Data Reference Sheet



した時間を除

At The Consumer Goods Forum (CGF), and particularly within my work in the End-to-End Value Chain and Product Data, we are focused on making things easier and more effective for businesses along the entire value chain. This could be to help make a business more sustainable, or it could be to support efficiency and reduce costs and waste along the supply chain, for example.

I am, therefore, excited about the introduction of the CGF **Data Reference Sheet (DRS)**, which aims to considerably speed up the coupling of any two databases. It does this by leveraging the increasingly available power of Machine Learning Tools and a "red pill".

Every single day, databases need to be connected and disconnected somewhere in any given supply chain, be it onboarding a new supplier or customer, or creating new connections among inner-company databases. Usually, this is achieved via an API adoption through a process called "mapping", client by client, item by item, attribute by attribute. An IT project is set up, the data models of the two databases are compared and then jointly mapped.

This consumes time and resources.

In the past, one way to "simplify" this process was to force all participants to apply and implement the same data model and field names hard coded into their software, which poses different challenges in the setup and requires all participants to map the API: that is, each single company is spending the same amount of money and time to make things compatible.

The author assumes that most companies would rather not change their data model and storage system and hence prefer to delegate much of the mapping and tedious tinkering to software that is enabled to automatically adopt and map any two databases.

This got us thinking. Is there an efficient "middle ground"? Can we leverage software and machine learning to do the heavy lifting and speed up the integration? Can we fully automate the coupling of any two databases?

We believe the answer is "YES!". The core component of achieving this is a simple process utilising the aforementioned CGF Data Reference Sheet.

In the next pages is a description – things are about to get technical – of the basic concept and some simple examples to highlight the mechanics.

BASIC CONCEPT

Creating a Data Reference Sheet

At every step in each supply chain business partners agree to exchange a certain amount of data.

They come together in a meeting and record all required information and their attributes on a sheet of paper.

They then fill in the values, or actual data, using fictitious but plausible values. They create the data sheet of a "fake" product, including using an artificial name / identifier for that piece, product or material that will stand out and likely be unique.

For a consumer goods product, this means collecting all typical attributes describing content, packaging and sustainability, etc, needed to drive business.

If all companies in that sector agree on the same "list", we create a reference sheet:

This sheet of paper is called the Data Reference Sheet.

STEP 1

A Data Reference Sheet is made public, posted on a website and downloadable by all interested parties. It is regularly updated and maintained.

Those companies interested in exchanging information are asked to "digest" the information into their database. This means they task someone to enter the fictitious information into their master data or ERP system, creating an artificial product set complete with the fictitious values printed on the sheet.

The specialist creating this new data set and entering the data needs to make necessary translations, and manually store the data of each field in their home database that best matches the Data Reference Sheet.

- If they do not find an equitable field or attribute in their own system as proposed by the Data Reference Sheet, they simply ignore it.
- If their own system asks to enter attributes as part of the record or data set that are not part of the Data Reference Sheet, they simply leave the field blank (if possible).

During this process of setting up the "alien" record set s/he actually "maps" the Reference Data Sheet with their own system. They do this without changing the attribute of field names, and without changing the data format, that is, how the data is actually stored in their system. For example, they would store a 'date' like they always would.

STEP 2

Download and digest the Data Reference Sheet adding a new record entry to your own database.

We will exemplify this with a simple case. It needs to be adopted to all fields selected by consensus.

Let's assume the artificial product name is "CGF DRS v2021" or similar, that is, significantly unique to avoid doubles in any given system with 99.9% probability.

Let's assume the Data Reference Sheet contains a "Best Before Date". A proposed fictitious value would be 13-01-2023.

- This clarifies that "13" can only be the day as months count up to 12 in the Western calendar.
- "01" hence denotes the month of January and also shows how to deal with single digits (fill space with "0").
- The year is a four-digit number.
- This leads a machine to deduce the DD-MM-YYYY structure, separating data bits by a dash "-". That is a text string.
- This date cannot be used again in the entire Data Reference Sheet. It needs to be unique.

Now let's assume the "other" database has digested the data record with "CGF DRS v2021" and stored the Best Before Date in a field called "Fecha De Caducidad" (Spanish for 'Best Before Date') and it looks like this : "20230113". That is, the system stores the date as an integer, to make it feasibly sortable.

Connecting Any Two Databases

Next, what needs to happen to successfully couple two databases that are aware of each other and know they have "swallowed" the same consensus Data Reference Sheet?

When two databases are connected, the first thing to do is to request a full record extraction of a product called "CGF DRS v2021". If such a product exists, the machine will produce a "record dump" with all attribute names and values and send it back. As all values are fictitious, no secrets are shared.

STEP 3

Simple Machine Driven "Handshaking"



The original machine that sent the request will now be able to "map" the received record on its own, utilising the knowledge of the CGF Data Reference Sheet.

How does this work?

- It knows that somewhere in this record it will find a unique date 13 January 2023. It does this by scanning all data fields with specific algorithms until it hits a field with "13", "Jan", "2023", etc. or "20230113". Machine Learning helps to identify all options.
- Once it found this date, it knows how to read the data structure: in this case here it would be an "integer", structured "yyyymmdd", without separator).
- It now maps the attribute name it finds in the file received, "Fecha De Caducidad ", to its own "Best Before Date" knowing this is how the other database names it.

Algorithms establish an automatic mapping mechanism

Fecha De Caducidad	Best Before Date
int{20230113}	string{13-01-2023}

machine was able to map both databases on all values that were stored. It will only miss those data that are part of the official CGF Data Reference Sheet but were not digested into the host system for a variety of reasons.

- for "Fecha De Caducidad" of any given product to obtain the Best Before Date.
- this date. From now on it can, without error, read out the BBD from other products as it is aware of the underlying data structure of the connected system.

This means, a Standard naming convention immediately obsolete, or, at minimum level, optional. enabled by a simple Data Reference Sheet.



Data Reference Sheet

Next Steps

We learned how leveraging today's power of machine learning and other tools can significantly speed up and automate the connection of any two databases in real time.

The process is enabled and reinforced by digesting a "pill" into each participating database. The pill is always the same and consists of a Data Reference Sheet developed in industry consensus. The sheet lists all "attributes" and "typical" unique artificial values that help to describe a data set that needs to be exchanged in a unique way.

Once all systems have been 'immunised' with this reference set, programmes can connect and adopt the necessary mapping to each system on the fly and in real time.

At the CGF, we are calling for action on creating a sample Data Reference Sheet and we are now looking for volunteers.

Tests connecting databases will be run in conjunction with DataPorts, the CGF's open-source data exchange protocol available under dataports.io. Please <u>contact me</u> if you'd like to learn more.

Author: Rudy Hagedorn Director, End-to-End Value Chain (+49) 17 16 93 55 83 r.hagedorn@theconsumergoodsforum.com



