**Insights using MySQL to explore Global Trade Item Number (GTIN) data structures**

Randy Lisbona, Marvin Scott, Vinh Le

**Abstract**

GS1 is a global organization that maintains standards for numbering systems used between trading partners. GTIN (Global Trade Item Number) is a family of product ID codes used worldwide between trading partners to identify products via linear or 2D barcodes on the product or packaging. Commonly referred to as UPC codes in the USA. This project will provide insight on the format and usage of GTIN product codes.

An open source GTIN subset > 100K records will be downloaded and installed on a local mySQL database instance. Open source documentation will be reviewed to better understand the GTIN data. A graphical schema of the database will be created with example data from each table to visualize the relationship between tables. Exploratory queries will be created to summarize selected fields such as brand, and packaging level. A summary list of mySQL commands used in the analysis will be provided along with impressions on ease of use, intuitiveness, and effectiveness.

Figure 1 - GTIN formats



# GS1.org History

GS1.org is the global standards organization that manages a system of unique identification numbers used between trading partners in over 150 countries worldwide. GS1 traces its roots to the Uniform Product Code Council (UCC) which was established in the USA in 1973 to manage UPC barcodes used in North America, and to the European Article Numbering Association (EAN International) established in 1977 to develop a compatible barcode identification system outside of North America. GS1 was launched in 2005 to combine the two standards organizations into one international organization. [1]

# GTIN formats

GTIN (Global Trade Item Number) is just one of the numbering systems managed by GS1:

* **Global Trade Item Number (GTIN).**
* Global Location Number (GLN).
* Serial Shipping Container Code (SSCC).
* Global Returnable Asset Identifier (GRAI).
* Global Individual Asset Identifier (GIAI).
* Global Service Relation Number (GSRN).
* Global Document Type Identifier (GDTI).
* Global Shipment Identification Number (GSIN).
* Global Item Number for Consignment (GINC).
* Global Coupon Number (GCN).
* Component / Part Identifier (CPID).

A GTIN consists of four parts, an optional Application Identifier, the Company Prefix, the Item Reference, and a check digit. Notice that price is not in the GTIN, the GTIN is used to lookup product price in an external database. See Figure 1.

* GTIN-8 is a truncated 8 digit GSI identification key used on packages with limited label space, such as chewing gum.
* GTIN-12 is 12 digit GSI identification key consisting of a U.P.C. company prefix, item reference, and check digit.
* GTIN-13 is a 13 digit GSI identification key consisting of a GS1 Company prefix, item reference, and check digit
* GTIN-14 is a 14 digit GSI identification key consisting of an indicator digit (1-9), GS1 company prefix, item key, and check digit.

## Indicator Prefix



Table 2.1 - Figure 3 Decoding

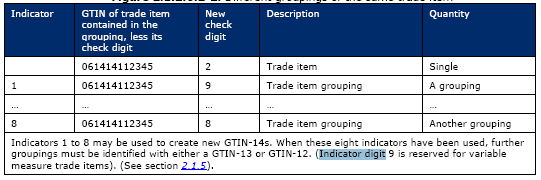


Figure 2 - Indicators

The Indicator Prefix Digit is used to define a grouping, or packaging level. For instance, in Figure 2, we see an example GTIN of 061414112345, let’s say this was a 2oz bag of candy corn. Adding an indicator prefix of “1” could denote a bulk pack of 12 2oz packages, indicator prefix “2” could be used for a 24 pack, indicator prefix “3” could be a “Gross” pack (144) of 2oz candy corn packages, and so on. If more than 8 packaging levels are needed, then a new GTIN-12 or GTIN-13 is required.

## Variable Measure Trade Items

Indicator prefix “9” is reserved for variable measure trade items. Variable measure items are trade items that cannot guarantee consistent weight, size, or length due to the production process (e.g. meat, bulk cheese).

## Data Carrier

A data carrier is a means of representing GS1 identification codes in a machine readable form. 1D Linear barcodes such as UPC-A or EAN-13 are widely used. Matrix symbols such as QR code and Data Matrix are also supported, as well as RFID tags.

## Examples

Figure 3 shows an example barcode from a box of chocolates. The format is GTIN-13 and the data carrier is an EAN-13 Linear barcode.



Figure 3 - EAN-13 barcode symbol

Table 2.1 shows decoded values from figure 3. The first three digits (400) indicate the country that issued the Global Company Prefix. In this case it was Germany. The country that issued the GCP is not necessarily the home country of the company. For instance, a US based company may get a GCP issued in a foreign country. The Prefix plus the company code 81555 results in a unique company code worldwide (40081555).

The item code is 02020, Companies assign their own item codes. The calculated check digit is 7. Check digit calculations are outlined in section 7.9 of the GS1\_General\_Specifications document. [1]

Figure 4 is a chocolate candy bar from the same company, note the item code of 03300 and check digit 9.

