Distributed Hybrid MDM, aka Virtual MDM Optional Add-on, for WhamTech® SmartData Fabric®
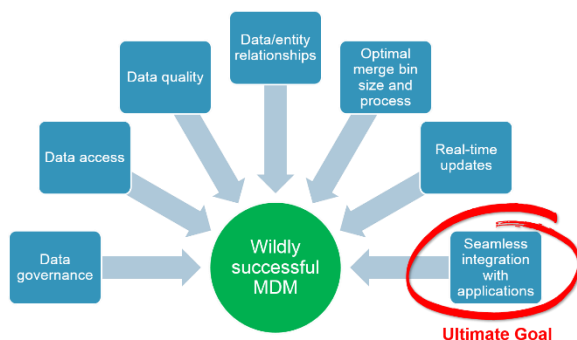
Revision 2.1

Contents

Introduction

Master Data Management (MDM) is a classic example of how conventional approaches do not fit well with real-life business problems, as they are generally independent systems that are not well integrated with the rest of the enterprise, including reporting, BI and analytics. Many organizations dump large amounts of data on a regular basis into a separate database for MDM or Master Patient Index (MPI) in healthcare, and work from there. In most, if not all, cases, it takes application writes or modifications to take advantage of, and integrate, MDM systems. Statistically, the main use of MDM systems is mailing lists for the marketing department – a considerable letdown from what could be achieved and, probably, what was used for justification for creating these systems in the first place. Ironically, well-implemented MDM is probably the single-most critical component of any successful and comprehensive enterprise application and/or analytics solution – a poor implementation creates a drag on business value, limits potential additional revenue opportunities and can result in significant losses.

Summary

WhamTech offers a unique and unconventional alternative approach to conventional MDM/MPI as part of SmartData Fabric® that enables a default, seamless and automatic integration of MDM/MPI with any and all data sources, e.g., operational/transactional (collectively, operational) systems. This integration of MDM/MPI works for almost any and all applications, including interoperability, reporting, BI, analytics, Big Data, Cloud and IoT. Applications do not need to be written or modified to take advantage of MDM/MPI, as normal queries from applications are automatically diverted to use master data as part of queries to individual data sources or multiple data sources. In addition, master data can be, optionally, substituted for the equivalent raw data in results from data sources. And one of the most important reasons for WhamTech's unique approach is that it enables secure and private master data to stay in the same environment as data sources, which is particularly important when multiple organizations are involved, for on-soil data retention requirements, data security and the European Union (EU) General Data Protection Regulation (GDPR).

Master data, along with other value-added data processes, helps bridge the gap between backend operational/transactional enterprise systems and frontend reporting, BI and analytics. This addresses one of the most difficult problems facing Big Data analytics on how to obtain a contiguous single, high quality view based on master data, e.g., of a customer or patient, including their associated data, across both operational, and reporting, BI and analytics systems. Also, the bridge enables feedback, in near real-time, from reporting, BI and analytics to operational systems, and thereby improves the customer/patient experience, increases revenue, lowers cost and increases profit.

It Should Be All About MDM

Master data is the best and often an improved version of key data that exists in data sources. As mentioned in the introduction, MDM is probably the single-most critical component of any successful and comprehensive data management strategy. Examples of solutions that master data are critical to, are:

- BI, reporting and analytics – to provide a complete understanding of the business and support valid decision-making

- Highly normalized semantic/ontological views of data used by link analyses and semantic/ontological model representations for knowledge management – to represent and interact with the knowledge base that comprises much of the value of the business

- Business Process Management (BPM) – to effectively monitor key performance indicators critical to a business and interact with processes to maximize the value of the business, improve existing processes and create new workflows

- Customer Relationship Management (CRM) – to view all data relevant to a customer/patient, regardless of its location, support reporting, BI and analytics, be fully informed and interact intelligently to improve a customer/patient experience, and maximize the value of a customer/patient and, by doing so, the business

- Data object-oriented application development – to represent legacy and other data in a modern architecture, provide great flexibility in the use of data, and support new workflows and applications (as mentioned below)

- Data/Web services/microservices – to develop new applications that can interact more effectively and fully with backend data sources, enable digital transformation, and support mobile and other modern applications

MDM should apply to all data objects, aka entities, important to the business, e.g., in healthcare HL7, a person who is a patient, a place that is a hospital and an organization that is a healthcare provider or payer. Master data and all other data in operational and other data sources should be connected and used as if in a single, very clean database.

