF_ PRINT | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_book_sku where OUT_OF_PRINT in (select OUT_OF_PRINT from DQ_LOV_BOOK_OUT_OF_PRINT)) a, (select count(sku_id) tot_row from v2_book_sku) b; | 102,479 | 103,449 | 0.9906 |
| PURCHASE_ TYPE | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_book_sku where PURCHASE_TYPE in (select PURCHASE_TYPE from DQ_LOV_BOOK_PURCHASE_TYPE)) a, (select count(sku_id) tot_row from v2_book_sku) b; | 102,192 | 103,449 | 0.9878 |
| Price | | | | |
| PRICE_LIST | select cnt, tot_row, cnt/ tot_row from (select count(sku_id) cnt from v2_book_dcs_price where PRICE_LIST in (select PRICE_LIST from DQ_LOV_BOOK_PRICE_LIST)) a, (select count(sku_id) tot_row from v2_book_dcs_price) b; | 321,417 | 321,417 | 1.0000 |

### 3.2.3 Absence of Contradiction

There 3 rules to assess absence of contradiction/ consistency as follows: :

i. Title - Category (see 3.1.2 for rules)
ii. Location - Price (see 3.1.3 for tules)
iii. Location&Format - Fulfillment Company Code (see 3.1.4 for rules)

1) Total
   a) Journal = avg(AOC-01, AOC-02) = avg(1, 0.9648) = 0.9824
   b) Book = avg(AOC-01, AOC-02) = avg(0.9999, 0.9084) = 0.9542

2) AOC-01
   a) Journal = avg ($rule_i$) = 1.00

i) Rule (i) = 1 (see 3.1.2)

ii) Rule (ii) = 1.00

- Location
  - Assessment using SQL execution :
    - select * from DQ_LOV_JOURNAL_SITE_ID where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| EST_AU_BS | 2,129 | 14,870 | 14.3174 |
| EST_MEA_BS | 2,129 | 14,870 | 14.3174 |
| EST_ASIA_BS | 2,129 | 14,870 | 14.3174 |
| EST_US_BS | 2,129 | 14,870 | 14.3174 |
| EST_UK_BS | 2,118 | 14,870 | 14.2434 |

  - Consistency = 1.00 (1st iteration)
- Price
  - Assessment using SQL execution :
    - select * from DQ_LOV_JOURNAL_PRICE_LIST where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| plist2310002 | 5,633 | 20,139 | 27.9706 |
| plist2310003 | 5,046 | 20,139 | 25.0559 |
| plist2310004 | 5,026 | 20,139 | 24.9566 |
| plist8380012 | 3,864 | 20,139 | 19.1867 |
| plist9730004 | 570 | 20,139 | 2.8303 |

  - Consistency = 1.00 (1st iteration)

iii) Rule (iii) = 1.00

- Location
  - Assessment using SQL execution :
    - select * from DQ_LOV_JOURNAL_SITE_ID where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| EST_AU_BS | 2,129 | 14,870 | 14.3174 |
| EST_MEA_BS | 2,129 | 14,870 | 14.3174 |
| EST_ASIA_BS | 2,129 | 14,870 | 14.3174 |
| EST_US_BS | 2,129 | 14,870 | 14.3174 |
| EST_UK_BS | 2,118 | 14,870 | 14.2434 |

  - Consistency = 1.00 (1st iteration)
- Format
  - Assessment using SQL execution :
    - select * from DQ_LOV_JOURNAL_PRODUCT_TYPE where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| Journal | 2,129 | 2,129 | 100.0000 |

- Consistency = 1.00 (1st iteration)

- Fulfillment Company Code
  - Assessment using SQL execution :
    - select * from DQ_LOV_JOURNAL_FULFILL_CC where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| DELTA-E-COMMERCE | 8,064 | 14,870 | 54.2300 |
| COPS-E-COMMERCE | 4,677 | 14,870 | 31.4526 |
| BOOKMASTER-E-COMMERCE | 2,129 | 14,870 | 14.3174 |

  - Consistency = 1.00 (1st iteration)

b) Book = avg ($rule_i$) = 0.9999

i) Rule (i) = 1 (see 3.1.2)

ii) Rule (ii) = 1.00

- Location
  - Assessment using SQL execution :
    - select * from DQ_LOV_BOOK_SITE_ID where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| EST_JP_BS | 42,230 | 292,647 | 14.4304 |
| EST_US_BS | 42,026 | 292,647 | 14.3606 |
| EST_ASIA_BS | 42,026 | 292,647 | 14.3606 |
| EST_MEA_BS | 41,938 | 292,647 | 14.3306 |
| EST_UK_BS | 41,826 | 292,647 | 14.2923 |

  - Consistency = 1.00 (1st iteration)
- Price
  - Assessment using SQL execution :
    - select * from DQ_LOV_BOOK_PRICE_LIST where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| plist2310002 | 103,246 | 321,417 | 32.1221 |
| plist2310003 | 102,326 | 321,417 | 31.8359 |
| plist8380012 | 85,521 | 321,417 | 26.6075 |
| plist2310005 | 21,611 | 321,417 | 6.7237 |
| plist2310004 | 8,713 | 321,417 | 2.7108 |

  - Consistency = 1.00 (1st iteration)

iii) Rule (iii) = 1

- Location
  - Assessment using SQL execution :
    - select * from DQ_LOV_BOOK_SITE_ID where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| EST_JP_BS | 42,230 | 292,647 | 14.4304 |
| EST_US_BS | 42,026 | 292,647 | 14.3606 |
| EST_ASIA_BS | 42,026 | 292,647 | 14.3606 |
| EST_MEA_BS | 41,938 | 292,647 | 14.3306 |
| EST_UK_BS | 41,826 | 292,647 | 14.2923 |

  - Consistency = 1.00 (1st iteration)

- Format
  - Assessment using SQL execution :
    - select * from DQ_LOV_BOOK_PRODUCT_TYPE where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| Ebook | 24,128 | 43,464 | 55.5126 |
| Printbook | 19,336 | 43,464 | 44.4874 |

  - Consistency = 1.00 (1st iteration)

- Fulfillment Company Code
  - Assessment using SQL execution :
    - select * from DQ_LOV_BOOK_FULFILL_CC where rownum<=5;

| Values | Records | Total | % |
|---|---|---|---|
| DELTA-E-COMMERCE | 148,688 | 292,647 | 50.8080 |
| COPS-E-COMMERCE | 103,064 | 292,647 | 35.2179 |
| BOOKMASTER-E-COMMERCE | 40,894 | 292,647 | 13.9738 |
| COPS-ECOMMEauRCE | 1 | 292,647 | 0.0003 |

  - Consistency = 0.9999

3) AOC-02
   a) Journal = avg ($rule_i$) = 0.9648
      i) Rule (i) = 1 (see 3.1.2)
      ii) Rule (ii) = 1 (see 3.1.3)
      iii) Rule (iii) = 0.8944(see 3.1.4)

   b) Book = avg ($rule_i$) = 0.9084

i) Rule (i) = 0.7908 (see 3.1.2)
ii) Rule (ii) = 0.9985 (see 3.1.3)
iii) Rule (iii) = 0.9360 (see 3.1.4)

4) AOC-03
   a) Journal : Data is unavailable
   b) Book : Data is unavailable

### 3.2.4 Absence of Repetition

1) AOR-01
   a) Journal
      - Total = avg(AOR per each type of entity) = 1.00
      - Method : use SQL :

| Entity | SQL | Unique Row | Total Row | Ratio |
|---|---|---|---|---|
| Journal Info | select tot, tot_row, tot/tot_row from (select count(product_id) tot from (select product_id, count(product_id) cnt from v2_journal_info group by product_id) where cnt=1) a, (select count(product_id) tot_row from v2_journal_info) b; | 2,129 | 2,129 | 1.0000 |
| Journal Regional Info | select tot, tot_row, tot/tot_row from (select count(product_id) tot from (select product_id, count(product_id) cnt from t_journal_els_prd_reg_info group by product_id,site_id) where cnt=1) a, (select count(product_id) tot_row from t_journal_els_prd_reg_info) b; | 14,870 | 14,870 | 1.0000 |
| Journal SKU | select tot, tot_row, tot/tot_row from (select count(sku_id) tot from (select sku_id, count(sku_id) cnt from v2_journal_sku group by sku_id) where cnt=1) a, (select count(sku_id) tot_row from v2_journal_sku) b; | 5,634 | 5,634 | 1.0000 |
| Journal Price | select tot, tot_row, tot/tot_row from (select count(sku_id) tot from (select sku_id, count(sku_id) cnt from v2_journal_dcs_price group by sku_id, price_list) where cnt=1) a, (select count(sku_id) tot_row from v2_journal_dcs_price) b; | 20,139 | 20,139 | 1.0000 |

   b) Book
      - Total = avg(AOR per each type of entity) = 1.00
      - Method : use SQL :

| Entity | SQL | Unique Row | Total Row | Ratio |
|---|---|---|---|---|
| Book Info | select tot, tot_row, tot/tot_row from (select count(product_id) tot from (select product_id, count(product_id) cnt from v2_Book_info group by product_id) where cnt=1) a, (select count(product_id) tot_row from v2_Book_info) b; | 43,464 | 43,464 | 1.0000 |
| Book Regional Info | select tot, tot_row, tot/tot_row from (select count(product_id) tot from (select product_id, count(product_id) cnt from t_Book_els_prd_reg_info group by product_id,site_id) where cnt=1) a, (select count(product_id) tot_row from t_Book_els_prd_reg_info) b; | 292,647 | 292,647 | 1.0000 |

| Entity | SQL | Unique Row | Total Row | Ratio |
|---|---|---|---|---|
| Book SKU | select tot, tot_row, tot/tot_row from (select count(sku_id) tot from (select sku_id, count(sku_id) cnt from v2_Book_sku group by sku_id) where cnt=1) a, (select count(sku_id) tot_row from v2_Book_sku) b; | 103,449 | 103,449 | 1.0000 |
| Book Price | select tot, tot_row, tot/tot_row from (select count(sku_id) tot from (select sku_id, count(sku_id) cnt from v2_Book_dcs_price group by sku_id, price_list) where cnt=1) a, (select count(sku_id) tot_row from v2_Book_dcs_price) b; | 321,417 | 321,417 | 1.0000 |

### 3.2.5 Accuracy inc. Currency

1) Total
   a) Journal = average(CPR-01, SC-01, SC-02, ACR-01, ACR-02) = 0.9139
   b) Book = average(CPR-01, SC-01, SC-02, ACR-01, ACR-02)

2) CPR-01, SC-01, SC-02 see (3.2.1,3.2.2)

| | CPR-01 | SC-01 | SC-02 |
|---|---|---|---|
| **Journal** | 0.8364 | 0.9979 | 0.9998 |
| **Book** | 0.9265 | 0.9990 | 0.9518 |

3) ACR-01
   a) Journal = 0.8750 (see 3.1.5)
   b) Book

4) ACR-02
   a) Journal =
      - ETL date : November 20, 2013
      - Start-time = 04:00:05
      - End-time = 07:21:08
      - Duration = 3h 21m 3s = 201.05 m
      - Ratio = 1 - 201.05/1,440 = 0.8604
   b) Book = Journal = 0.8604

# 4 Performance Assessment with Google Analytics and Sales Data

The assessment of data quality with 2013 performance data (visits and sales) is conducted to find correlation between poor data and performance, and to determine the threshold values.