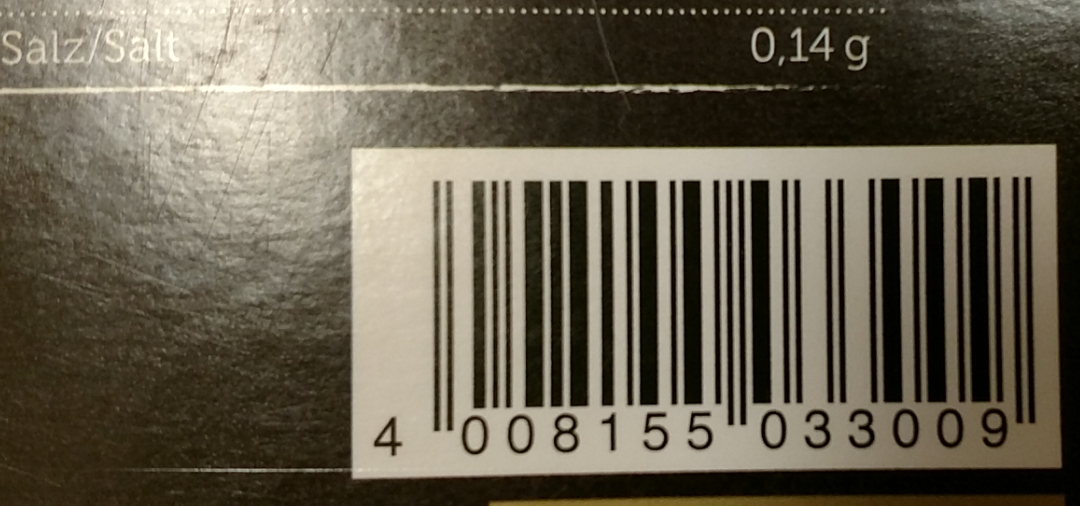


Figure 4- EAN-13 barcode symbol 2

Figure 5 shows a GTIN-12 encoded in a UPC-A linear barcode. 8 indicates the GPC was assigned in the USA, 84912 is the GPC for Post Consumer Brands, LLC. The Item number is 00471 and it was from a 581-gram box of Grape Nuts Cereal.



Figure 5 - UPC-A barcode symbol

# Product Open Data

GS1 does not maintain a worldwide list of Global Trade Item numbers (GTIN), GS1 simply assigns Global Company Code prefixes to ensure unique GTIN values. Each company using GTIN maintains their own list of item numbers and shares this with their trading partners. There are several initiatives to create a global open source GTIN list. For this project the Product Open Data (POD) database was used. This is far from a complete database in its current state, but does demonstrate the type of data linked to GTIN’s.

## Download and install POD

The POD database was downloaded and installed locally in MySQL. [2] The Specification file was also downloaded. [3] Queries were written to pull sample records from each table.

## POD Observations

Thirteen tables containing 2.6 million rows were found in the POD database. Example records were extracted from each table and key fields were examined to generate a schema. See Appendix B, Figure B1. Discrepancies from the schema in the specification file found with the POD database were resolved using key field names and matching records in the POD database. While there were at least a few records in each table, matching records in related tables proved to be disappointing. See Appendix C, figure C1 for a summary of matching records in the 7 largest tables. Of the 2.6 million original rows, only 52 were linked between 6 tables, 0 for 7 or more tables. The website seems to imply POD is an attempt at a global GTIN registry but it appears it may have been abandoned in early 2014, the last revision date on the database dump.

Figure 6 - Record counts and key fields in POD database



## Metadata

Marvin, please add paragraph here



Figure 7 - Object Metadata

And other charts. Vinh also add your charts and a write up describing the chart.

# Summary

Without Global Company Prefixes and Global Trade item numbers, there would be a high chance of a single item number being used by more than one company. It is hard to imagine how global trade would work without Global Trade Item Numbers and the GS1 organization in place to ensure unique numbering worldwide. The POD database gave an interesting peak into the detailed data related to GTIN, but it’s small size and fractured nature make it not very useful other than learning about GTIN.

# References

|  |  |
| --- | --- |
| [1] | GS1.org, "The global Language of Business," 11 2016. [Online]. Available: http://www.gs1.org/gs1-source/latest GS1\_General\_Specifications.pdf. |
| [2] | "Product Open Data," 01 01 2014. [Online]. Available: http://www.product-open-data.com/download/ "POD Database - Dump". [Accessed 11 2016]. |
| [3] | "Product Open Data," 13 11 2013. [Online]. Available: http://www.product-open-data.com/download/ "POD Database - Specifications". [Accessed 11 2016]. |

**Appendix A Data sources and related works**

Github Repository : <https://github.com/rlisbona/MSDS-7330-Term-Paper-1>

Github contains files used in this project.

Table A1 - Data Sources

|  |  |
| --- | --- |
| Product Open Data – Subset of GTIN | <http://www.product-open-data.com/en/1-home.html> |
| POD database SQL Create and Load | <http://www.product-open-data.com/docs/pod_web_2014.01.01_01.sql.gz> |
| POD database Specification | <http://www.product-open-data.com/docs/POD-SPECS-2013.11.13_01.xlsx> |

Table A1 shows links to data sources downloaded for this project

Table A2 -- Related Works

|  |  |
| --- | --- |
| Open EAN/GTIN Database | http://opengtindb.org/ |
| Outpan | https://www.outpan.com/ |
| Datakick | https://www.datakick.org/ |
| EAN-Search | http://www.ean-search.org/ |

Table A2 shows websites similar to Product Open Data. Most of these limit searches to just a few GTIN records at a time and do not have a database to download.