However, conventional MDM systems are, typically, independent and separate from data access, and applications need to be written or modified to work with them. Whereas, an ideal MDM system should be completely integrated with normal data access and used without applications needing to be aware of it.

Three Conventional Approaches to MDM

There are three conventional approaches to MDM that attempt to integrate with other data, typically, operational data. These three conventional approaches reflect the three conventional approaches to data integration in general, namely:

1. Centralized database/data warehouse/data mart, aka master data repository, and copies of operational and other data – all data in one centralized database

2. Federated data access, aka master data registry – leave all data in sources

3. Hybrid centralized database repository and federated data access – some master data in a separate database and pointers to associated data in sources

We will examine each of these approaches with their own advantages and disadvantages in detail, as follows:

1. Conventional centralized master data repository and data

Typically, master data repositories are populated with master data on a regular basis and in many cases, these repositories are also populated with the latest aggregated operational/transactional data, resulting in a single database containing both types of data. This approach has advantages and disadvantages:

ADVANTAGES

- Single best version of master data

- Directly queryable

DISADVANTAGES

- Latency of data (not the latest)

- Data security and privacy, e.g., Personally Identifiable Information (PII) - master data is copied from data sources – this is particularly important when multiple organizations are involved, for on-soil data retention requirements, data security and the EU GDPR

- Large amounts of work and associated cost of ETL and master data merge operations, as the associated schemas tend to be complex, and the ongoing cost of maintenance and modifications tend to be high

- No direct connections with data source systems – there may be indirect connections through various IDs relating back to source data systems

- Applications have to be written or modified to use centralized master data repositories

- Normal operational/transactional applications cannot take advantage of master data, at least, not without being written or modified to do so

- Increasingly common problems of legal, compliance and privacy constraints being placed on sensitive data storage, which is particularly true for global enterprises that face national laws prohibiting sensitive data from leaving countries, aka on-soil data retention requirements

2. Conventional federated master data registry and federated access to data

Pure master data registries avoid the need to separately maintain and store master data by retaining pointers to master data in sources, using these pointers to retrieve master data as needed, and combining retrieved master data along with other data retrieved through federated data access. This approach has advantages and disadvantages:

ADVANTAGES

- Use of keys and link tables to point to master data in data sources – avoids ETL into a complex schema and synchronization

- Data security and privacy, e.g., Personally identifiable information (PII) - master data stays in data sources – this is particularly important when multiple organizations are involved

- Operational applications indirectly maintain master data – not a separate external application

- Existing applications can automatically use and substitute master data – with difficulty

- New applications can be written to take advantage of registry MDMs easier than repository MDMs

DISADVANTAGES

- Requires query access to, typically, multiple data sources to retrieve master data – subject to data source availability, query capabilities and data quality

- Query loads on adapters and data sources

- Frequent queries on source systems to check for updates if no changed data capture

- Bandwidth and query performance - scalability

- Retrieved source data requires cleansing, transformation, standardization, masking, tokenization and/or encryption, which can impact performance

- Typically, pointers tend to be limited to the best single record for an entity instead of a combination of pointers to multiple attributes in multiple records