### 4.1 Completeness

#### 4.1.1 Google Analytics

1) Book
   - Year : 2013
   - Book : 31,949 (of 38,237) = 83.55% visited books assessed
   - Visit : 1,818,477 (of 1,966,341) = 92.48% visits assessed
   - Assessment using SQL to dq_ast_comp_book_ga.

| Min | Max | #Product | #Visit | #Trx | Bounces (%) |
|---|---|---|---|---|---|
| 0 | 0.5 | - | - | - | - |
| 0.5 | 0.7 | 4,789 | 44,332 | 274 | 79.5742 |
| 0.7 | 0.8 | 11,134 | 418,071 | 954 | 79.4955 (-0.08) |
| 0.8 | 0.9 | 14,511 | 1,117,871 | 5,328 | 76.7111 (-2.78) |
| 0.9 | 1 | 1,515 | 238,203 | 2,303 | 74.0095 (-2.7) |

   - Assessment Result:
     a. Most visits are for product with completeness between 0.8 and 0.9. This could be affected by a better search result when the product has that range of completeness
     b. The visits to the product pages with higher completeness have a lower bounce rate (± 2.7%). An increase of completeness level could provide a lower bounce rate that leads to a higher conversion rate.
     c. At least, the ATG team needs to increase the completeness level into minimum 0.8 to raise the visits and lower the bounce rate.

2) Journal
   - Year : 2013
   - Journal : 1,185 (of 1,123) = 83.55% visited journal assessed
   - Visit : 7,113 (of 7,175) = 92.48% visits assessed
   - Assessment using SQL to dq_ast_comp_journal_ga.

| Min | Max | #Product | #Visit | #Trx | Bounces (%) |
|---|---|---|---|---|---|
| 0 | 0.5 | - | | | |
| 0.5 | 0.7 | 4 | 9 | - | 100.0000 |
| 0.7 | 0.8 | 43 | 371 | 2 | 55.5156 |
| 0.8 | 0.9 | 302 | 1,104 | 12 | 67.1893 |
| 0.9 | 1 | 836 | 5,629 | 33 | 65.8322 |

   - Assessment Result:

a. Most visits are for product with completeness between 0.8 and 0.9. This could be affected by a better search result when the product has that range of completeness
b. The visits to the product pages with higher completeness have a lower bounce rate (+ 1.36%). An increase of completeness level could provide a lower bounce rate that leads to a higher conversion rate. The products with 0.7-0.8 have the lowest bounce rate but those also have a low visit number.
c. At least, the ATG team needs to increase the completeness level into minimum 0.8 to raise the visits and lower the bounce rate.

### 4.1.2 Sales

1) Book
   - Year : 2013
   - Book : 6,350(of 6,536) = 97.15% sold books assessed
   - Price : USD 7,857,208.91 (of USD 9,263,705.93) = 84.81% sales revenue assessed
   - Assessment using SQL to DQ_AST_COMP_BOOK_SALES

| Min | Max | #Product | Qty | Total(USD) |
|---|---|---|---|---|
| 0 | 0.5 | - | | |
| 0.5 | 0.7 | 382 | 4,540 | 277,588.21 |
| 0.7 | 0.8 | 1,430 | 4,589 | 543,451.76 |
| 0.8 | 0.9 | 3,777 | 18,404 | 5,765,484.64 |
| 0.9 | 1 | 761 | 5,494 | 1,270,684.31 |

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 0 |
| 50 | 70 | 9,941 |
| 70 | 80 | 15,116 |
| 80 | 90 | 16,820 |
| 90 | 100 | 1,586 |
| Total Records | | 43,463 |

   - Assessment Result:
     a. Most of the transactions are for products with completeness between 0.8 and 0.9. This information supports the previous assessment on Google Analytics result where this completeness range provides the highest visits.
     b. The products with 0.9-1 completeness have a better sales rate per item(7.2) compared with the products with 0.8-0.9 completeness (4.8).
     c. The ATG team needs to increase the completeness level into minimum 0.8 to increase the sales

2) Journal
   - Year : 2013

- Journal : 815 (of 822) = 99.15% sold books assessed
- Price : USD 13,898,474.0 (of USD 13,900,414.0) = 99.98% sales revenue assessed
- Assessment using SQL to DQ_AST_COMP_BOOK_SALES

| Min | Max | #Product | Qty | Total(USD) |
|---|---|---|---|---|
| 0 | 0.5 | - | | |
| 0.5 | 0.7 | 3 | 4 | 792,699 |
| 0.7 | 0.8 | 28 | 280 | 1,870,290 |
| 0.8 | 0.9 | 239 | 1,024 | 2,307,580 |
| 0.9 | 1 | 545 | 2,216 | 8,927,906 |

| Min(%) | Max(%) | Records |
|---|---|---|
| - | 50 | 0 |
| 50 | 70 | 118 |
| 70 | 80 | 114 |
| 80 | 90 | 621 |
| 90 | 100 | 1,276 |
| Total Records | | 2,129 |

- Assessment Result:
  a. Most of the transactions are for products with completeness between 0.9 and 0.1.
  b. The ATG team needs to increase the completeness level into minimum 0.9 to increase the sales

### 4.2 Consistency (Title - Category)

#### 4.2.1 Google Analytics

1) Book
   - Year : 2013
   - Book : 31,949 (of 38,237) = 83.55% visited books assessed
   - Visit : 1,818,477 (of 1,966,341) = 92.48% visits assessed
   - Assessment using SQL to DQ_AST_CORR_BOOK_GA

| Correct Category | Count of PRODUCT_ID | Average of BOUNCE | Sum of VISIT | Sum of TRX |
|---|---|---|---|---|
| FALSE | 11,880 | 77.0949 | 766,996 | 4,856 |
| TRUE | 20,069 | 78.5078 | 1,051,481 | 4,003 |
| Grand Total | 31,949 | 77.9824 | 1,818,477 | 8,859 |

   - Assessment Result:
     a. Products with correct Title - Category have more visits, however bounce rates is till high.

#### 4.2.2 Sales

1) Book

- Year : 2013
- Book : 6,350(of 6,536) = 97.15% sold books assessed
- Price : USD 7,857,208.91 (of USD 9,263,705.93) = 84.81% sales revenue assessed
- Assessment using SQL to DQ_AST_CORR_BOOK_SALES

| Correct Category | Count of PRODUCT_ID | Sum of TOTAL | Sum of QTY |
|---|---|---|---|
| FALSE | 3,190 | 5,085,613.48 | 18,924 |
| TRUE | 3,160 | 2,771,595.42 | 14,103 |
| Grand Total | 6,350 | 7,857,208.91 | 33,027 |

- Assessment Result:
  a. Correctness of the category does not provide a better sales performance.

### 4.3 Syntactical Correctness

#### 4.3.1 Google Analytics

1) Book (SC-02)
   - Year : 2013
   - Book : 31,950 (of 38,237) = 83.56% visited books assessed
   - Visit : 1,818,489 (of 1,966,341) = 92.58% visits assessed
   - Book Affected by SC-02 (incorrect date) : 3,777 (of 5,384) = 70.15% incorrect books assessed, i.e. ± 30% of books with incorrect date were not visited
   - Assessment using SQL to DQ_AST_CORR_SC02_BOOK_GA

| Correct Date | Count of PRODUCT_ID | Avg BOUNCE RATE | Sum of VISIT | Sum of TRX |
|---|---|---|---|---|
| FALSE | 3,777 | 76.7422 | 108,267 | 604 |
| TRUE | 28,173 | 78.1493 | 1,710,222 | 8,255 |
| Grand Total | 31,950 | 77.4457 | 1,818,489 | 8,859 |

   - Assessment Result:
     a. Products with correct date have more visits, however there is no advantage in bounce rates
     b. It should also be noted that the composition of incorrect date are as follows:

| SITE | Incorrect | Total | Ratio (%) |
|---|---|---|---|
| EST_EU_BS | 6 | 41,826 | 0.014 |
| EST_MEA_BS | 6 | 41,938 | 0.014 |
| EST_AU_BS | 5,378 | 40,894 | 13.151 |
| EST_US_BS | 14 | 42,026 | 0.033 |
| EST_JP_BS | 14 | 42,230 | 0.033 |
| EST_UK_BS | 6 | 41,826 | 0.014 |
| EST_ASIA_BS | 14 | 42,026 | 0.033 |

#### 4.3.2 Sales

1) Book (SC-02)
   - Year : 2013
   - Book : 6,350(of 6,536) = 97.15% sold books assessed
   - Price : USD 7,857,208.91 (of USD 9,263,705.93) = 84.81% sales revenue assessed

- Book Affected by SC-02 (incorrect date) : 545 (of 5,384) = 10.12% incorrect books assessed i.e ± 90% books with incorrect date were not sold
- Assessment using SQL to DQ_AST_CORR_SC02_BOOK_SALES

| Correct Category | Count of PRODUCT_ID | Sum of TOTAL | Sum of QTY |
|---|---|---|---|
| FALSE | 545 | 682,537.71 | 2,121 |
| TRUE | 5,805 | 7,174,671.20 | 30,906 |
| Grand Total | 6,350 | 7,857,208.91 | 33,027 |

- Assessment Result:
  a. Correctness of the date provides a better sales performance.
  b. The composition of sites with incorrect date are as in 4.3.1

# 5 Metrics Assessment Summary

## 5.1 Relationships among Metrics

The relationship between the DQ metrics, DQ attributes, and the business problems are depicted in Figure 13. The DQ attributes and metrics are all mapped for the preventive and reactive measures. The business problems could be mapped directly to the DQ attributes or mapped with only several required DQ metrics.