**Appendix B POD database Schema**

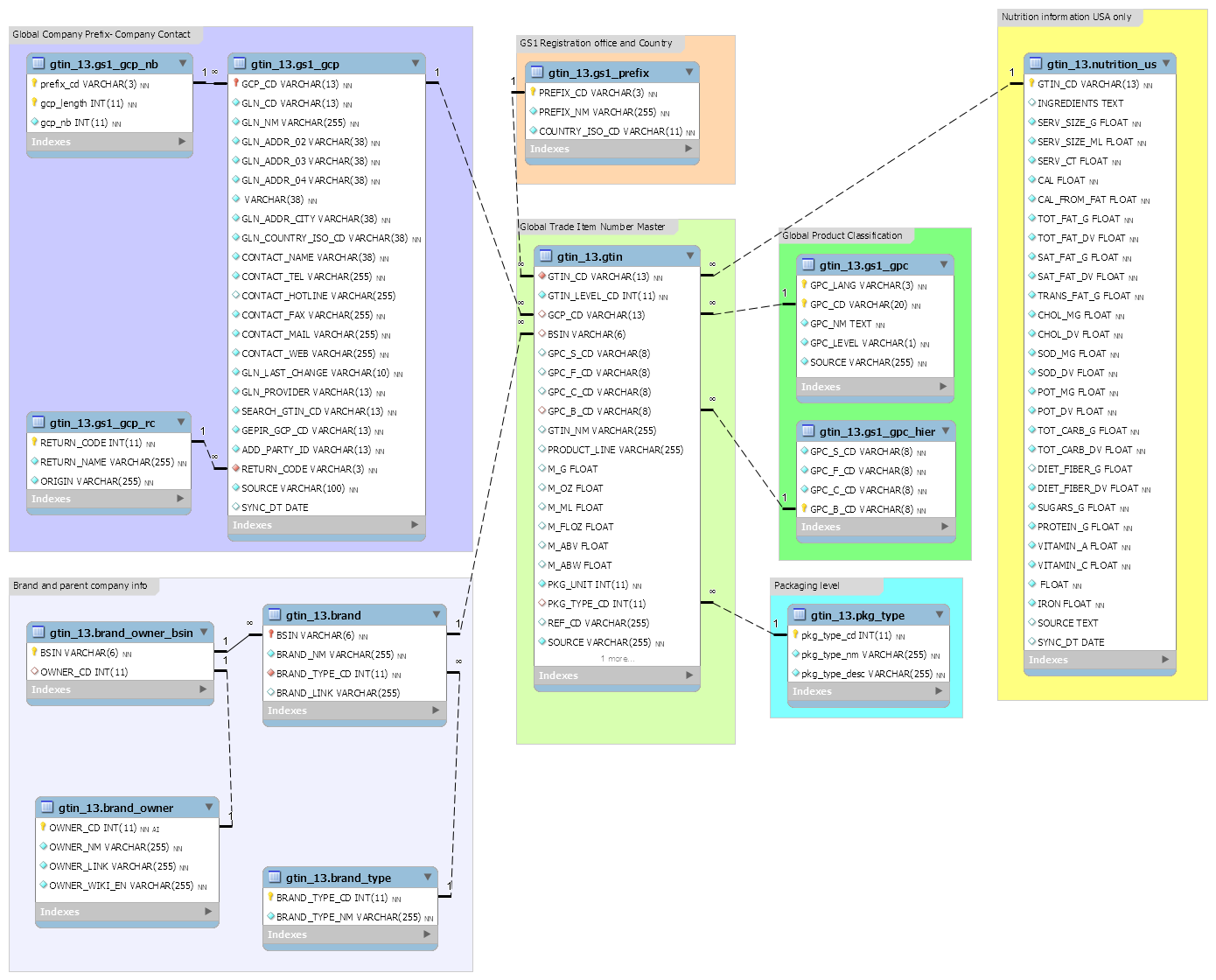


Figure B1 - POD database Schema

**Appendix C – MySql Exploratory code**

Figure C1 – Exploratory Joins

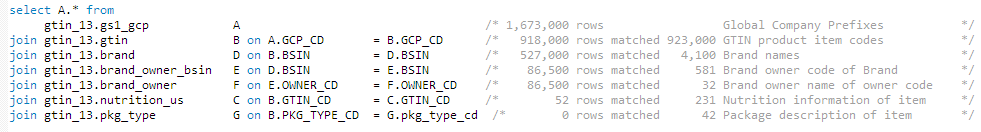


Figure C1 shows exploratory join record counts. As more tables were added to the join, fewer and fewer rows matched between all tables. The conclusion from this query is the example data in the POD GTIN database is not a complete subset of GTIN.