- Difficult to maintain historic master data

3. Conventional hybrid centralized master data repository, registry and federated data access

This an approach favored by conventional federated data access vendors, such as Cisco Data Virtualization®, fka Composite Software®, ("Composite") and Denodo Technologies® ("Denodo"), that recognize that the pure conventional federated master data registries approach, as listed in approach 2 above, has issues. Composite, Denodo and other vendors, overcome some of the issues associated with either a pure centralized master data repository or a pure federated master data registry by combining a centralized master data repository with a registry and federated access to source data, as illustrated in the diagrams from Composite and Denodo below:
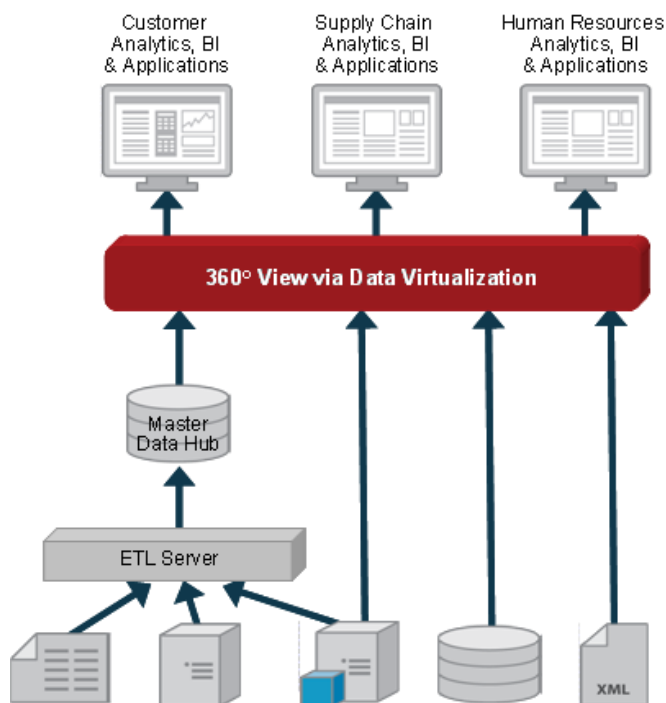

Figure 1: Composite's hybrid centralized master data repository, registry and federated data access
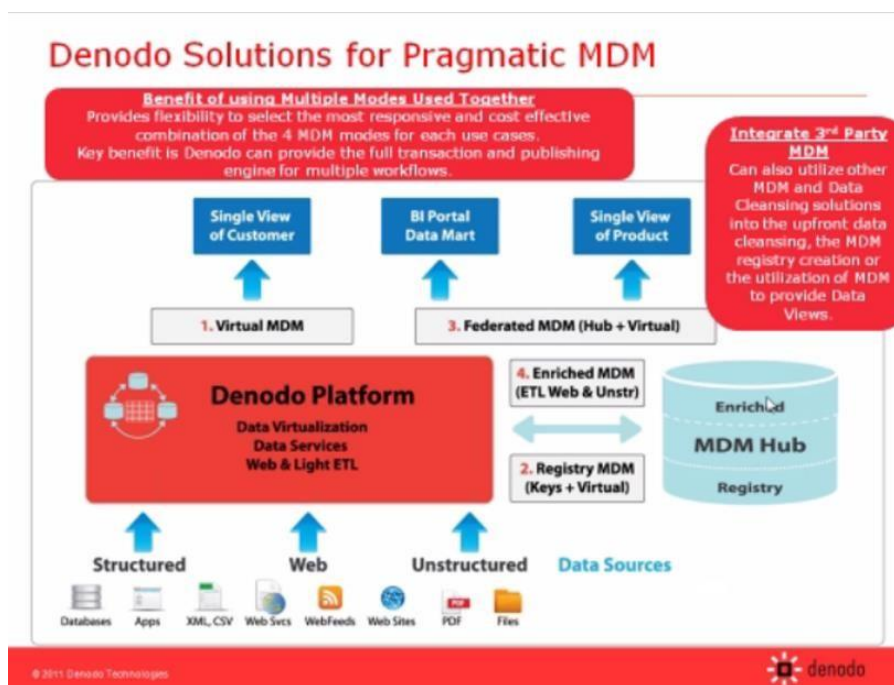
Figure 2: Denodo's hybrid centralized master data repository, registry, and federated data access

This approach attempts to leverage the important advantages and overcome the important disadvantages of both the centralized master data repository and federated master data registry approaches – a balanced combination.