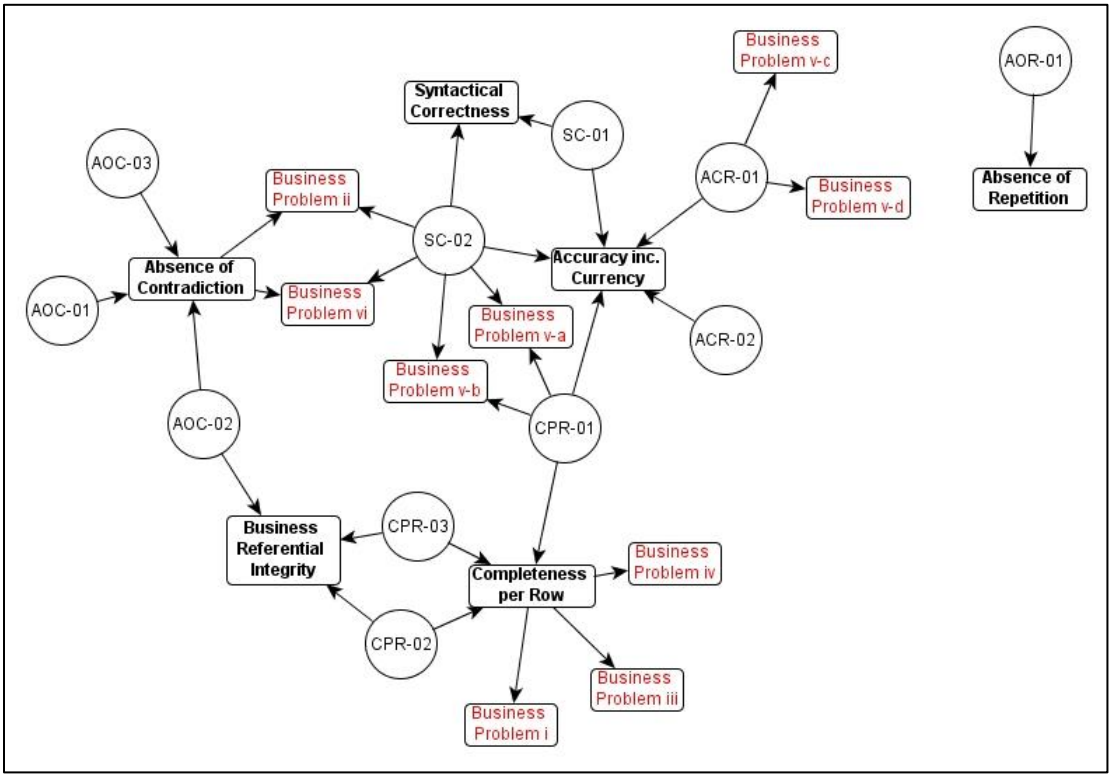


**Figure 13 Metrics Relationships**

## 5.2 Metrics Assessment Result Summary

The summary of DQ assessment in section 3 is in Table 13. The threshold values are determined by these following values:

a. Sales and Google Analytics data assessment in section 4
b. Using assessment result that is considered as historical data
c. Using a value based on the importance of the field and DQ attributes in the application

**Table 13 Metrics Assessment Result Summary**

| No | DQ Attributes | Method | Assessment Result | | Threshold 1 | Threshold 2 | Description for Threshold |
|---|---|---|---|---|---|---|---|
| | | | Book | Journal | | | |
| | Business Problems | | | | | | |
| 1. | Completeness | PROC, QUERY | 0.9265 | 0.8364 | 0.8 | 0.9 | Using GA and Sales Assessment |
| 2. | Consistency (Title-Category) | QUERY | 0.7908 | 1.00 | 0.8(1) | 0.9(1) | Using completeness |
| 3. | Accuracy (Location-Price) | PROC, QUERY | 0.9989 | 1.00 | 1 | 1 | ▪ It must be accurate 100%. Inaccurate data will result to |
| 4. | Accuracy (Location/ Format | PROC, | 0.9552 | 0.9261 | 1 | 1 | |

| No | DQ Attributes | Method | Assessment Result | | Threshold 1 | Threshold 2 | Description for Threshold |
|---|---|---|---|---|---|---|---|
| | | | Book | Journal | | | |
| | Type - Fulfillment Company Code) | QUERY | | | | | non-displayed or non-fulfilled products<br>▪ Use historical data: mean + standard deviation of 3 assessment results. |
| 5. | Accuracy (ISN) | QUERY | - | 0.8750 | 1 | 1 | |
| | Preventive and Reactive | | | | | | |
| 6. | Completeness per Row | PROC, QUERY | 0.9265 | 0.8364 | 0.8 | 0.9 | See 1 |
| 7. | Syntactical Correctness | PROC, QUERY | 0.9754 | 0.9989 | 0.9754 | 0.9754 | Use historical data: mean + standard deviation of 3 assessment results. |
| 8. | Absence of Contradiction | PROC, QUERY | 0.9542 | 0.9824 | 0.9340 | 0.9340 | |
| 9. | Absence of Repetition | QUERY | 1.00 | 1.00 | 1 | 1 | |
| 10. | Accuracy inc. Currency | PROC, QUERY | | 0.9139 | 0.9139 | 0.9139 | |

*PROC=Procedure/ Function

### 5.3 Metrics Requirement Assessment

An assessment using the criteria for the data quality metrics in (Table 14) is conducted to provide the requirement compliance level for each data quality metrics.

**Table 14 Criteria for DQ Metrics Requirements**

| No | Requirement | Code | Valuation Criteria | Method |
|---|---|---|---|---|
| | Generic | | | |
| 1 | Acceptability | DQ-R-03 | 3 = the threshold value is derived from a performance report assessment<br>2 = the threshold value is derived from best practices | Performance assessment to assess the correlation |
| 2 | Normalization | DQ-R-05 | 3 = normalized value in 0-1<br>2 = not normalized value | Assess the DQ Metrics value |
| 3 | Interval Scale | DQ-R-06 | 3 = ratio scale<br>2 = interval scale<br>1 = ordinal scale | Assess the DQ Metrics value |
| | Methods related | | | |
| 4 | Feasibility | DQ-R-07 | 3 = could be developed using simple tasks like SQL query<br>2 = need to use a simple programming (e.g. PL/SQL) to develop<br>1 = cannot be developed | Database Assessment |
| 5 | Reproducible | DQ-R-08 | 3 = tested<br>1 = not tested | Database Assessment |
| | Value and Methods | | | |

| No | Requirement | Code | Valuation Criteria | Method |
|---|---|---|---|---|
| 9 | Aggregation | DQ-R-09 | 3 = aggregation at row, table, and database level<br>2 = aggregation at field, table and database level<br>1 = no aggregation | Database Assessment |

The result of evaluation for each DQ metrics is in Table 15. The AOC-03 metrics has a low value for reproducible it is not tested due to data unavailability.

**Table 15 DQ Metrics Assessment Result**

| No | Requirement | Value | CPR-01 | CPR-02 | CPR-03 | SC-01 | SC-02 | AOC-01 | AOC-02 | AOC-03 | ACR-01 | ACR-02 | AOR-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Generic | | | | | | | | | | | | |
| 1 | Acceptability | 3.5 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| | Methods related | | | | | | | | | | | | |
| 2 | Feasibility | 4 | 2 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 3 | 3 | 3 |
| 3 | Reproducible | 3.5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 3 | 3 |
| | Value and Methods | | | | | | | | | | | | |
| 4 | Aggregation | 3.5 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

**Appendix 1. Assumptions : Valid Product and SKU**

| Name | SQL |
|---|---|
| PRODUCT | |
| V_VALID_PRD_ID | SELECT DISTINCT product_id<br>FROM DCSX_PRODUCT<br>WHERE LOWER (product_id) LIKE 'est_%'<br>AND LOWER (TRIM (product_type)) IN ('printbook', 'ebook', 'journal'); |
| V_VALID_SKU_ID | SELECT DISTINCT sku_id<br>FROM DCSX_SKU<br>WHERE LOWER (sku_id) LIKE 'est_%' AND<br>(REPLACE (SUBSTR (sku_id, 1, 23), 'SKU', 'PK') IN (SELECT DISTINCT PRODUCT_ID<br>FROM V_VALID_PRD_ID) OR<br>REPLACE (SUBSTR (sku_id, 1, 28), 'SKU', 'PK') IN (SELECT DISTINCT PRODUCT_ID FROM<br>V_VALID_PRD_ID)); |
| V_ELS_PRODUCT_REGIONAL_INFO | SELECT a."REGIONAL_INFO_ID",<br>a."PUB_DATE",<br>a."PUB_STATUS",<br>a."SHORT_DESCRIPTION",<br>a."KEY_FEATURE",<br>a."LONG_DESCRIPTION",<br>a."AUDIENCE",<br>a."QUOTES",<br>a."FULFILLMENT_COMPANY_CODE",<br>b.PRODUCT_ID,<br>b.SITE_ID<br>FROM (SELECT REGIONAL_INFO_ID,<br>PUB_DATE,<br>PUB_STATUS,<br>SHORT_DESCRIPTION,<br>KEY_FEATURE,<br>LONG_DESCRIPTION,<br>AUDIENCE,<br>QUOTES,<br>FULFILLMENT_COMPANY_CODE<br>FROM ELS_PRODUCT_REGIONAL_INFO<br>WHERE **active = 1**<br>AND REGIONAL_INFO_ID IN (SELECT DISTINCT<br>PRODUCT_REGIONAL_INFO_ID<br>FROM V_ELS_PRD_REG_MAP))a,<br>V_ELS_PRD_REG_MAP b<br>WHERE a.REGIONAL_INFO_ID = b.PRODUCT_REGIONAL_INFO_ID; |
| V_DCS_PRICE | SELECT DISTINCT "SKU_ID",<br>"PRICE_LIST",<br>"LIST_PRICE",<br>"PRICE_LIST_ID"<br>FROM (SELECT SKU_ID, PRICE_LIST, LIST_PRICE<br>FROM DCS_PRICE<br>WHERE LOWER **(price_list) IN** ('plist2310002',<br>'plist2310003','plist2310005','plist2310004',<br>'plist8380012','plist9730004') |

