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

Randy Lisbona, Marvin Scott, Vinh Le

**Abstract**

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. This project will provide insight on the format and usage of GTIN product codes using a subset of GTIN.

An open source GTIN subset > 100K records (.csv, .sql, or JSON format) will be downloaded and installed on a local mySQL database instance. Open source documentation will be reviewed to better understand the GTIN data. The schema will be checked for normalization and modified if necessary, Primary and foreign key indexes will be created/modified as needed. 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.

**History**

GS1.org is the global standards organization that manages a system of unique identification numbers used between trading partners 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 format**

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. See Figure 1.



Figure - GTIN formats

* GTIN-8 is a truncated 8 digit GSI identification key used on packages with limited label space
* 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**

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 groupings are needed, then a new GTIN-12 or GTIN-13 is required for the item.

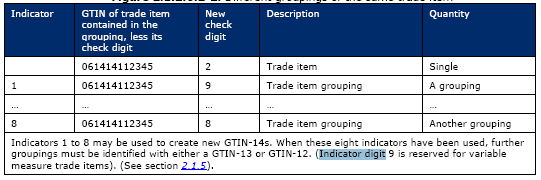


Figure - Indicators

**Variable Measure Trade Items**

Indicator “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).

the GTIN dataset, (open source, 1M + records, a variety of tables. Where we got our tables, steps to load it. Steps to import BSIN code in JSON <http://brand-okfn.herokuapp.com/brand/>

Investigate feasibility of doing this project on webhosted mySQL like <https://www.000webhost.com/>

Background on GTIN from [www.gs1.org](http://www.gs1.org)

, who uses it, can we find/estimate worldwide usage. Describe various formats for GTIN =(companycode+Itemcode) How many unique companies and itemID’s can it encode. We have GTIN-13, there are a few others. Appears that there is not an open source version of the complete database, trading partners publish their data to companies selling the products? Can we explain why? Explain why price is not in this database.

EER diagram– Randy working on this. Steps in MySQL to create EER, Reverse engineering the schema in MySQL didn’t bring in the tables. Data Export to SQL script, import mysql reverse engineer script does work, steps to explore and define relationships. Need explanation of nearly empty skeleton tables found with just a few records. Base our queries on the

EER Graphic

Queries: Basic stats on each table, records, keys, do we need to add indexes? File size of zipped, unzipped, and MySQL database. Create several

example summary queries, a few charts:

1. ItemCount by Company
2. Brand Count by company
3. Average Package size ?

Use Excel PowerQuery for charts, need to install SQL connector <https://support.office.com/en-gb/article/Connect-to-a-MySQL-database-Power-Query-8760c647-88b9-409d-b312-6ea8f84a269b?ui=en-US&rs=en-GB&ad=GB>

Insights:

1. Average package size
2. Min max avg number of items per company.
3. Brands per company
4. Ease of using MySQL for relationships, queries, data exploration

Summary, what we learned with each of us working on our own local copy, vs a web version if we can get that working.

Max length of paper = 4 pages

# References

|  |  |
| --- | --- |
| [1] | GS1.org, "The global Language of Business," 11 2016. [Online]. Available: http://www.gs1.org/gs1-source/latest GS1\_General\_Specifications.pdf. |

Appendix

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

Figure 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> |
| BSIN – Brand to Product (JSON) | <http://brand-okfn.herokuapp.com/brand/api/v1/brand/?format=json> |
| GS1.org – Standards Organization for GTIN | <http://www.gs1.org/> |
|  |  |

Figure - Related Works

|  |  |
| --- | --- |
| None found so far |  |
|  |  |



Figure 5 -POD Database Record Counts

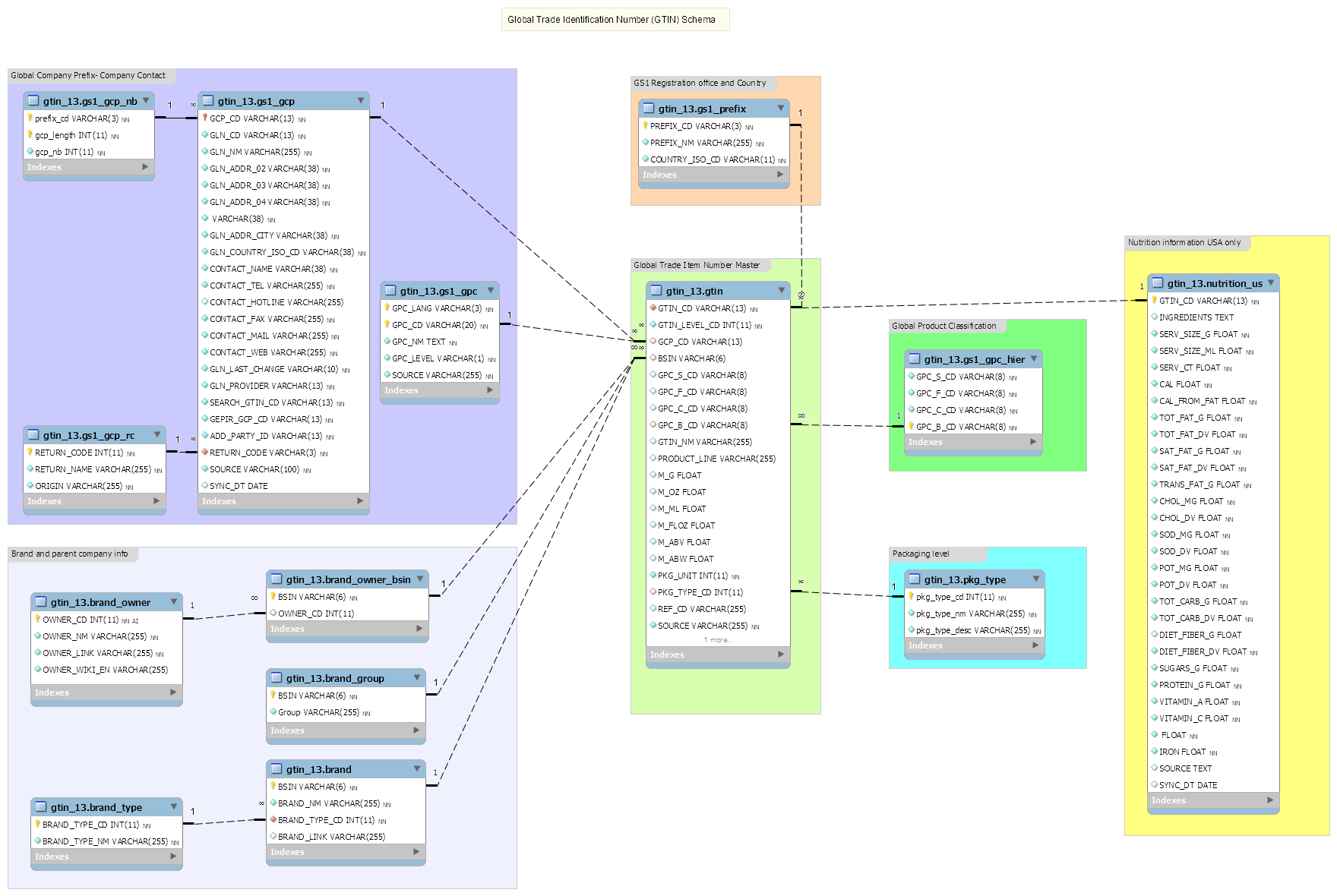


Figure - GTIN Schema