ADVANTAGES

These are the main advantages of the pure centralized repository and pure federated registry approaches:

- Single best version can be updated on a regular basis, even in near real-time, subject to changed data capture

- Directly queryable

- Existing applications can automatically use and substitute master data – with difficulty

DISADVANTAGES

The disadvantages remain, to a greater extent, of approach 1. Conventional centralized master data repository and data, with the exception that master data is now connected to operational data source systems for updates and combination. The disadvantages remain as:

- Data security and privacy issues, e.g., Personally identifiable information (PII) or master data, is copied from data sources – this is particularly important when multiple organizations are involved

- Latency of data (not the latest), unless subject to changed data capture

- Large amount of work and associated cost of ETL and master data merge operations, as the associated schemas tend to be complex, and the ongoing cost of maintenance and modifications is a constant burden

- Increasingly common problems of legal, compliance and privacy constraints being placed on sensitive data storage, which is particularly true for global enterprises that face national laws prohibiting sensitive data from leaving countries

In addition, the disadvantages of federated data access remain for approach 2. Conventional federated master data registry and federated data access:

- Subject to data source availability, query capabilities and data quality

- Query loads on data sources through adapters

- Frequent queries on source systems to check for updates if there is no changed data capture

- Bandwidth and query performance - scalability

- Retrieved source data requires cleansing, transformation, standardization, masking, tokenization and/or encryption, which can severely impact performance

It is acknowledged that approach 3. Conventional hybrid centralized master data repository, registry and federated data access, is an improvement over either a pure centralized repository or pure centralized federated registry. In fact, WhamTech has taken hybrid approach 3 and extended and considerably improved it to retain the advantages of the hybrid approach AND overcome ALL of the disadvantages, as described next.

WhamTech's Unconventional Hybrid Distributed Limited Repository, Registry and Federated Data Access

Although WhamTech can implement conventional MDM and conventional Master Patient Indexes (MPIs), it offers a unique and unconventional alternative approach. WhamTech is realizing success in enabling a real-time hybrid approach that combines the best of the conventional approaches, including 3. Conventional hybrid centralized master data repository, registry and federated data access discussed earlier, yet overcomes the worst of these approaches.

Master data and master data indexes are an extension of the WhamTech SmartData Fabric®

WhamTech enabled the above-mentioned real-time hybrid approach by extending its SmartData Fabric® to add master data and master data indexes at the federated data access adapter index level. Out of the three conventional approaches described earlier, the conventional hybrid centralized master data repository, registry and federated data access approach exemplified by Composite, Denodo and others, represents a good balance between having a curated physical "golden record" and federated data access for the vast majority of other data. WhamTech extends this concept by distributing, storing and indexing limited master data of relevance local to data source WhamTech SmartData Fabric® adapters. Only master data directly relevant to an individual data source is stored and indexed locally. Not all master data need be distributed and stored, however, it may be indexed – some master data can stay in sources, and indexed and accessed as needed. Complex entities and frequently accessed entities, such as persons/customers/patients, benefit from being stored, indexed and associated with relevant data sources. Simple and less frequently accessed entities may not benefit from being stored locally and indexed, and indeed, some master data may be stored in reference data sources that can be accessed as needed.