| Name | SQL |
|---|---|
| | AND sku_id IN (SELECT DISTINCT sku_id FROM V_VALID_SKU_ID))a,<br>(SELECT PRICE_LIST_ID<br>FROM DCS_PRICE_LIST<br>WHERE ( end_date IS NULL<br>OR (end_date IS NOT NULL AND end_date >= SYSDATE))<br>AND LOWER (PRICE_LIST_ID) IN<br>('plist2310002','plist2310003','plist2310005','plist2310004', 'plist8380012', 'plist9730004')) b<br>WHERE a.price_list = b.price_list_id; |
| BOOK OR JOURNAL | |
| V_VALID_JOURNAL_ID | SELECT DISTINCT product_id<br>FROM v_valid_prd_id<br>WHERE product_id IN (SELECT DISTINCT product_id FROM v_dcsx_product<br>WHERE LOWER (product_type) = 'journal'); |
| V_VALID_BOOK_ID | SELECT DISTINCT product_id<br>FROM v_valid_prd_id<br>WHERE product_id IN (SELECT DISTINCT product_id<br>FROM v_dcsx_product<br>WHERE LOWER (product_type) IN ('printbook', 'ebook')); |
| V_VALID_JOURNAL_SKU_ID | SELECT DISTINCT sku_id<br>FROM V_VALID_SKU_ID<br>WHERE REPLACE (SUBSTR (sku_id, 1, 23), 'SKU', 'PK') IN (SELECT DISTINCT PRODUCT_ID<br>FROM V_VALID_JOURNAL_ID); |
| V_VALID_BOOK_SKU_ID | SELECT DISTINCT sku_id<br>FROM V_VALID_SKU_ID<br>WHERE REPLACE (SUBSTR (sku_id, 1, 28), 'SKU', 'PK') IN (SELECT DISTINCT PRODUCT_ID<br>FROM V_VALID_BOOK_ID); |
| BOOK-SKU | BOOK.PRODUCT_ID = REPLACE(SUBSTR(BOOK.SKU_ID,1,28),'SKU',PK') |
| JOURNAL-SKU | JOURNAL.PRODUCT_ID = REPLACE(SUBSTR(JOURNAL.SKU_ID,1,23),'SKU',PK') |

## Appendix 2. Database Objects for DQ Assessment

**Table 16 Database Objects for DQ Assessment**

| No | Name | Type | # Records | # (unique) Records | Purpose |
|---|---|---|---|---|---|
| | PRODUCT | | | | |
| 1 | V_VALID_PRD_ID | VIEW | 55,050 | 55,050 | Data Selection |
| 2 | V_VALID_SKU_ID | VIEW | 109,083 | 109,083 | Data Selection |
| 3 | V_DCSX_PRODUCT | VIEW | 55,050 | 55,050 | Data Selection |
| 4 | V_ANG_PRODUCT | VIEW | 34,129 | 34,129 | Data Selection |
| 5 | V_ELS_PRODUCT | VIEW | 55,050 | 55,050 | Data Selection |
| 6 | V_ELS_PRODUCT_EN | VIEW | 52,437 | 52,437 | Data Selection |
| 7 | V_PRD_PRNT_CATS | VIEW | 54,652 | 54,652 | Data Selection |
| 8 | V_ELS_PRD_REG_MAP | VIEW | 385,238 | 55,034 | Data Selection |
| 9 | V_ELS_PRODUCT_REGIONAL_INFO | VIEW | 307,517 | 45,593 | Data Selection |
| 10 | V_DCSX_SKU | VIEW | 109,083 | 109,083 | Data Selection |
| 11 | V_ELS_SKU | VIEW | 108,113 | 108,113 | Data Selection |
| 12 | V_ELS_DCSX_SKU | VIEW | 109,034 | 109,034 | Data Selection |
| 13 | V_DCS_PRICE | VIEW | 341,556 | 109,047 | Data Selection |
| | BOOK OR JOURNAL | | | | |
| 14 | V_VALID_JOURNAL_ID | VIEW | 2,273 | 2,273 | Data Selection |
| 15 | V_VALID_BOOK_ID | VIEW | 52,777 | 52,777 | Data Selection |
| 16 | V_VALID_JOURNAL_SKU_ID | VIEW | 5,634 | 5,634 | Data Selection |
| 17 | V_VALID_BOOK_SKU_ID | VIEW | 103,449 | 103,449 | Data Selection |
| 18 | T_BOOK_DCSX_PRODUCT | TABLE | 52,777 | 52,777 | Data Selection |
| 19 | V2_JOURNAL_DCSX_PRODUCT | VIEW | 2,273 | 2,273 | Data Selection |
| 20 | V2_JOURNAL_ANG_PRODUCT | VIEW | 2,273 | 2,273 | Data Selection |
| 21 | T_BOOK_ELS_PRODUCT | TABLE | 52,777 | 52,777 | Data Selection |
| 22 | T_JOURNAL_ELS_PRODUCT | TABLE | 2,273 | 2,273 | Data Selection |
| 23 | T_BOOK_ELS_PRODUCT_EN | TABLE | 50,169 | 50,169 | Data Selection |
| 24 | V2_JOURNAL_ELS_PRODUCT_EN | VIEW | 2,268 | 2,268 | Data Selection |
| 25 | V2_BOOK_DCSX_SKU | VIEW | 103,449 | 103,449 | Data Selection |
| 26 | V2_JOURNAL_DCSX_SKU | VIEW | 5,634 | 5,634 | Data Selection |
| 27 | V2_BOOK_ELS_DCSX_SKU | VIEW | 103,400 | 103,400 | Data Selection |
| 28 | V2_JOURNAL_ELS_DCSX_SKU | VIEW | 5,634 | 5,634 | Data Selection |
| 29 | V2_BOOK_ELS_SKU | VIEW | 102,479 | 102,479 | Data Selection |
| 30 | V2_JOURNAL_ELS_SKU | VIEW | 5,634 | 5,634 | Data Selection |
| 31 | V2_BOOK_PRD_PRNT_CATS | VIEW | 52,401 | 52,401 | Data Selection |
| 32 | V2_JOURNAL_PRD_PRNT_CATS | VIEW | 2,251 | 2,251 | Data Selection |
| | [BOOK/ JOURNAL] INFO AND SKU | | | | |
| 33 | V2_BOOK_DCS_PRICE | VIEW | 321,417 | 103,414 | Data Selection |
| 34 | V2_JOURNAL_DCS_PRICE | VIEW | 20,139 | 5,633 | Data Selection |
| 35 | T_BOOK_ELS_PRD_REG_INFO | TABLE | 292,647 | 43,464 | Data Selection |

| No | Name | Type | # Records | # (unique) Records | Purpose |
|---|---|---|---|---|---|
| 36 | T_JOURNAL_ELS_PRD_REG_INFO | TABLE | 14,870 | 2,129 | Data Selection |
| 37 | V2_BOOK_INFO | VIEW | 43,464 | 43,464 | Data Selection |
| 38 | V2_JOURNAL_INFO | VIEW | 2,129 | 2,129 | Data Selection |
| 39 | V2_BOOK_SKU | VIEW | 103,449 | 103,449 | Data Selection |
| 40 | V2_JOURNAL_SKU | VIEW | 5,634 | 5,634 | Data Selection |
| | LOV | | | | |
| 41 | DQ_LOV_BOOK_PRODUCT_TYPE | VIEW | | | List of Values |
| 42 | DQ_LOV_BOOK_IMPRINT | VIEW | | | List of Values |
| 43 | DQ_LOV_BOOK_WEB_PT_NAME | VIEW | | | List of Values |
| 44 | DQ_LOV_BOOK_CATEGORY_ID | VIEW | | | List of Values |
| 45 | DQ_LOV_BOOK_SKU_TYPE | VIEW | | | List of Values |
| 46 | DQ_LOV_BOOK_FULFILL_CC1 | VIEW | | | List of Values |
| 47 | DQ_LOV_BOOK_OUT_OF_PRINT | VIEW | | | List of Values |
| 48 | DQ_LOV_BOOK_PURCHASE_TYPE | VIEW | | | List of Values |
| 49 | DQ_LOV_BOOK_PRICE_LIST | VIEW | | | List of Values |
| 50 | DQ_LOV_BOOK_PUB_DATE | VIEW | | | List of Values |
| 51 | DQ_LOV_BOOK_PUB_STATUS | VIEW | | | List of Values |
| 52 | DQ_LOV_BOOK_FULFILL_CC | VIEW | | | List of Values |
| 53 | DQ_LOV_BOOK_SITE_ID | VIEW | | | List of Values |
| 54 | DQ_LOV_JOURNAL_PRODUCT_TYPE | VIEW | | | List of Values |
| 55 | DQ_LOV_JOURNAL_WEB_PT_NAME | VIEW | | | List of Values |
| 56 | DQ_LOV_JOURNAL_CATEGORY_ID | VIEW | | | List of Values |
| 57 | DQ_LOV_JOURNAL_SKU_TYPE | VIEW | | | List of Values |
| 58 | DQ_LOV_JOURNAL_FULFILL_CC1 | VIEW | | | List of Values |
| 59 | DQ_LOV_JOURNAL_OUT_OF_PRINT | VIEW | | | List of Values |
| 60 | DQ_LOV_JOURNAL_PURCHASE_TYPE | VIEW | | | List of Values |
| 61 | DQ_LOV_JOURNAL_PRICE_LIST | VIEW | | | List of Values |
| 62 | DQ_LOV_JOURNAL_PUB_STATUS | VIEW | | | List of Values |
| 63 | DQ_LOV_JOURNAL_FULFILL_CC | VIEW | | | List of Values |
| 64 | DQ_LOV_JOURNAL_SITE_ID | VIEW | | | List of Values |
| | SOURCE ID | | | | |
| 65 | O_JOURNAL_ID_PRICE_2013 | TABLE | 2,212 | 2,212 | |
| 66 | O_BOOK_ID_PRICE_2013 | TABLE | | | |
| | DQ ASSESSMENT | | | | |
| | COMPLETENESS | | | | |
| 67 | DQ_COMP_BOOK | TABLE | 43,463 | 43,463 | Temp Table |
| 68 | DQ_COMP_JOURNAL | TABLE | 2,129 | 2,129 | Temp Table |
| 69 | DQ_COMP_BOOK_INFO | TABLE | 43,464 | 43,464 | Temp Table |
| 70 | DQ_COMP_BOOK_REG | TABLE | 292,647 | 43,464 | Temp Table |
| 71 | DQ_COMP_PRD_BOOK | TABLE | 43,464 | 43,464 | Temp Table |

