

SPONSORED BY:

MOBILE



Transform Your Business With Data

[ABOUT US](#) [REGISTER](#) [LOGIN](#)

[HOME](#) [BLOGGERS](#) [MESSAGES](#) [POLLS](#) [WEBINARS](#) [RESOURCES](#) [VIDEOS](#) | [STRATEGY](#) [ANALYTICS](#) [TECHNOLOGY](#) [FB](#) [TW](#) [LI](#) [G+](#) [RSS](#)

[FINANCIAL SERVICES](#) [HEALTHCARE](#) [GOVERNMENT](#) [EDUCATION](#) [RETAIL](#)

BIG-DATA PLATFORMS & TOOLS

Big Data Pools



Robert Plant, Associate Professor, School of Business Administration, University of Miami
 3/13/2013
 Comment
 3 comments

[Login](#)
 50%
 50%

Data pools are among the most important technical systems facilitating B2B trading between companies. However, they are relatively unknown by the broader community outside of the electronic data interchange (EDI) supply chain mechanism.

Five basic steps

The concept behind data pools is that businesses need to exchange information about products with each other, such as the product ID, the volume of the packaging, ingredients, weight, MRRP, barcode and RFID information, etc. Data pools follow five basic steps: first, suppliers place their product data into a common standard (such as EAN.UCC); second, they upload their data to a "source" data pool. This source pool can, if the company is big enough, be developed internally or can be outsourced to a third party specializing in data pools. The third step involves the source data pool sending meta-data about each item in the data pool to a "global registry," a database that holds information on items from every contributing company's data pools. With the data pool metadata established, the global registry's customers can now search for product information sets that they are interested in and place that in their own "recipient" data pool. The fifth step involves the synchronization of the "source" and "recipient" data pools; after this has been done the two pools are automatically synched every time a change to a product's data is made.

Commercial data pools are therefore natural big data repositories, central to which is **GS1**, which has global operations; GS1-US, for example, serves more than 200,000 companies, and a subsidiary, **1SYNC**, has more than six million Global Trade Numbers in its Global Data Synchronization Network (GSDN), through which it processes nearly 60 million transactions annually.

The P&G example

Tying into a data pool requires companies to create substantial data sets for their products. For example, P&G, which has 300 brands, would need to create data pool records with every SKU associated with each brand and its associated meta-data, and then perform updates on these attributes when a change occurs. This is in essence a Master Data Management problem, but one that can, if errors are embedded in the data pool, have a cascading effect across the recipient data pool users.

Ensuring big data pool MDM integrity is not an easy process; but the following three steps can be useful in maintaining high-quality governance of MDM data sets.

1. First, adhere to standards. For MDM it is vital that the organization adhere to the industry standard for its sector. This will enable systems, vendors, tools, and human capital to become uniformly aligned, leveraging the knowledge across the industry group as a whole. For example, the financial industry needs to comply with Basel II, MiFID, and PII, while in healthcare, HIPAA dominates. Adherence to standards also allows a clear directive to be given to the IT organization, without which slow erosion of standards can take place and a blurring of requirements occurs.
2. Second, create a controls policy. Data will need to be modified and changed over time, as will data types and additions made. A strict authorization policy needs to be put in place to ensure that this is done within accordance to the standards defined in the IT organization. A tracking and logging process also needs to be put into place to ensure compliance and accountability.
3. Third, try to automate the data management processes as much as possible. The validation and verification processes then only have to be undertaken once, as automated systems can operate over the whole MDM solution. Product master data will itself be drawn from a variety of sources including solutions from external data sources such as ERP and the supply chain, which may be error prone. Additionally, data that is being added from a merged data set such as in an acquisition needs carefully developed automated ETL processes to be enacted.

MDM data that is tied into data pools can not only be "big," but more importantly, its "size to impact" ratio is very big. As recently reported in the Financial Times (and repeated outside the paywall by Yext) "**data errors cost \$10.3B**" through incorrect or missing information, so if your description of a product is wrong in the MDM, that gets transformed through the data pool process, and then magnified through the synchronization of the global registry, resulting in missed sales and lost market opportunities. As such, it is imperative that the quality dimension of the data in the data set be measured for all databases and that governance policies are enacted in order to maintain the quality levels appropriate for its use.

Related posts:

- [In-Memory Databases: Fact or Fiction](#)

MORE BLOGS FROM ROBERT PLANT

A Review: Big Data in 2013
 39 comments
 Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

What Would Seymour Cray Think of Big Data?
 13 comments
 How would the tech legend respond to the 4 Vs?