The following diagrams illustrate WhamTech SmartData Fabric® configurations for four data sources, consisting of three federated adapters with indexes for three data sources and a conventional federated adapter with no indexes for one data source. Figure 3 is a basic configuration without a distributed master data repository and master data indexes. Figure 4 is the same configuration but with the repository and indexes
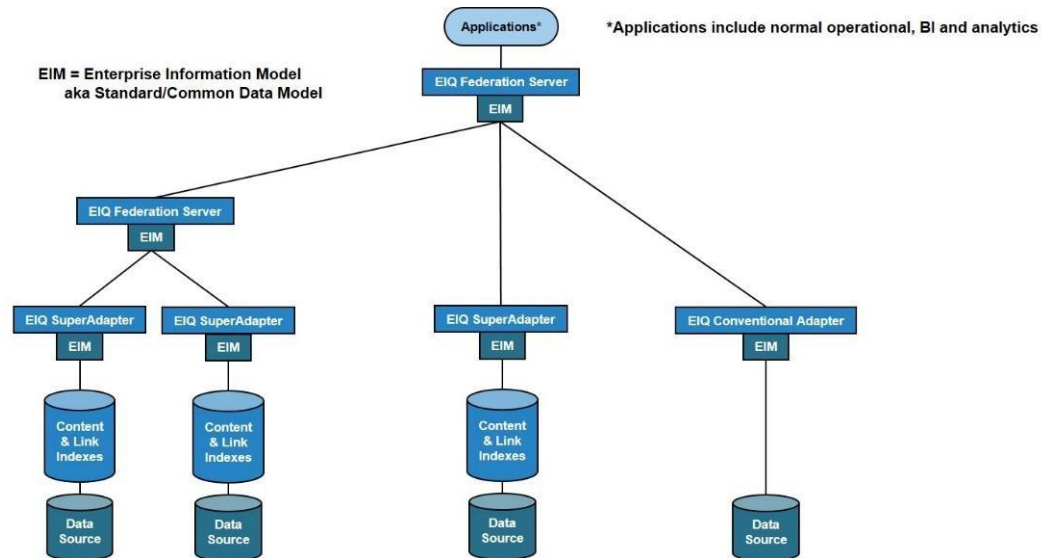
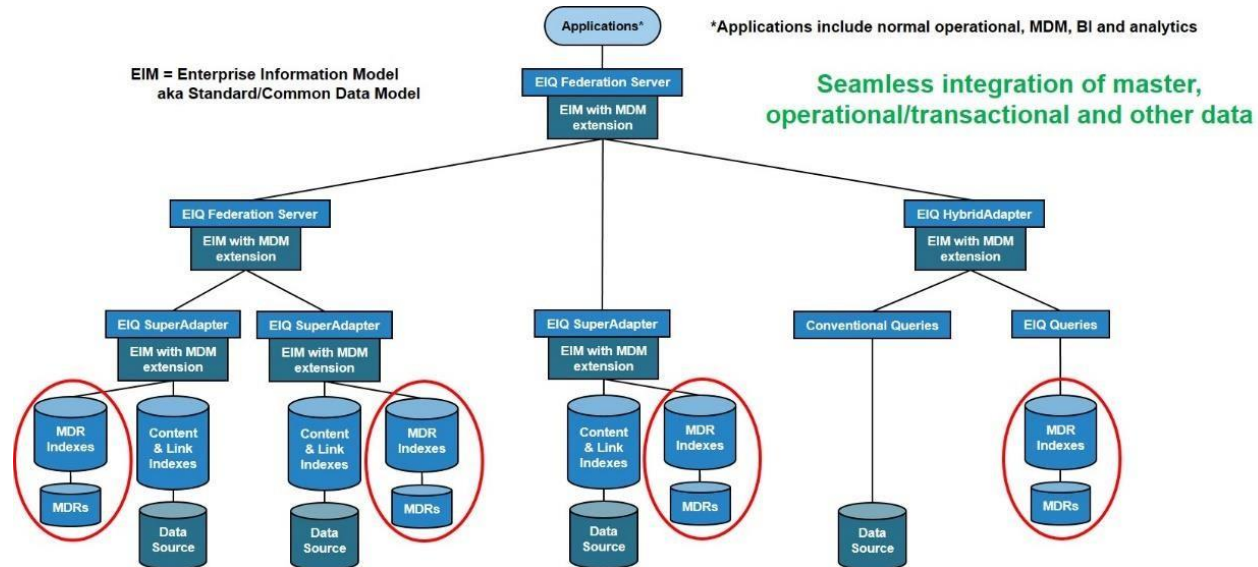**Figure 3: Basic WhamTech SmartData Fabric Configuration without MDM**

Figure 4: WhamTech SmartData Fabric Configuration with Virtual MDM

The seamless integration of master data with source data for almost any and all applications without writing or modifying applications to accommodate master data, is the main benefit of a number of benefits of the WhamTech distributed hybrid approach, as listed in the following table:

Distributed Hybrid	Limited Repository	Full Registry
• Data security and privacy, e.g., PII, master data stays in the same environment as sources • Adapter-level rules can integrate master data automatically – no application rewrites • Local processing - fast and secure • Use master data at multiple levels • Add/remove data source adapters with little or no impact • No central bottleneck or dependencies • Can still be centrally managed and consolidated	• Best master data may not exist as such in data sources • Master data immediately available – avoids (1) access to multiple sources and (2) source data transformation • Allows retention of historic master data	• Updateable • Traceable • Avoids repository for some master data

Figure 5: Benefits of WhamTech 's distributed repository, registry and federated data access hybrid MDM approach

Unlike any conventional federated data access approaches for master and/or other data, WhamTech addresses all seven keys to successful MDM, as illustrated in the diagram below:
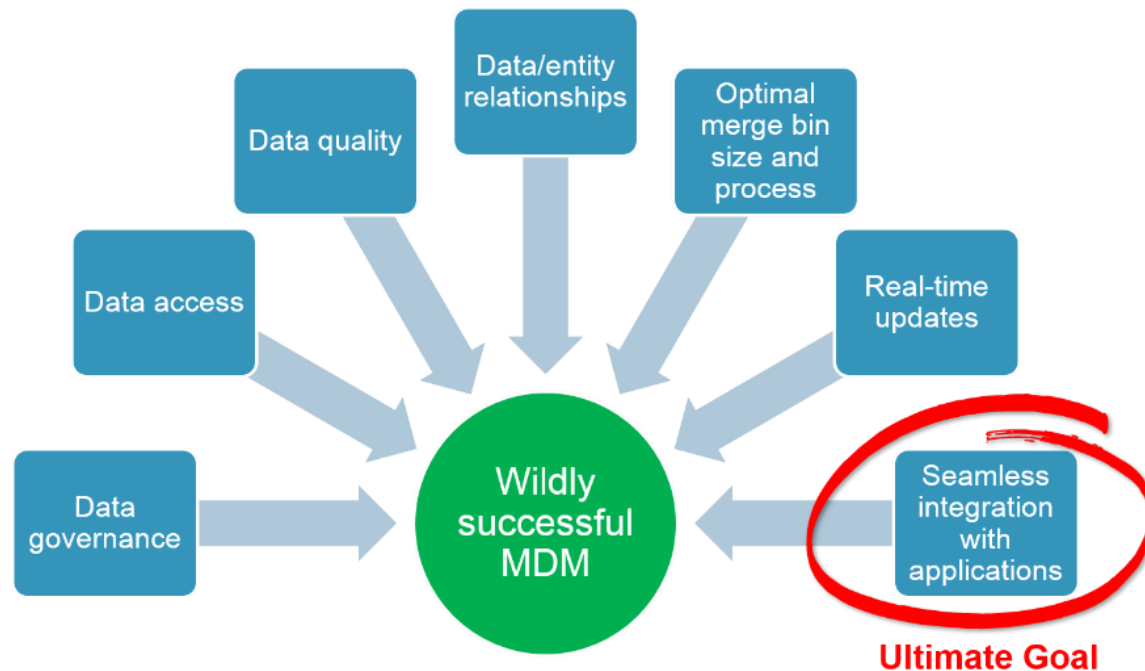

Figure 5: WhamTech addresses all seven keys to successful MDM

Seamless integration of MDM/MPI

The main benefit, by far, is the **SEAMLESS AND AUTOMATIC INTEGRATION OF MASTER DATA WITH SOURCE DATA FOR ALMOST ANY AND ALL APPLICATIONS WITHOUT WRITING OR MODIFYING APPLICATIONS TO ACCOMMODATE MASTER DATA**. That is, applications can automatically query on master data and have master data substituted for the equivalent raw data in results from data sources.

