Customer Facing Product Data (CFPD)

Standards & Principles, Data Quality KPI and Audit Protocol

Version 1.2x
Edited for sharing externally
References to internal organisational structures removed
Auguhst 2006

Table of Contents

1 Introduction and objective

This document lays out a series of operational principles and information standards that address the delivery of sustainable improvements in the quality of product information that *Company* shares with it's customers.

These standards and principles will involve the implementation of processes, organisation and practices that extend wider than the attributes included in the scope of this project. However, the increase in demand from customers for improvements in Customer Facing Product Data (CFPD) will act as a catalyst to bring focus and general improvements to the information management programs in *Company*.

Regional Management is responsible for the implementation of these standards and principles.

It is anticipated that regions will incorporate them into existing or planned information management programs.

1.1 Business Driver

An industry wide business need for an overall improvement in the quality of data shared by manufacturers with their customers has been identified. Poor quality product data reduces the efficiency of internal processes and negatively impacts customer relationships, undermining collaborative programs[1].

Although this issue has been highlighted by the recent growth in data synchronisation initiatives between manufacturers and retailers, it is not just linked to data synchronisation programs.

For the majority of product data shared by manufacturers with customers, clear industry standards and implementation guidelines exist. The majority of manufacturers, including *Company*, are aware of these standards and generally believe that they are implementing them correctly. However data quality across the industry is at an unacceptably low level. When data is wrong, significant inefficiency and cost is generated. The value of this waste is considerable. Estimates of the cost involved are between 0.1–0.5% of sales.

[1] Retailer reaction to poor quality data has, in some cases, included refusal to list products with inaccurate or incomplete data.

1.2 The Industry Response

The wide-spread nature of this problem has resulted in considerable focus being placed by the leading industry organisations on addressing this issue.

As part of the industry work, a standardised data inspection protocol and a (self-)certifiable data quality management process has been developed through the Joint Business Planning Group[2] (JBP). The GS1 organisation will manage the JBP developed protocols as the 'GS1 Data Quality protocol'[3].

A significant aspect to the GS1 Data Quality protocol is the focus that it gives to the importance of internal Information Management processes and organisational structures in maintaining high quality data. The approach is similar to that used for ISO[4] certification; a *Company* being 'certified' for 'sustainable data quality' would need to demonstrate that they have the internal capability to manage information adequately and then, through a sample physical inspection, that the information that they have is correct. The combination of the two; good information management practice and accurate sample data, would indicate that this is a trading partner from whom a retailer can trust the product data being received.[5]

The *Company* standards presented herein are aligned to the industry program.

1.3 *Company* Position

Company has played a lead role in the industry program and so will benefit from having a insight into industry activity and thinking. This learning has been incorporated into this program.

The quality of *Company* product data being shared with customers is not particularly better or worse than our peer group.

There are, across *Company*, several programs underway to address the information management of product data shared with customers. Regional (or sub-regional) programs are in progress in North America, Latin America and Europe. Country based programs are, likewise underway in several countries.

It is, however, believed that some common, global *Company* standards and principles are applicable and would ensure that *Company* has a consistent, coherent approach for customer facing product data that would be aligned with industry standards and industry direction.

Hence the justification for these CFPD standards and principles.

2 Scope, Implementation, Auditing, Ownership and Governance

2.1 Scope of Attributes

The standards and principles contained in this document apply to a subset of the total attributes relating to products that are shared (or published) to retail customers.

---

[2] The Joint Business Planning Group includes the following industry organisations: GCI, GS1, AIM, GMA, FMI, ECR Europe, CIES

[3] As of June 2006, the GS1 Data Quality Protocol was being managed with the GDSN division and is planned to be operational in Q3 2006.

[4] 'International Organization for Standards', responsible for standards ranging from ISO9000 to the A-series paper sizes.

[5] This approach is not yet, as of June 2006, fully supported by all parties across the industry.

These attributes are detailed in the table on the following page.