| No | Name | Type | # Records | # (unique) Records | Purpose |
|---|---|---|---|---|---|
| 72 | DQ_COMP_BOOK_SKU | TABLE | 103,449 | 103,449 | Temp Table |
| 73 | DQ_COMP_BOOK_PRICE | TABLE | 321,417 | 103,414 | Temp Table |
| 74 | DQ_COMP_SKU_BOOK | TABLE | 103,414 | 103,414 | Temp Table |
| 75 | DQ_COMP_JOURNAL_INFO | TABLE | 2,129 | 2,129 | Temp Table |
| 76 | DQ_COMP_JOURNAL_REG | TABLE | 14,870 | 2,129 | Temp Table |
| 77 | DQ_COMP_PRD_JOURNAL | TABLE | 2,129 | 2,129 | Temp Table |
| 78 | DQ_COMP_JOURNAL_SKU | TABLE | 5,634 | 5,634 | Temp Table |
| 79 | DQ_COMP_JOURNAL_PRICE | TABLE | 20,139 | 5,633 | Temp Table |
| 80 | DQ_COMP_SKU_JOURNAL | TABLE | 5,633 | 5,633 | Temp Table |
| 81 | PROC_DQ_COMP_JOURNAL_INFO | PROCEDURE | NA | NA | DQ Logic |
| 82 | PROC_DQ_COMP_JOURNAL_REG | PROCEDURE | NA | NA | DQ Logic |
| 83 | PROC_DQ_COMP_JOURNAL_SKU | PROCEDURE | NA | NA | DQ Logic |
| 84 | PROC_DQ_COMP_JOURNAL_PRICE | PROCEDURE | NA | NA | DQ Logic |
| 85 | PROC_DQ_COMP_JOURNAL | PROCEDURE | NA | NA | DQ Logic |
| 86 | PROC_DQ_COMP_BOOK_INFO | PROCEDURE | NA | NA | DQ Logic |
| 87 | PROC_DQ_COMP_BOOK_REG | PROCEDURE | NA | NA | DQ Logic |
| 88 | PROC_DQ_COMP_BOOK_SKU | PROCEDURE | NA | NA | DQ Logic |
| 89 | PROC_DQ_COMP_BOOK_PRICE | PROCEDURE | NA | NA | DQ Logic |
| 90 | PROC_DQ_COMP_BOOK | PROCEDURE | NA | NA | DQ Logic |
| | CONSISTENCY | | | | |
| 91 | DQ_CONS_BOOK_TITLE1_CAT | VIEW | | | DQ Logic |
| 92 | DQ_CONS_JOURNAL_TITLE1_CAT | VIEW | | | DQ Logic |
| 93 | DQ_CONS_SITE_PRICE_BOOK | VIEW | | | DQ Logic |
| 94 | DQ_CONS_SITE_PRICE_JOURNAL | VIEW | | | DQ Logic |
| 95 | DQ_CONS_FULFILLMENT_BOOK | VIEW | | | DQ Logic |
| 96 | DQ_CONS_FULFILLMENT_JOURNAL | VIEW | | | DQ Logic |
| | CORRECTNESS | | | | |
| 97 | DQ_CORR_BOOK_INFO | VIEW | | | DQ Logic |
| 98 | DQ_CORR_BOOK_REG_INFO | VIEW | | | DQ Logic |
| 99 | DQ_CORR_BOOK_PRICE | VIEW | | | DQ Logic |
| 100 | DQ_CORR_JOURNAL_INFO | VIEW | | | DQ Logic |
| 101 | DQ_CORR_JOURNAL_REG_INFO | VIEW | | | DQ Logic |
| 102 | DQ_CORR_JOURNAL_PRICE | VIEW | | | DQ Logic |
| | PERFORMANCE ASSESSMENT | | | | |
| 103 | O_GA_BOOK_2013 | TABLE | 118,290 | 38,237 | |
| 104 | O_GA_JOURNAL_2013 | TABLE | 2,609 | 1,223 | |
| 105 | O_BOOK_SALES_2013 | TABLE | 36,835 | 6,536 | |
| 106 | O_JOURNAL_SALES_2013 | TABLE | 3,512 | 822 | |
| 107 | DQ_AST_COMP_BOOK_GA | VIEW | 31,949 | 31,949 | |
| 108 | DQ_AST_ COMP_JOURNAL_GA | VIEW | 1,185 | 1,185 | |

| No | Name | Type | # Records | # (unique) Records | Purpose |
|---|---|---|---|---|---|
| 109 | DQ_AST_ COMP_BOOK_SALES | VIEW | | | DQ Logic |
| 110 | DQ_AST_ COMP_JOURNAL_SALES | VIEW | | | DQ Logic |
| 111 | DQ_AST_CORR_BOOK_GA | VIEW | | | DQ Logic |
| 112 | DQ_AST_CORR_BOOK_SALES | VIEW | | | DQ Logic |
| 113 | DQ_AST_CORR_JOURNAL_GA | VIEW | | | DQ Logic |
| 114 | DQ_AST_CORR_JOURNAL_SALES | VIEW | | | DQ Logic |
| 115 | DQ_AST_CORR_SC-02_BOOK_GA | VIEW | | | DQ Logic |
| 116 | DQ_AST_CORR_SC-02_BOOK_SALES | VIEW | | | DQ Logic |

**Appendix 3. List of Values Assessment**

1) Book
   a) Info
      i) PRODUCT_TYPE

| Values | Records | Total | % |
|---|---|---|---|
| Ebook | 24,128 | 43,464 | 55.5126 |
| Printbook | 19,336 | 43,464 | 44.4874 |

      ii) IMPRINT

| Values | Records | Total | % |
|---|---|---|---|
| Academic Press | 14,728 | 43,464 | 33.8855 |
| Elsevier Science | 5,101 | 43,464 | 11.7361 |
| Saunders | 3,762 | 43,464 | 8.6554 |
| Elsevier | 2,075 | 43,464 | 4.7741 |
| Butterworth-Heinemann | 2,066 | 43,464 | 4.7534 |
| Woodhead Publishing | 1,716 | 43,464 | 3.9481 |
| Churchill Livingstone | 1,634 | 43,464 | 3.7594 |
| Mosby | 1,624 | 43,464 | 3.7364 |
| North Holland | 1,506 | 43,464 | 3.4649 |
| Morgan Kaufmann | 1,037 | 43,464 | 2.3859 |
| Elsevier Masson | 950 | 43,464 | 2.1857 |
| Newnes | 894 | 43,464 | 2.0569 |
| William Andrew | 829 | 43,464 | 1.9073 |
| URBFI | 764 | 43,464 | 1.7578 |
| Syngress | 742 | 43,464 | 1.7072 |
| Chandos Publishing | 710 | 43,464 | 1.6335 |
| Pergamon | 425 | 43,464 | 0.9778 |
| Gulf Professional Publishing | 343 | 43,464 | 0.7892 |
| Saunders Ltd. | 297 | 43,464 | 0.6833 |
| Anderson | 241 | 43,464 | 0.5545 |
| Elsevier Editora Ltda. | 239 | 43,464 | 0.5499 |
| JAI Press | 226 | 43,464 | 0.5200 |
| Elsevier India | 193 | 43,464 | 0.4440 |
| Mosby Ltd. | 183 | 43,464 | 0.4210 |
| Elsevier Urban & Partner | 183 | 43,464 | 0.4210 |
| Digital Press | 177 | 43,464 | 0.4072 |
| Churchill Livingstone Australia | 173 | 43,464 | 0.3980 |
| Elsevier Srl | 158 | 43,464 | 0.3635 |
| Bailliere Tindall | 135 | 43,464 | 0.3106 |
| Gulf Publishing Company | 55 | 43,464 | 0.1265 |
| Mosby Australia | 46 | 43,464 | 0.1058 |
| Hanley & Belfus | 43 | 43,464 | 0.0989 |

| Values | Records | Total | % |
|---|---|---|---|
| Books for Midwives | 37 | 43,464 | 0.0851 |
| ChemTec Publishing | 25 | 43,464 | 0.0575 |
| ACACL | 17 | 43,464 | 0.0391 |
| Mosby/JEMS | 16 | 43,464 | 0.0368 |
| Oily Press | 15 | 43,464 | 0.0345 |
| CMP | 14 | 43,464 | 0.0322 |
| Vincentz | 10 | 43,464 | 0.0230 |
| Mosby Canada | 9 | 43,464 | 0.0207 |
| Saunders Australia | 8 | 43,464 | 0.0184 |
| Elsevier Poland | 8 | 43,464 | 0.0184 |
| Saunders Canada | 6 | 43,464 | 0.0138 |
| AP Cell | 6 | 43,464 | 0.0138 |
| BH/Optician | 6 | 43,464 | 0.0138 |
| Made Simple | 5 | 43,464 | 0.0115 |
| Merck Manuals | 5 | 43,464 | 0.0115 |
| Architectural Press | 4 | 43,464 | 0.0092 |
| AMIRSYS | 4 | 43,464 | 0.0092 |
| Harcourt India | 3 | 43,464 | 0.0069 |
| Abington Publishing | 3 | 43,464 | 0.0069 |
| BH/BCLA | 3 | 43,464 | 0.0069 |
| Laxton's | 1 | 43,464 | 0.0023 |
| Wright | 1 | 43,464 | 0.0023 |
| The Lancet | 1 | 43,464 | 0.0023 |
| Pergamon Flexible Learning | 1 | 43,464 | 0.0023 |
| Trends | 1 | 43,464 | 0.0023 |

iii) WEB_PRODUCT_TYPE_NAME

| Values | Records | Total | % |
|---|---|---|---|
| Electron | 20,694 | 43,464 | 47.6118 |
| Hardcover | 9,298 | 43,464 | 21.3924 |
| E-book | 3,065 | 43,464 | 7.0518 |
| Paperback | 2,795 | 43,464 | 6.4306 |
| BASIC TEXT - CLOTH | 2,285 | 43,464 | 5.2572 |
| BASIC TEXT - PAPER | 2,230 | 43,464 | 5.1307 |
| NULL | 1,946 | 43,464 | 4.4773 |
| EXPERT CONSULT | 572 | 43,464 | 1.3160 |
| CD-ROM | 71 | 43,464 | 0.1634 |
| SPIRAL BOUND | 70 | 43,464 | 0.1611 |
| DVD | 61 | 43,464 | 0.1403 |
| PACKAGE | 54 | 43,464 | 0.1242 |
| Hardcover w/ CD-ROM | 47 | 43,464 | 0.1081 |