Master data leveraged to access and integrate data sources

All successful data management solutions go through a cycle of data discovery to insight and, ultimately, decision-making, as per the diagram below:
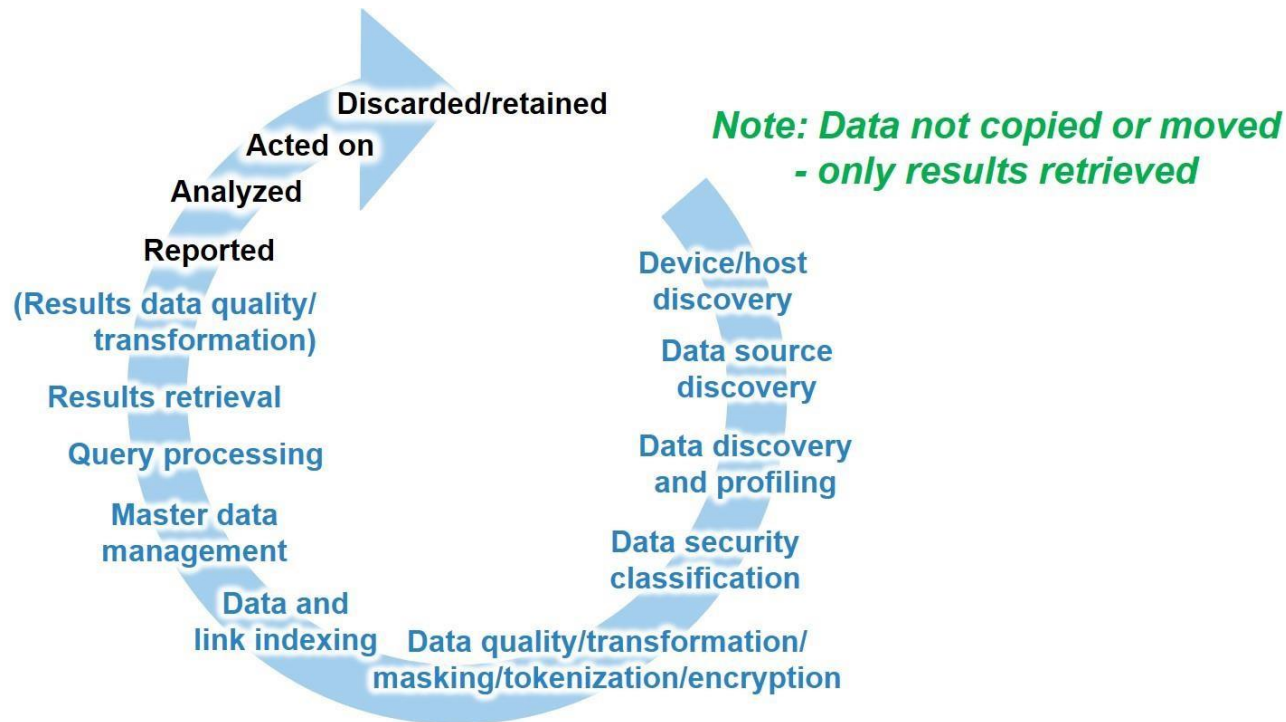

Figure 6: Successful data management cycle from discovery to decision-making

Using master data to access and integrate data sources, such as operational/transactional systems, is a major benefit for interoperability, e.g., Health Information Exchanges (HIEs), reporting, BI, analytics, Big Data, Cloud, and almost any and all applications.