This subset was defined as part of the industry JBP program and reflects those attributes that currently cause the most widespread process failure issues when used within the retailer environment. As an active participant in the industry program *Company* will continue to push for this subset to be expanded to include attributes that have larger material impact on *Company*. This subset should, therefore, not be considered a 'final' list but is a 'first' list. The range of attributes will grow over time and whilst the standards described herein are generally applicable across all attributes there may be additional standards included, or modifications made, as the scope of attributes increases.

The table of attributes also includes reference to the appropriate GS1 global industry standard that addresses that attribute (when applicable). These industry standards are enhanced frequently. Various different language formats and supporting material are also made available frequently through GS1 and other industry organisations. The table, therefore, refers to the primary, or general heading for the standard used by GS1. An Intranet site will be developed by Global Customer Development, associated to this framework, that links to the current versions of industry standards and to appropriate supporting material.

The table also includes a summary of typical data quality issues related to that attribute that have been observed in *Company* or by our peer companies. This should not be read as a definitive analysis of root causes but as illustrative examples.

2.2 Implementation

Responsibility for the implementation of these *Company* Standards and Principles lies with Regional Management. It is anticipated that this will be enacted through the RIO of through a similar regional information function.

Where a region or country chooses to outsource any process that affects compliance with any of the standards and principles, then the region shall ensure control over such processes. Control of such outsourced processes shall be identified within the Regional Data Management Framework (see Section 3).

These standards and principles are applicable to all regions[6]. Regions are required to commit to a time line to reach compliance with these standards and principles.

2.3 Auditing

The standards and principles defined in section 3 of this document will be provided to *Company* Internal Audit for the purpose of auditing regions or countries for compliance, should compliance be deemed an audit priority.

[6] Standards #C1, #C2 and #C3 are not applicable in some circumstances and geographies. This is noted in the definition of the standards in Section 3.

Table 1: Attributes, GS1 Standards, Common Quality Issues

Attribute	*Applicable GS1 Standard*	*Common Quality Issues*
GTIN - Global Trade Item Numbers (formerly EAN and UPC codes)	• GTIN Standards • GTIN Allocation Rules	GTINs are applied to >97% of *Company* consumer items and >95% of trade items[7]. In most of Western Europe and North America they have been in widespread use for decades. Generally the allocation rules are implemented. However, in some cases, a new GTIN is not allocated to changed product which meets the criteria for a new GTIN for a variety of reasons including: - Retailer requests to reduce administration costs - Fear of OOS resulting from failure by the retailer to align ordering systems - De-listing of slow moving products after a changed GTIN brings the item to the attention of the buyer - Retailer slotting or listing fees (although this appears to be generally overstated) An industry project, sponsored by *Company* for the GCI, is being started (May '06) to address these common industry issues with the correct allocation of GTINs.
Classification Category Code	• GPC Standards ('Brick' level only[8])	The GPC standard is not generally used in *Company* and so GPC brick codes are not applied at source. These codes are generally only applied to product information that is being published through the Global Data Synchronisation Network (GDSN) – including e-Cat[9] as they are mandatory for registering all products in the GDSN. Due to the codes being added manually, or through data mapping programs, at the point of publication, it is highly likely that some will be applied incorrectly. At this stage it is not easy to judge how many as the classification data is yet to be widely used by retailers.
Trade Item Description	• Industry development of Trade Item Description standard is in progress.	Trade Item Description is often manually generated in local language, according to local rules. With regional and global product sourcing, this can lead to misalignment of descriptions and language of the same product in different systems or no local language support at all in multi-language environments.
- **Net Content** - **Depth** - **Width** - **Height** - **Gross Weight** Including respective **unit of measure**	• Package Measurement Rules • Agreed tolerances	- Transposing of attributes (eg: depth and width) - Incorrect scale (eg: 100cm instead of 10cm) - Information not updated after changes are made to the physical packaging - Information copy/pasted from other GTINs and not corrected
- Quantity of sales items **- Quantity of next lower level trade item** **- Quantity of trade items per pallet layer** **- Quantity of layers per pallet**	No applicable GS1 standard	- Information not available at the time that the record was created and not added later - Pallet configurations changed and information not updated

[7] Measured through the Customer e-Business scorecard. Items not allocated GTINs tend to be in countries (mostly in Asia) were GTINs are not used. It is, therefore, not a *Company* problem in not having 100% coverage.
[8] The full GS1 Global Product Classification (GPC) Standards include *bricks*, *attributes* and *values*.
[9] e-Cat is the name for the *Company* technology infrastructure that supports GDS

2.4 Ownership and Governance of *Company* CFPD Standards and Principles

The quality of CFPD that is shared with customers is ultimately the responsibility of the Customer Development function although it is understood that the creation and management of that information falls across many functions with CD often playing a minor operational role.