| Values | Records | Total | % |
|---|---|---|---|
| EVOLVE E-ONLY | 37 | 43,464 | 0.0851 |
| ONLINE COURSES | 36 | 43,464 | 0.0828 |
| Paperback w/ CD-ROM | 28 | 43,464 | 0.0644 |
| CARD KITS | 25 | 43,464 | 0.0575 |
| EVOLVE P+E PKG | 21 | 43,464 | 0.0483 |
| Loose-leaf | 18 | 43,464 | 0.0414 |
| BOOK-HARDCOVER | 16 | 43,464 | 0.0368 |
| Poster | 14 | 43,464 | 0.0322 |
| Spiral bound | 12 | 43,464 | 0.0276 |
| SALE ON-LINE MATERIAL | 10 | 43,464 | 0.0230 |
| Digital or multimedia/Electron | 8 | 43,464 | 0.0184 |
| Diskette | 6 | 43,464 | 0.0138 |
| Spiral bound w/ CD-ROM | 5 | 43,464 | 0.0115 |
| E-DITION WITH BOOK | 4 | 43,464 | 0.0092 |
| Audio | 4 | 43,464 | 0.0092 |
| LABORATORY MANUAL | 3 | 43,464 | 0.0069 |
| Cards | 3 | 43,464 | 0.0069 |
| SLS | 3 | 43,464 | 0.0069 |
| Software package | 3 | 43,464 | 0.0069 |
| MONOGRAPH - CLOTH | 3 | 43,464 | 0.0069 |
| PAPERBACK | 3 | 43,464 | 0.0069 |
| HARDCOVER | 2 | 43,464 | 0.0046 |
| AP TEXTBOOK - PAPER | 1 | 43,464 | 0.0023 |
| WEB STANDALONE | 1 | 43,464 | 0.0023 |
| Softcover | 1 | 43,464 | 0.0023 |
| Slides | 1 | 43,464 | 0.0023 |
| Book/Hardback | 1 | 43,464 | 0.0023 |
| Digital | 1 | 43,464 | 0.0023 |
| PDA | 1 | 43,464 | 0.0023 |
| EXPERT CONSULT ONLINE | 1 | 43,464 | 0.0023 |
| OTHER | 1 | 43,464 | 0.0023 |
| HANDBOOK | 1 | 43,464 | 0.0023 |
| NON SALE ON-LINE MATERIAL | 1 | 43,464 | 0.0023 |
| MONOGRAPH - PAPER | 1 | 43,464 | 0.0023 |

iv) CATEGORY_ID (Top-10)

| Values | Records | Total | % |
|---|---|---|---|
| EST_SA-8-85 | 1,587 | 43,464 | 3.6513 |
| EST_IMP-23 | 1,098 | 43,464 | 2.5262 |
| EST_SA-7-54 | 949 | 43,464 | 2.1834 |
| EST_SA-8-75 | 935 | 43,464 | 2.1512 |

| | | | |
|---|---|---|---|
| EST_SA-8-83 | 846 | 43,464 | 1.9464 |
| EST_SA-9-91-206 | 788 | 43,464 | 1.8130 |
| EST_IND-20 | 677 | 43,464 | 1.5576 |
| EST_SA-9-92-254 | 615 | 43,464 | 1.4150 |
| EST_SA-8-82-141 | 573 | 43,464 | 1.3183 |
| EST_SA-4-30 | 548 | 43,464 | 1.2608 |

b) Product Regional Info

i) PUB_DATE

| Values | Records | Total | % |
|---|---|---|---|
| 1799 | 91,159 | 292,647 | 31.1498 |
| 2013 | 19,360 | 292,647 | 6.6155 |
| 2012 | 18,589 | 292,647 | 6.3520 |
| 2010 | 15,358 | 292,647 | 5.2480 |
| 2011 | 15,296 | 292,647 | 5.2268 |
| 2009 | 14,069 | 292,647 | 4.8075 |
| 2008 | 13,323 | 292,647 | 4.5526 |
| 2007 | 13,073 | 292,647 | 4.4672 |
| 2006 | 10,901 | 292,647 | 3.7250 |
| 2005 | 9,158 | 292,647 | 3.1294 |
| 2004 | 7,861 | 292,647 | 2.6862 |
| 2003 | 7,290 | 292,647 | 2.4911 |
| 2001 | 6,415 | 292,647 | 2.1921 |
| 2002 | 6,336 | 292,647 | 2.1651 |
| 1900 | 5,438 | 292,647 | 1.8582 |
| 2000 | 5,389 | 292,647 | 1.8415 |
| 1999 | 4,867 | 292,647 | 1.6631 |
| 1998 | 4,808 | 292,647 | 1.6429 |
| 2014 | 4,478 | 292,647 | 1.5302 |
| 1997 | 3,852 | 292,647 | 1.3163 |
| 1996 | 3,614 | 292,647 | 1.2349 |
| 1995 | 2,859 | 292,647 | 0.9769 |
| 1994 | 1,923 | 292,647 | 0.6571 |
| 1993 | 1,212 | 292,647 | 0.4142 |
| 1992 | 1,180 | 292,647 | 0.4032 |
| 1991 | 987 | 292,647 | 0.3373 |
| 1990 | 777 | 292,647 | 0.2655 |
| 1988 | 433 | 292,647 | 0.1480 |
| 1989 | 430 | 292,647 | 0.1469 |
| 1987 | 294 | 292,647 | 0.1005 |
| 1986 | 262 | 292,647 | 0.0895 |

| Values | Records | Total | % |
|---|---|---|---|
| 2015 | 202 | 292,647 | 0.0690 |
| 1983 | 201 | 292,647 | 0.0687 |
| 1984 | 185 | 292,647 | 0.0632 |
| 1985 | 162 | 292,647 | 0.0554 |
| 1982 | 146 | 292,647 | 0.0499 |
| 1981 | 131 | 292,647 | 0.0448 |
| 1980 | 125 | 292,647 | 0.0427 |
| 1977 | 80 | 292,647 | 0.0273 |
| 1975 | 76 | 292,647 | 0.0260 |
| 1978 | 58 | 292,647 | 0.0198 |
| 1979 | 47 | 292,647 | 0.0161 |
| 1971 | 41 | 292,647 | 0.0140 |
| 1972 | 26 | 292,647 | 0.0089 |
| 1973 | 18 | 292,647 | 0.0062 |
| 1976 | 17 | 292,647 | 0.0058 |
| 1955 | 16 | 292,647 | 0.0055 |
| 1964 | 14 | 292,647 | 0.0048 |
| 1967 | 12 | 292,647 | 0.0041 |
| 1966 | 12 | 292,647 | 0.0041 |
| 1969 | 12 | 292,647 | 0.0041 |
| 1961 | 10 | 292,647 | 0.0034 |
| 1970 | 9 | 292,647 | 0.0031 |
| 1974 | 9 | 292,647 | 0.0031 |
| 2020 | 9 | 292,647 | 0.0031 |
| 2016 | 9 | 292,647 | 0.0031 |
| 1962 | 7 | 292,647 | 0.0024 |
| 1959 | 7 | 292,647 | 0.0024 |
| 2019 | 6 | 292,647 | 0.0021 |
| 1957 | 6 | 292,647 | 0.0021 |
| 2021 | 3 | 292,647 | 0.0010 |

ii) PUB_STATUS

| Values | Records | Total | % |
|---|---|---|---|
| 132 | 276,309 | 292,647 | 94.4172 |
| 133 | 11,937 | 292,647 | 4.0790 |
| 127 | 4,401 | 292,647 | 1.5039 |

iii) FULFILLMENT_COMPANY_CODE

| Values | Records | Total | % |
|---|---|---|---|
| DELTA-E-COMMERCE | 148,688 | 292,647 | 50.8080 |
| COPS-E-COMMERCE | 103,064 | 292,647 | 35.2179 |

| | | | |
|---|---|---|---|
| BOOKMASTER-E-COMMERCE | 40,894 | 292,647 | 13.9738 |
| COPS-ECOMMEauRCE | 1 | 292,647 | 0.0003 |

iv) SITE_ID

| Values | Records | Total | % |
|---|---|---|---|
| EST_JP_BS | 42,230 | 292,647 | 14.4304 |
| EST_US_BS | 42,026 | 292,647 | 14.3606 |
| EST_ASIA_BS | 42,026 | 292,647 | 14.3606 |
| EST_MEA_BS | 41,938 | 292,647 | 14.3306 |
| EST_UK_BS | 41,826 | 292,647 | 14.2923 |
| EST_EU_BS | 41,707 | 292,647 | 14.2516 |
| EST_AU_BS | 40,894 | 292,647 | 13.9738 |

c) SKU

i) SKU_TYPE

| Values | Records | Total | % |
|---|---|---|---|
| Physical | 25,464 | 103,449 | 24.6150 |
| DRM | 20,535 | 103,449 | 19.8504 |
| VST | 19,575 | 103,449 | 18.9224 |
| GL | 19,148 | 103,449 | 18.5096 |
| E3 | 9,975 | 103,449 | 9.6424 |
| EBSD | 8,702 | 103,449 | 8.4119 |
| NULL | 49 | 103,449 | 0.0474 |
| DF | 1 | 103,449 | 0.0010 |

ii) FULFILLMENT_COMPANY_CODE

| Values | Records | Total | % |
|---|---|---|---|
| DELTA-E-COMMERCE | 69,111 | 103,449 | 66.8068 |
| NULL | 25,636 | 103,449 | 24.7813 |
| CRM-E-COMMERCE | 8,702 | 103,449 | 8.4119 |

iii) OUT_OF_PRINT

| Values | Records | Total | % |
|---|---|---|---|
| 0 | 102,479 | 103,449 | 99.0623 |
| -99 | 970 | 103,449 | 0.9377 |

iv) PURCHASE_TYPE

| Values | Records | Total | % |
|---|---|---|---|
| Digital or multimedia/Electron | 68,860 | 103,449 | 66.5642 |
| Book/Hardback | 15,214 | 103,449 | 14.7068 |
| Digital or multimedia/Online f | 9,045 | 103,449 | 8.7434 |
| Book/Paperback | 8,393 | 103,449 | 8.1132 |
| NULL | 1,257 | 103,449 | 1.2151 |