The 360 Degree View of a Customer
 56 comments
 Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

Lessons for Big Data From President Obama's Healthcare Implementation
 24 comments
 Three big data lessons for CEOs from HealthCare.gov's bumpy start.

Is Big Data Ruining Sports?
 29 comments
 From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

[More from Robert Plant](#)



FLASH POLL

[All Polls](#)

DIGITAL AUDIO



Big-Data Platforms & Tools Radio

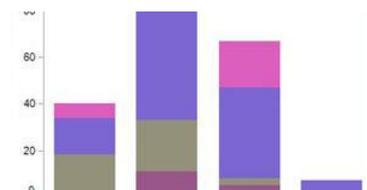
Latest Archived Broadcast
Who Can See From Your Platform?

We discuss the best internal setups to make sure the right sets of eyes are watching the data development.
[Listen](#)

[Full schedule](#) [Archived shows](#) [Download to calendar](#)

DATA VISUALIZATION SHOWCASE

Visualization: What's Stopping Big Data Improvement?



Survey results show bigger companies still confident in spite of massive IT overhaul requirements.
[Explore this data here.](#)

[More Data Visualization Showcase](#)

BDR IN YOUR INBOX

Enter email

- [Islands in the Stream of Extreme Data](#)

— Robert Plant, Associate Professor, [School of Business Administration, University of Miami](#)

[Email This](#) [Print](#) [Comment](#)

COMMENTS

[Newest First](#) | [Oldest First](#) | [Threaded View](#)



Saul Sherry, User Rank: Blogger
3/20/2013 | 5:25:58 AM

Re: Cloud

@RHRoop in the EHR case does that bring in a whole host of privacy concerns that the likes of P&G wouldn't have had to struggle with in creating these pools?

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



RHRoop, User Rank: Bit Player
3/14/2013 | 10:55:12 PM

Cloud

Very interesting. I'm working in the cloud space and we talk about standards such as HIPAA but from a compliance not competitive point of view. From a sales perspective this is very useful as it encourages me to talk wih potential customers from the point of view as compliance as a market maker rather than a cost. I was just talking today with a BPO that is looking at opportunities that result from the Affordable Care Act. EHR and the ability to share data with a customer care organization that will enable after care follow up will be a definite competitive advantage to insurers and providers.

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)



Saul Sherry, User Rank: Blogger
3/13/2013 | 1:02:43 PM

A reluctance to share

Robert, will we see companies opting into this... what is the P&G reason for involvement? Being that size do they stand to gain as much as smaller companies accessing the same pool?

[Login](#)



50% 50%

[Reply](#) | [Post Message](#) | [Edit/Delete](#) | [Messages List](#) | [Start a Board](#)

FOLLOW US ON TWITTER

Big Data Republic @BigDataRepublic 15 Apr

AllJoyn: A Common Language For Internet Of Things - Open-source lingua franca aims to let IoT devices and services... [ow.ly/2FHbY7](#)

Retweeted by Yakamoto

Expand

Hajime Ozaki @hajozaki 22h

BDR Top 20 #BigData David Smith David Smith List of top tweeters on Big Data from @BigDataRepublic [wp.me/p4Aavi-2](#)

Show Summary

Marcus Borba @marcusborba 28 Apr

#BigData #Analytics #BusinessIntelligence #FS @informationweek @AllAnalytics @BigDataRepublic @BigDataNetwork @BigDataClub

LIKE US ON FACEBOOK

Big Data Republic

Like

4,168 people like Big Data Republic.

Facebook social plugin

ACCOLADES

Media Pioneer Award Winner

BIG DATA REPUBLIC

[ABOUT US](#) | [CONTACT US](#) | [HELP](#) | [REGISTER](#)
[TWITTER](#) | [FACEBOOK](#) | [LINKED IN](#) | [GOOGLE+](#) | [RSS](#)

in partnership with

powered by

bigdatarepublic : /bigdatarepublic/section/2635 : /bigdatarepublic/section/2635/260142



UBM TECH

OUR MARKETS: [Business Technology](#) | [Electronics](#) | [Game & App Development](#)

Working With Us: [Advertising Contacts](#) | [Event Calendar](#) | [Tech Marketing Solutions](#) | [Corporate Site](#) | [Contact Us](#) / [Feedback](#)

[Terms of Service](#) | [Privacy Statement](#) | [Copyright © 2014 UBM Tech, All rights reserved](#)