(a) Review of these Standards and Principles

The CFPD standards and principles are valid from the date of publication and will be reviewed and reconfirmed, modified or extended one year after publication.

Interim changes to the *Company* standards will be kept to a minimum during the initial twelve months but it should be recognised that revisions and improvements to the industry standards that underpin these *Company* standards will occur and those changes will need to be incorporated into the *Company* standards.

(b) Exception Management Process

Section 3 of this document lays out in detail the standards and principles for CFPD. These have been grouped into three logical blocks:

Group A: Regional Data Management Framework
Group B: Attribute Ownership and Management Standards
Group C: Implementation of Industry Defined Standards

Group A and B are, primarily, principles, based on industry and *Company* best practice, by which regional information management and operational information management will be implemented. Therefore **no exceptions** will be permitted for Group A or B.

Exceptions to the standards included in **Group C** will, generally, only be allowed if complying with the standard would expose *Company* to risk including, for example, financial loss, deterioration of customer relationship or legal liability.

The regional body responsible for implementing the CFPD standards and principles will also be responsible for ensuring that operations within the region comply with Group C standards. Work currently underway in *Company* and in the industry[10] indicates that in many cases non-compliance issues with the industry standards referred to in Group C can be resolved with increased education.

If the regional body is unable to resolve the issue and maintain compliance with the standards then they will be required to submit a request for an exception. This request should fully document the reason for the request and should identify the risk to *Company* from compliance.

The request will be reviewed and, if deemed valid, an exception will be granted. Exceptions will be formulated as an 'exception rule' that will be added to the standard allowing regional resolution of similar exception requests.

[10] For example, the GCI GTIN Compliance project which includes *Company* participation. As of June 2006.

3 *Company* CFPD Standards and Principles

For clarity the *Company* standards and principles have been grouped into three logical blocks.

A. Regional Data Management Framework

These standards and principles address the structures (organisational, process and policy) that regions will be required to put in place to ensure good information management of CFPD.

The standards and principles in this section are:

#A1 A regional CFPD policy will be established
#A2 Operational processes and procedures that impact CFPD will be identified and documented
#A3 A continuous improvement program for CPFD will be established
#A4 The CFPD program and CFPD procedures will be reviewed regularly against the GS1 Data Quality protocols
#A5 Audit trails will be established for the creation or change to any CFPD attributes

It is recognised that these standards and principles are applicable to most types of data, not just CFPD. However, due to the particular and current need for improvement in CFPD quality, the regions will be required to demonstrate that these standards and principles are in place for the CFPD attributes (as listed in Section 2.1) as a minimum.

B. Attribute Ownership and Management Standards

The standards and principles included in this group are applicable at a local/regional level and concern the detailed management of attributes. These standards and principles, therefore, would be included in the Regional Data Management Framework for CFPD. 'Local' will, most likely, mean a country BU level or country category level but the decision as to what level is appropriate would be made by the region.

The standards and principles in this section are:

#B1 Data owners must be identified for each CFPD attribute
#B2 One data source is identified for each CFPD attribute (per item)
#B3 CFPD attributes are validated at point of data origin
#B4 Data owners periodically re-confirm the correctness of the data
#B5 Sample data inspection of physical product measurements is carried out periodically

C. Implementation of Industry Defined Standards

This group represents the external, industry standards that are mandated for use within *Company* and, thus, are considered as *Company* standards.