| Values | Records | Total | % |
|---|---|---|---|
| Digital or multimedia/CD-ROM | 168 | 103,449 | 0.1624 |
| Book/Spiral bound | 156 | 103,449 | 0.1508 |
| Miscellaneous printed material | 99 | 103,449 | 0.0957 |
| Digital or multimedia/DVD | 71 | 103,449 | 0.0686 |
| Book/Other book | 52 | 103,449 | 0.0503 |
| Book/Book | 35 | 103,449 | 0.0338 |
| Book/Loose-leaf | 22 | 103,449 | 0.0213 |
| Video/Video | 20 | 103,449 | 0.0193 |
| Digital or multimedia/Diskette | 17 | 103,449 | 0.0164 |
| Audio/Audio | 12 | 103,449 | 0.0116 |
| Digital or multimedia/Digital | 6 | 103,449 | 0.0058 |
| Promotional and trade-only/Shr | 6 | 103,449 | 0.0058 |
| Digital or multimedia/Other di | 4 | 103,449 | 0.0039 |
| Video/Videodisk | 3 | 103,449 | 0.0029 |
| Film/Slides | 3 | 103,449 | 0.0029 |
| Mixed media and retail packs/M | 2 | 103,449 | 0.0019 |
| Book/Board book | 2 | 103,449 | 0.0019 |
| Maps/Sheet map, flat | 1 | 103,449 | 0.0010 |
| Video/Other video format | 1 | 103,449 | 0.0010 |

d) Price

i) PRICE_LIST

| Values | Records | Total | % |
|---|---|---|---|
| plist2310002 | 103,246 | 321,417 | 32.1221 |
| plist2310003 | 102,326 | 321,417 | 31.8359 |
| plist8380012 | 85,521 | 321,417 | 26.6075 |
| plist2310005 | 21,611 | 321,417 | 6.7237 |
| plist2310004 | 8,713 | 321,417 | 2.7108 |

2) Journal

a) Info

i) PRODUCT_TYPE

| Values | Records | Total | % |
|---|---|---|---|
| Journal | 2,129 | 2,129 | 100.0000 |

ii) WEB_PRODUCT_TYPE_NAME

| Values | Records | Total | % |
|---|---|---|---|
| Journal | 2,125 | 2,129 | 99.8121 |
| NULL | 4 | 2,129 | 0.1879 |

iii) CATEGORY_ID (Top-10)

| Values | Records | Total | % |
|---|---|---|---|

| | | | |
|---|---|---|---|
| EST_SA-9-91-206 | 50 | 2,129 | 2.3485 |
| EST_SA-6-42 | 47 | 2,129 | 2.2076 |
| EST_SA-8-82-141 | 44 | 2,129 | 2.0667 |
| EST_SA-9-91-205 | 39 | 2,129 | 1.8318 |
| EST_SA-8-73 | 38 | 2,129 | 1.7849 |
| EST_SA-13-126-375 | 37 | 2,129 | 1.7379 |
| EST_SA-9-91 | 37 | 2,129 | 1.7379 |
| EST_SA-7-50-45 | 35 | 2,129 | 1.6440 |
| EST_SA-9-92-254 | 33 | 2,129 | 1.5500 |
| EST_SA-9-91-247 | 33 | 2,129 | 1.5500 |

b) Product Regional Info

i) PUB_STATUS

| Values | Records | Total | % |
|---|---|---|---|
| 132 | 14,870 | 14,870 | 100.0000 |

ii) FULFILLMENT_COMPANY_CODE

| Values | Records | Total | % |
|---|---|---|---|
| DELTA-E-COMMERCE | 8,064 | 14,870 | 54.2300 |
| COPS-E-COMMERCE | 4,677 | 14,870 | 31.4526 |
| BOOKMASTER-E-COMMERCE | 2,129 | 14,870 | 14.3174 |

iii) SITE_ID

| Values | Records | Total | % |
|---|---|---|---|
| EST_AU_BS | 2,129 | 14,870 | 14.3174 |
| EST_MEA_BS | 2,129 | 14,870 | 14.3174 |
| EST_ASIA_BS | 2,129 | 14,870 | 14.3174 |
| EST_US_BS | 2,129 | 14,870 | 14.3174 |
| EST_UK_BS | 2,118 | 14,870 | 14.2434 |
| EST_EU_BS | 2,118 | 14,870 | 14.2434 |
| EST_JP_BS | 2,118 | 14,870 | 14.2434 |

c) SKU

i) SKU_TYPE

| Values | Records | Total | % |
|---|---|---|---|
| PJPROMIS | 2,948 | 5,634 | 52.3252 |
| EJSD | 2,087 | 5,634 | 37.0430 |
| PJARGI | 599 | 5,634 | 10.6319 |

ii) FULFILLMENT_COMPANY_CODE

| Values | Records | Total | % |
|---|---|---|---|
| DELTA-E-COMMERCE | 2,948 | 5,634 | 52.3252 |
| CRM-E-COMMERCE | 2,087 | 5,634 | 37.0430 |

| | | | |
|---|---|---|---|
| ARGI-E-COMMERCE | 598 | 5,634 | 10.6141 |
| NULL | 1 | 5,634 | 0.0177 |

iii) OUT_OF_PRINT

| Values | Records | Total | % |
|---|---|---|---|
| 0 | 5,634 | 5,634 | 100.0000 |

iv) PURCHASE_TYPE

| Values | Records | Total | % |
|---|---|---|---|
| Journal | 5,634 | 5,634 | 100.0000 |

d) Price

i) PRICE_LIST

| Values | Records | Total | % |
|---|---|---|---|
| plist2310002 | 5,633 | 20,139 | 27.9706 |
| plist2310003 | 5,046 | 20,139 | 25.0559 |
| plist2310004 | 5,026 | 20,139 | 24.9566 |
| plist8380012 | 3,864 | 20,139 | 19.1867 |
| plist9730004 | 570 | 20,139 | 2.8303 |

**Appendix 4. Data Quality Metrics Specification for e-commerce**

Measuring Point : e-commerce Database [Attribute 3]

**Table 17 Metrics Specification for Business Problems**

| No | Business Problem | Business Impact | Data Defect | DQ Dimensions | Attribute |
|---|---|---|---|---|---|
| i | Customer does not buy a product | Potential revenue loss | Incomplete information in e-commerce system (Book database) [Attribute 10] | Completeness per row [Attribute 2] | All in websites data model [Attribute 9] |
| | ▪ Measurement Method : [Attribute 1-2, 4-8]<br>1. CPR-01 : Ratio of Record with non-blank or non-null field in Product Repository<br>2. CPR-02 : Ratio of NON-Parentless Record in Product Repository (e.g. SKU is referenced by a Product)<br>3. CPR-03 : Ratio of NON-Childless Record in Product Repository (e.g. Product has SKU)<br>4. TOTAL : 70%xCPR-01 + 15%xCPR-02 +15%xCPR-03<br>▪ Frequency : Daily [Attribute 7]<br>▪ Value : [0-1] [Attribute 5-6]<br>▪ Expected Threshold : | | | | |
| | | | | | |
| ii | Customer could not browse the site conveniently | Customer dissatisfaction | Ambiguous data in Book database (taxonomy mapping problem) | Absence of contradictions (consistency) | Subject, Parent Category |
| | There are 2 problems here :<br>a. Wrong mapping -> consistency issue<br>b. Different taxonomy -> taxonomy mapping issue<br>For the problem (a) we can use this measurement:<br>▪ Measurement Method :<br>1. AOC-01 : Ratio of reasonable fields (subject related)<br>Reasonable field : field that has the same top-5 values on the basis of its distribution compared with previous data<br>2. AOC-02 : Ratio of record which adhere business rule<br>example : Title-Subject : Ratio (i) = number of records with Title (i) and Subject (i)/ number of records with Title (i)<br>3. SC-02 : Ratio of record which has non deviated value in Product Repository (List of Values)<br>4. TOTAL : 15%xAOC-01 + 15%xSC-02 + 70%xAOC-02<br>▪ Frequency : Daily<br>▪ Value : [0-1]<br>▪ Expected Threshold : | | | | |
| iii | Unable to run marketing campaign using AdWords and Email channel | Potential revenue loss | Incomplete information in e-commerce system | Completeness per row | All in marketing data model |
| | ▪ see (i) mapping problem | | | | |
| iv | Internet user could not find the data in top result using search engine | Potential revenue loss | Incomplete information in e-commerce system | Completeness per row | All in websites data model |
| | ▪ see (i) mapping problem | | | | |

| No | Business Problem | Business Impact | Data Defect | DQ Dimensions | Attribute |
|---|---|---|---|---|---|
| v | Offering unavailable product | Customer dissatisfaction, unrecognized revenue, ineffective marketing, and potential revenue loss | a. Inaccurate data in e-commerce system (Journal database) | Accuracy inc. currency | Saleable/ Availability in a Region |
| | | | b. Incomplete data in e-commerce system (Journal database) | Completeness, Business Referential Integrity | Fulfillment system |
| | | | c. Inconsistent data from Journal database and e-commerce system | Absence of contradiction, Accuracy incl. currency | Product data |
| | | | d. Inaccurate data in e-commerce system | | Product Data |

- Data defect : (a) (Marketing Restriction)
- Measurement Method :
  1. CPR-01: Ratio of Record with non-blank or non-null for availability fields in Product Repository
  2. SC-02: Ratio of record which has non deviated value in Product Repository (List of Values)
  3. AOC-02 : Ratio of record which adhere business rule

     rule1 :

     Journal

| Site | Currency |
|---|---|
| EST_UK_BS | GBP/ EUR |
| EST_AU_BS | USD |
| EST_ASIA_BS | USD |
| EST_US_BS | USD |
| EST_JP_BS | JPY/ USD |
| EST_MEA_BS | USD |
| EST_EU_BS | EUR |

     Book

| Site | Currency |
|---|---|
| EST_UK_BS | GBP/ EUR |
| EST_AU_BS | AUD/ USD |
| EST_ASIA_BS | USD |
| EST_US_BS | USD |
| EST_JP_BS | JPY/ USD |
| EST_MEA_BS | USD |
| EST_EU_BS | EUR |

  4. TOTAL = 15%xCPR-01 + 15%xSC-02 + 70%xAOC-02
- Frequency : Daily
- Value : [0-1]
- Expected Threshold :

- Data defect [b]
- Measurement Method :
  1. CPR-01 : Ratio of Record with non-blank or non-null field for fulfilment fields in Product Repository
  2. SC-02: Ratio of record which has non deviated value in Product Repository (List of Values)
  3. AOC-02 : Ratio of record which adhere business rule

     rule: Journal

| Site | Print Journal | | eJournal |
|---|---|---|---|
| | PJROMIS | PJARGI | EJSD |
| EST_AU_BS | DELTA | ARGI | CRM |
| EST_EU_BS | DELTA | - | CRM |
| EST_MEA_BS | DELTA | ARGI | CRM |
| EST_UK_BS | DELTA | - | CRM |
| EST_JP_BS | DELTA | - | CRM |