The standards (in *Company* terms) in this section are:

#C1 GTIN codes will be allocated and/or changed according to appropriate GS1 standards
#C2 A GS1 Global Product Classification ‘brick’ code will be assigned to all traded items
#C3 Physical measurement of product for data origination, verification or audit, shall be undertaken in accordance with the GS1 Package Measurement rules
#C4 Information Management processes and procedures that impact CFPD will conform to the GS1 Data Quality protocols
#C5 Trade Item Description

Group A: Regional Data Management Framework

Standard #A1	**A regional CFPD policy will be established**
Business reason	The majority of data quality issues relate to how CFPD is managed in the organisation. Establishing a clear regional policy for managing CFPD brings focus and a structured plan to resolving this issue.
Description	The CFPD policy communicates the commitment to professionally manage customer product data in the region and lays out the individual elements of the CFPD program. Management shall provide evidence of its commitment to the development and implementation of a CFPD management framework by: - Communicating to the organisation the importance of meeting this requirement - Communicating to the organisation the data quality objectives on a yearly basis - Communicating quarterly the status of data quality measurement against the target This commitment is laid out in the CFPD policy. The policy will be published and be generally available across the region. It will form the basis of local programs. Top management's commitment is required to the CFPD framework, its objectives and rigorous implementation. This standard is analogous to the ISO 9000 standard for product quality

Standard #A2	**Operational processes and procedures that impact CFPD will be identified and documented**
Business reason	Clarity on which processes and procedures impact CFPD is needed in order to ensure that these processes and procedures support, and are aligned to, the data quality program
Description	The internal CFPD processes shall be described including the data creation and maintenance process of product attributes. This can be done as a part of the description of the total data creation and maintenance process and procedures. Management of CFPD runs across multiple functional departments. A clear organisational structure with defined owners, responsibilities and procedures is critical to the medium to longer-term vision for consistent high quality data to flow through the global supply chain. It guarantees a certain quality level of the product data.

Standard #A3	**A continuous improvement program for CPFD will be established**
Business reason	To avoid deterioration of data quality business focus needs to be given to continuous review and improvement.
Description	The program is analogous to the TPM programs run in SC. Main elements of the Continuous Improvement Program are: - Regular reporting of the standardised Data Quality KPI (see Section 4) - Develop a loss tree which shows impact and detailed root causes of poor CFPD quality - Prioritise the root causes and address the largest opportunity with a focused improvement project with small team - Roadmap progress towards the required result - Communicate success throughout the organisation - Make this a continuous improvement cycle by repeating above steps The relevant part of the business is continuously trained in developing appropriate skills and competencies to recognise the effect of bad data for the complete organisation and to be able to improve data management processes and procedures.

Standard #A4	**The CFPD Program and CFPD Procedures will be reviewed regularly against the industry Data Quality protocols**
Business reason	Regular review of the CFPD programs and procedures will identify gaps that need to be addressed and will identify aspects of the program that need realignment.
Description	Internal reviews will be carried out at planned intervals of, initially, not less than once per year to determine whether the data management system conforms to the planned arrangements, the requirements of this standard and the data quality management system requirements established by the organisation, and is effectively implemented and maintained. Review programs shall be planned, established, implemented and maintained taking into consideration the importance of the data management system processes and the results of previous reviews. The organisation shall establish and maintain documented review procedure that addresses: ▪ Responsibilities and requirements for planning and conducting reviews, reporting results and retaining associated records, ▪ Determination of review criteria, scope, frequency and methods. The data quality management system documentation shall include: ▪ Documented statements of a data quality management policy and data quality management objectives ▪ A data quality management manual ▪ Documented procedures Records needed by the organisation to ensure the effective planning, operation and control of its data quality management processes.
External ref:	GS1 Data Quality protocols

Standard #A5	**Audit trails will be established for the creation or change to any CFPD attributes**
Business reason	Audit trails provide a time based record of all changes made to attribute content. This enables identification of what information was available to customers at any point in time and, therefore, supports issues resolution.
Description	Records must be kept in place within the regions and OpCo's of the changes made to the attributes in scope. It should include the named person, the system in which the change was made and the date when the change was made. It is usual for audit trail will be automatically available through core systems like SAP.

Group B: Attribute Ownership and Management Standards

Standard #B1	**Data owners must be identified for each CFPD attribute**
Business reason	Data ownership provides a single point of responsibility for data entry, maintenance and accuracy of product attributes. This also supports the continuous improvement process by allowing feed-back on data quality measurement to the relevant data owners.
Description	For each product attribute a data owner is identified. Data ownership is recorded in a regional database and the list of owners is kept up to date and is freely accessible in the regional organisation. Data owners must be named individuals with their functional roles defined. The data owner is responsible for ensuring that all data creators understand how standards are to be implemented. Data owners are responsible for execution of the standards as described in standard #A4. Data owners are responsible to take preventative action to avoid potential data inaccuracies. These actions include: - Determine data inaccuracies and their causes - Evaluate the need for action to prevent occurrence of data inaccuracies - Determine and implement necessary actions - Keep records of results of action taken - Review preventive action taken. Data owners are also responsible to take corrective action to eliminate the cause of data inaccuracies. These actions include: - Review data inaccuracies (including user feedback) - Determine the causes of data inaccuracies - Evaluate the need for action to ensure that data inaccuracies do not recur - Determine and implement action needed - Correct data in the product master data - Record the result of action taken and - Review corrective action taken.

Standard #B2	**One identified data source for each CFPD attribute (per item)**
Business reason	Avoidance of data conflicts in downstream systems and processes generating unnecessary data queries. CFPD management is more effective and efficient when the data is created and maintained in one system. Difficulty arises when there are multiple systems used for creating the same data.
Description	Only one master data source is recognised as driving systems to secure the “best” source of data for the *Company* and to avoid data reconciliation between two or more sources. Where separate data management systems exist, the organisation shall ensure that the information provided by these systems is consistent and that one system be identified as the unique data source.

Standard #B3	**CFPD attributes are validated at point of data origin**
Business reason	Minimises corrective measures further down the data chain where the cost (and complexity) of correction potentially increases.
Description	The data management framework should put maximum validation at the data input side. This is done by checking the input against a set of up-to-date validation rules. If product data changes are required a message is sent back to the data owners for immediate correction. Data owners, in co-ordination with regional information offices, determine the rules to be applied. Tolerances should be set and reported to management. Validation rules are well documented, reviewed quarterly by management for updates (new and changed validation rules). Tools are developed to create alerts when wrong data is being inputted. Adoption of these tools to be included in all regions implementation plans. Reports like numbers of alerts per attribute and the total number of required changes will be developed and are used across the business. Reviews of these reports, and any additions required, is a continuous improvement process (see standard #A3).

Standard #B4	**Data owners periodically re-confirm the correctness of the data**
Business reason	Positive reconfirmation of data accuracy drives review and correction.
Full Definition	For each product attribute a data owner must positively confirm the accuracy of data at least once every six months.

Standard #B5	**Sample data inspection of physical product measurements is carried out periodically (see also #C4 and KPI definition)**
Business reason	Next to regular Data Quality KPI reporting, regular data inspection is used as a means to enhance data accuracy and avoid measuring queries.
Full Definition	The region shall establish and maintain a procedure for physical product measurement and data generation in accordance with GS1 standards and recommendations. The GS1 standards and recommendations include: ▪ Statistically relevant sample size ▪ Methods for measuring product attributes ▪ Measuring equipment ▪ Measuring location and conditions ▪ Personnel to perform the measurements ▪ Method for the recording of measurement data ▪ Measurement output format The region may choose to perform inspection internally or may choose a certified third party to perform the inspection in accordance with the GS1 Data Quality protocol. It should be noted that *Company* does not support the use of third party 'measurement services' mandated by customers or by national GS1 organisations.[11]
External reference	GS1 Data Quality protocols GS1 Package Measurement Rules

[11] The *Company* policy for this, issued as part of the Customer eBusiness Program, can also be found on Intranet site supporting these standards and principles.