<table>
<tr><th>No</th><th>Business Problem</th><th>Business Impact</th><th>Data Defect</th><th>DQ Dimensions</th><th>Attribute</th></tr>
<tr><td></td><td colspan="5">
<table>
<tr><td>EST_US_BS</td><td>DELTA</td><td>ARGI</td><td>CRM</td></tr>
<tr><td>EST_ASIA_BS</td><td>DELTA</td><td>ARGI</td><td>CRM</td></tr>
</table>
<br>rule: Book
<table>
<tr><th rowspan="2">Site</th><th>Print Book</th><th colspan="2">eBook</th></tr>
<tr><th>Physical</th><th>EBSD</th><th>Others</th></tr>
<tr><td>EST_AU_BS</td><td>BOOKMASTER</td><td>CRM</td><td>DELTA</td></tr>
<tr><td>EST_EU_BS</td><td>DELTA</td><td>CRM</td><td>DELTA</td></tr>
<tr><td>EST_MEA_BS</td><td>DELTA</td><td>CRM</td><td>DELTA</td></tr>
<tr><td>EST_UK_BS</td><td>DELTA</td><td>CRM</td><td>DELTA</td></tr>
<tr><td>EST_JP_BS</td><td>COPS</td><td>CRM</td><td>DELTA</td></tr>
<tr><td>EST_US_BS</td><td>COPS</td><td>CRM</td><td>DELTA</td></tr>
<tr><td>EST_ASIA_BS</td><td>COPS</td><td>CRM</td><td>DELTA</td></tr>
</table>
4. TOTAL : 15%xCPR-01 + 15%xSC-02 + 70%xAOC-02<br>
▪ Frequency : Daily<br>
▪ Value : [0-1]<br>
▪ Expected Threshold :
</td></tr>
<tr><td></td><td colspan="5">
▪ Data defect [c,d]<br>
▪ Measurement Method :<br>
1. ACR-01 : number of unique ISN in Journal database should be available for e-commerce/ number of unique ISN in e-commerce system -> min/ max, number of unique ISBN in Book database should be available for e-commerce / number of unique ISBN in e-commerce system -> min/ max.<br>
2. ACR-03: Ratio of Record with exact same value for ISN in Product Repository with data source (Journal database). Ratio of Record with exact same value for ISBN in Product Repository with data source (Book database)<br>
3. TOTAL : Average of (ACR-01, ACR-03)<br>
▪ Frequency : Daily<br>
▪ Value : [0-1]<br>
▪ Expected Threshold :
</td></tr>
<tr><td>vi</td><td>Products are not included in the marketing campaign</td><td>Potential revenue loss</td><td>Taxonomy mapping problem</td><td>Absence of contradiction</td><td>Subject</td></tr>
<tr><td></td><td colspan="5">see (ii) mapping problem</td></tr>
</table>

**Table 18 Metrics Specification for Preventive and Reactive Measures**

| No | ID | Measurement Method | Value | Freq. | Attribute |
|---|---|---|---|---|---|
| **Completeness per row (horizontal completeness)** [Attribute 2,10] | | | | | |
| 1 | CPR-01 [Attribute 1-2, 4-8] | Sebastian-Coleman (Field completeness - non-null able fields), DAMA, Peralta (Semantic Correctness Ratio Metric)<br>**result** = 1 - (Number of row with empty non-null able field divided with number of all row) [Attribute 4] | [0-1] [Attribute 5-6] | Daily [Attribute 7] | ▪ String and Numeric<br>▪ All in websites data model [Attribute 9] |
| 2 | CPR-02 | Sebastian-Coleman (Parent/child referential integrity)<br>**result** = 1 - (Number of unreferenced row (parentless | [0-1] | Daily | ▪ String and Numeric |

| No | ID | Measurement Method | Value | Freq. | Attribute |
|---|---|---|---|---|---|
| | | row) divided with number of all rows) | | | ▪ All in websites data model |
| 3 | CPR-03 | Sebastian-Coleman (Child/parent referential integrity)<br>**result** = 1 - (Number of row with empty non-null able reference field and non-exist reference field value (childless row) divided with number of all rows) | [0-1] | Daily | ▪ String and Numeric<br>▪ All in websites data model |
| 4 | CPR-04 | result = 70%xCPR-01 + 15%xCPR-02 +15%xCPR-03 | [0-1] | Daily | |
| | NOTE :<br>▪ | | | | |
| **Syntactical correctness (conformity)** | | | | | |
| 5 | SC-01 | Sebastian-Coleman (Validity check, single field, detailed results); Peralta (Syntactic Correctness Ratio Metric)<br>**result** = 1 - (Number of row with non-standard value or format divided with number of all rows)<br>▪ Standard Format: Top-3 string pattern on the basis of distribution OR defined business rule (postcode is 4 char, dash, 2 numeric : ZZZZ-99)<br>▪ Standard Value: Top-3 value on the basis of distribution OR between min-max value of previous data OR defined business rule (price is >=0) | [0-1] | Daily | ▪ Numeric: between min-max value<br>▪ String and Numeric: business rule, string patter, top-3 value |
| 6 | SC-02 | Sebastian-Coleman (Validity check, single field, detailed results:), Peralta (Syntactic Correctness Deviation Metric)<br>**result** = 1 - (Number of row with deviated value divided with number of all rows)<br>▪ Non-deviated value: there is a similar value at reference table with similarity>=0.8 for example (Levenshtein distance/length of longer string) <=0.2 OR Jaro-Winkler distance>=0.8.<br>▪ Similarity=1 for numeric type field | [0-1] | Daily | String and Numeric (deviation=0) |
| 7 | SC-03 | **result** = Average (SC-01, SC-02)<br>if the reference table for SC-02 is not available then SC-03 = SC-01 | [0-1] | Daily | String and Numeric (deviation=0) |
| | NOTE:<br>▪ SC-01 : Non LoV, Incorrect values include: non empty value that could be considered as blank, e.g., Text:"UNKNOWN", Text:"EMPTY", Date: "1/1/1900 00:00:00", Numeric:"0" ?<br>▪ Reference Table (LoV) in PIM for SC-02 : Business Classification, Country, Imprint, Language, Legal Entity, Page Count Type, Product Distribution Type, Product Manifestation Type, Product Type, Publisher, Region, State, Subject Area, Subject Area Type | | | | |
| **Absence of contradictions (consistency) and normative consistency** | | | | | |
| 8 | AOC-01 | Sebastian-Coleman (Consistent column profile)<br>**result** = 1 - (number of non-reasonable fields divided with number of all fields)<br>▪ Reasonable field : field that has the same top-5 values on the basis of its distribution compared with previous data | [0-1] | Daily | String |
| 9 | AOC-02 | Sebastian-Coleman (Consistent dataset content, distinct | [0-1] | Daily | String and |

| No | ID | Measurement Method | Value | Freq. | Attribute |
|---|---|---|---|---|---|
| | | count of represented entity, with ratios to record counts; Consistent cross table multi columns profile:)<br>▪ Rules : Title - Category, Price - Location, Location/ Type - Fulfillment Company Code<br>**result =** average (all ratio per avail) | | | Numeric |
| 10 | AOC-03 | Sebastian-Coleman (Consistent record counts by aggregated date)<br>▪ val1 : (M-1 rows/M-2 rows) ; val2 : (last year M-1 rows/ M-2 rows)<br>▪ val3 = val1/ val2<br>▪ minVal = min(val1,val2,val3); maxVal=max(val1,val2,val3)<br>▪ rawVal = not (minVal or maxVal)<br>▪ **result =** (rawVal-minVal) / (maxVal-minVal)<br>Quarterly : change M with Q | [0-1] | Monthly or Quarterly | String and Numeric |
| 11 | AOC-04 | **result =** Average (SC-02, AOC-01, AOC-02, AOC-3) | [0-1] | Daily | String and Numeric |
| | NOTE:<br>▪ | | | | |
| **Absence of repetitions (free of duplicates)** | | | | | |
| 12 | AOR-01 | **result =** 1 - (Number of duplicate row divided with number of all unique rows)<br>Unique = unique ISBN for book or unique ISSN for journal | [0-1] | Daily | String and Numeric |
| | NOTE:<br>▪ | | | | |
| **Business referential integrity (integrity)** | | | | | |
| 13 | BRI-01 | **result =** Average (CPR-02, CPR-03, AOC-02) | [0-1] | Daily | String and Numeric |
| | NOTE: This measurement could be ignored since it is composed from other measurement's components. | | | | |
| **Accuracy incl. currency** | | | | | |
| 14 | ACR-01 | Peralta (Semantic Correctness Ratio Metric), DAMA.<br>**Result** = average(number of unique book ISBN in Book database + number of unique book ISBN in French site XML/ number of unique book ISBN in e-commerce system, number of unique ISSN in Journal database/ number of unique ISSN in e-commerce system) | [0-1] | Daily | String and Numeric |
| 15 | ACR-02 | DAMA, Sebastian-Coleman (Timely delivery of data for processing, Timely availability of data for access)<br>▪ Ratio 1 : 1- (min difference of time data in Journal database/Book database/French XML with time data in e-commerce system)/ 720<br>▪ Ratio 2 : 1- (min difference of time data in e-commerce system and time data in Web)/ 720<br>▪ **result** = average(Ratio 1, Ratio 2)<br>▪ 720 minutes = 12 hours, if difference>720 then Ratio (i) = 0 | [0-1] | Daily | String and Numeric |
| 16 | ACR-04 | **Result** = average(CPR-01, SC-01, SC-02, ACR-01, ACR-02) | [0-1] | Daily | String and Numeric |

| No | ID | Measurement Method | Value | Freq. | Attribute |
|---|---|---|---|---|---|
| | NOTE:<br>▪ ACR-01 : Book database, Journal database, and French Site XML is considered as the “Real World”. In MDM, the repository holds the golden record. There could be other “Real Word” entities if the architectural type is Transaction Hub. | | | | |

## Appendix 5. Workshop Documents

## References

[1] Sebastian-Coleman, Laura. Measuring Data Quality for Ongoing Improvement. Morgan Kaufmann. 2013