Group C: Implementation of Industry defined standards

Standard #C1	**GTIN codes will be allocated and/or changed according to the appropriate GS1 standards**
Business reason	GTINs, also know as EAN and UPC codes, are the foundation for many business critical applications and processes ranging from warehouse management, the order-to-cash cycle and through to the point of sale.
Description of GTINs	Global Trade Item Numbers (GTINs) are a global industry standard for product identification that incorporates the existing UPC and EAN product identification standards[12]. GTINs are used by manufacturers and retailers throughout North America, Europe and Australasia. GTINs are used in most Latin American markets and are growing is usage in Asia and Africa. GTINs are managed by the local GS1 organisations. Barcodes, used on consumer and trade items, are carriers of the GTIN number. In markets where GTINs have been adopted, the use of GTINs is a mandatory standard for all *Company* consumer and trade items. *Company* is an active supporter of GS1 standards and encourages the use of GTINs in those markets where they are not currently adopted.
Description of Standard	Each region must ensure that GTINs are issued and maintained in compliance with the industry standards, as described in the GS1 published 'GTIN Allocation Rules'. The GTIN is the primary internal identifier of a Consumer or Traded Unit and must be treated and stored in *Company* systems as a key reference entity and not merely as a physical attribute. Data structures in applications and databases must be able to store 14 digit, globally unique GTINs.
Applicability	This standards is applicable in all countries that have any use of GTINs.
External reference	GS1 GTIN Allocation Rules

Standard #C2	**A Global Product Classification (GPC) 'brick' code will be applied to all traded items**
Business reason	The industry Global Data Synchronisation (GDS) process requires the use of the GPC 'brick' code. This is to enable buyers to narrow GDS searching to categories appropriate to their requirements.
Description of GPC	The Global Product Classification (GPC) standard provides a common language to allow the definition of categories and development of product

[12] The GTIN is a 14 digit non-negative integer. Existing EAN and UPC numbers are converted to GTINs by adding leading zeros to make 14 digits.

	hierarchy. The foundation of GPC is the *brick,* which defines groups of similar products. An example would be 'still wine'. A *Brick* can be further characterized by *attributes* specific to the *brick*, such as 'colour'. The GPC also has maintained list of acceptable *values* for *brick attributes*, such as 'red', 'white', 'rosé'…
Description of standard	All traded items must be classified with the appropriate GPC *brick* code in accordance with the industry standards, as described in the GS1 published GPC Access Guide. The use of the 'Unclassified' or 'null' GPC *brick* code is not permitted. Data structures in applications and databases must be able to store 8 digit GPC *brick* code as an attribute for each consumer and trade unit. The implementation of the GPC *attributes* and *values* is not currently mandatory in *Company*.
Applicability	This standard will apply to all items that are being published to any external data synchronisation process including, but not limited to, the following: ▪ *Company* e-Cat solution, ▪ Any national/regional GS1 data pool or ▪ Any retailer owned data pool. This is applicable even if the national/regional GS1 data pool or retailer owned data pool does not mandate the use of the GPC brick code.
External reference	GS1 Global Product Classification Standard – GPC Basics GS1 Global Product Classification Standard – GPC Access Guide

Standard #C3	**Physical measurement of product for data origin, verification or audit, shall be undertaken in accordance with the 'GS1 Package Measurement rules for data alignment'**
Business reason	To ensure the quality of data on the physical aspects of a product, such as dimension and weight, that is communicated to our customers Additional benefits can be realised internally through improved quality of dimension and weight data, including in improved efficiency in inventory management and transport utilisation.
Description	The use of the 'GS1 Package Measurement rules' is mandatory for all consumer and trade items. The rules define how product at unit, case and pallet levels, should be measured and weighed.
External reference	GS1 General Specification Package Measurement Rules (for UK) Package Measurement Rules for Data Alignment

Standard #C4	**Information Management processes and procedures that impact CFPD will conform to the GS1 Data Quality protocols**
Business reason	A certifiable quality standard based on process will support sustainable improvements in product data quality rather than by the use of measurement services (which add significant cost and complexity to the manufacturer supply chain). In addition, an industry agreed protocol brings consistency to retailer requirements.
Descriptionn	The *Company* standards #A1 to #A4 described herein should all be implemented in conformance with the GS1 Data Quality protocols. *Company* will, regionally, proceed with the published certification process[13] to confirm conformance.
External ref.	GS1 Data Quality Protocols

[13] As of June 2006, industry work is underway, expected to complete before the end of the year, on allowing a self-certification option.

Standard #C5	**Trade Item Description**
Business reason	Trade Item Description is an established, consistent and industry recognised approach for describing products. It is also one of the four mandatory attributes for uniquely identifying a set of data associated with a GTIN as part of the Global Data Synchronisation process. Internally, harmonisation of Trade Item Description across countries in a region reduces risk of duplicate records and having similar product with different descriptions.
Description	The Trade Item Description is generated from four attributes: ▪ Brand (mandatory) ▪ Sub Brand (optional) ▪ Functional Name (mandatory) ▪ Variant (optional) It replaces the Item Description Long and Medium found in the previous model. An industry review of this standard is currently underway.
Applicability	Application of this standard is, currently, **not mandatory.** Regions are, however, advised to start incorporating compliance with this standard as it currently exists in order to be prepared for when it does become mandatory. This is of particular importance in countries engaging in any external data synchronisation activities including, but not limited to: ▪ *Company* e-Cat solution, ▪ Any national/regional GS1 data pool or ▪ Any retailer owned data pool.
External ref.	To be advised

4 KPI for Data Quality

4.1 Definition of Data Quality KPI

The Data Quality KPI is a very simple measure that will allow *Company* to track the progress of data quality improvements over time.

Data Quality will be calculated at a data record level (where a data record contains product information about a single GTIN) and for all the attributes included in the scope of this program and detailed in the table in Section 2. This gives equal importance to all in-scope attributes in the data record.

Data Quality also gives equal weight to the **completeness** of a record (all attributes contain values) and the **correctness** of a record (the values in each attribute are correct).

The KPI will be calculated as follows and reported as a percentage:

$$\textit{Data Quality} = \frac{\textit{The number of product records that contain no missing or incorrect data}}{\textit{the total number of product records being measured}}$$

So, for example, a Data Quality KPI of 70% measured on 200 product data records would indicate that 60 product data records contained at least either one missing attribute value or one wrong data value.

No weighting is given to the importance of one product or another (such as by sales value) so Data Quality can be simply aggregated cumulatively, by data record, up to country, regional or global level.

4.2 Applicability

Regions will be required to report the Data Quality KPI at country and region level for all products that are commercially available in countries that are engaged in any external data synchronisation activities including, but not limited to:

- *Company* e-Cat solution,
- Any national/regional GS1 data pool
- Any retailer owned data pool

If a country is engaged in data synchronisation activities but only for a subset of products then it will still be required to report Data Quality for all products commercially available in that country.

4.3 Automated Measurement of Data Quality and Sample Inspection

Routines can be implemented into regional and local systems that can automatically measure the completeness of product data records and can measure the correctness, within defined parameters, of the product data record attribute values. It is assumed that regions will implement such routines within their core systems to measure Data Quality.

The defined parameters, by which the correctness of the data would be measured, would be based on rules for syntax and logic. This might include, for example, that a value for the height of a case of a product would be:

- a positive integer
- in a range between 5cm and 100cm
- equal or higher than the height of the equivalent consumer unit
- less than the height of the equivalent pallet.

Although the resulting data can used as a guide measure for Data Quality (and can be used within the region or country to measure improvement) it can not be used to give a true measure of Data Quality. This is because an attribute value may pass the validation test but is still incorrect. For example, and using the example above, the value in a data record for the height of the case might be listed as 12cm which meets the validation rules but, on inspection, the actual height of the case is seen to be 15cm.

To calculate the true Data Quality a full physical audit of all products followed by a comparison with the data product records would need to be undertaken.

Realising the impracticability of this, regions will be required to submit Data Quality results that can be based on automated measurement accompanied by the results of sample physical inspection as described in Standard #B5.

The results that will be submitted by each region will, therefore, be as follows, per country:

- the total number of records being measured
- the number of product records that contain no missing or incorrect data (as measured automatically)

- the number of consumer and trade units that were inspected
- the number of consumer and trade units that, after inspection, shown to have incorrect or incomplete

4.4 Collecting and Reporting the KPI

Regions are required to implement the Data Quality KPI for all countries and for all consumer and trade units in their region.

4.5 Additional Data Quality Measures

Regions should consider the use of other optional Data Quality measures to support the building of understanding of common Data Quality issues and more detailed monitoring of improvements.