Master data, along with other value-added data processes, bridges the gap between backend operational/transactional enterprise systems and more frontend analytics, as per the diagram below:
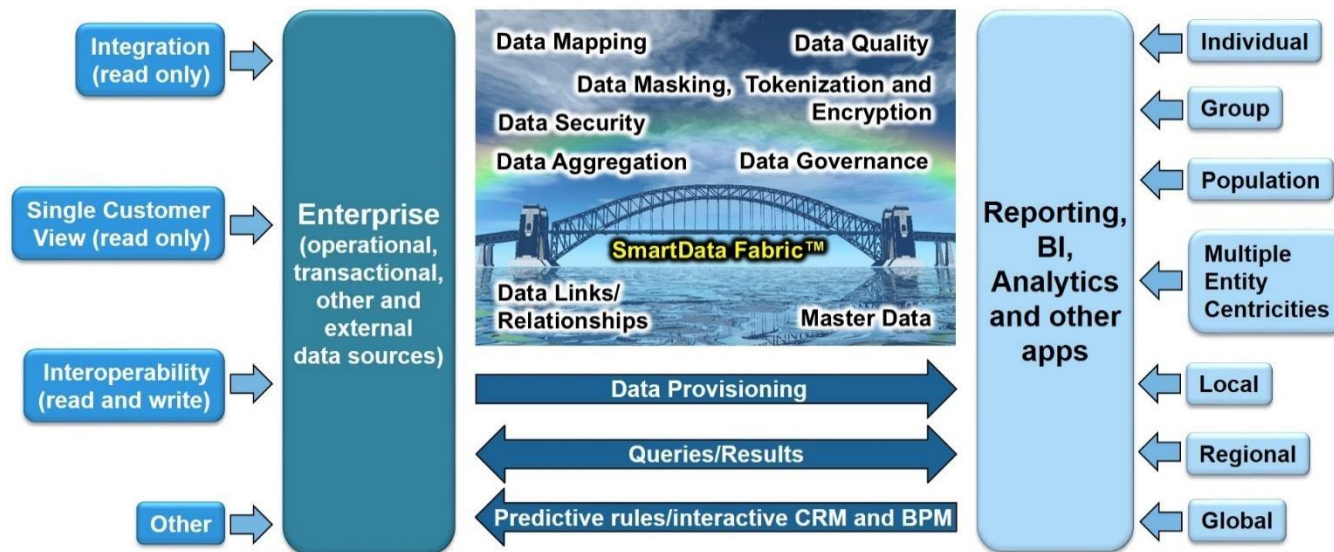

Figure 7: Master data, along with other value-added data processes, bridges the gap between backend systems and frontend analytics

This addresses one of the most difficult problems facing Big Data analytics on how to retain a contiguous single, high quality view of a customer or patient, as an example, and their associated data, across both operational/transactional and analytics systems. It also addresses how to feedback conclusions from analytics to operational systems in near real-time, and thereby improve the customer/patient experience, increase revenue, lower cost and increase profit.

Comparison of the Three Conventional Approaches and WhamTech's Unconventional Hybrid Approach

Benefit	Conventional centralized master data repository plus data	Conventional federated master data registry and federated data access	Conventional hybrid centralized master data repository, registry and federated data access	WhamTech's unconventional hybrid distributed repository, registry and federated data access
Latest master data	✖	✔ (depends on how frequently checked/updated)	✖	✔
Latest other data	✖	✔	✔	✔
Security and privacy – master data stays in data sources	✖	✔	✖	✔ (same environment as data sources in adapters)
NO Extract, Transform and Load (ETL)	✖	✔	✖ (described as "light")	✔
Direct connect with data sources	✖	✔	✔	✔
Applications need NO write or modification for – seamless integration of master data with other data	✖	✖	✖	✔
NO data source access required for master data	✔	✖	✔	✔
NO query load on data sources for master data	✔	✖	✔	✔

Benefit	Conventional centralized master data repository plus data	Conventional federated master data registry and federated data access	Conventional hybrid centralized master data repository, registry and federated data access	WhamTech's unconventional hybrid distributed repository, registry and federated data access
NO query load on data sources for other data	✓	✗	✗	✓ (minimal)
High query performance	✓	✗	✗	✓
Master data requires NO cleansing, transformation, standardization, masking, tokenization and/or encryption	✓	✗	✓	✓
Allow multiple data source records to be used to construct a single master data record	✓	✗	✓	✓
Easily maintain historic master data	✓	✗	✓	✓

Conclusions

The main conventional approach to MDM/MPI is (1) a centralized master data repository, which works if all other data is centralized and there is an integration of the master data with other data. However, this approach is rarely completely implemented and has been less than successful in most cases, resulting in the primary use of MDM systems as mailing lists for marketing– a considerable letdown from what could be achieved and probably what was justified for creating these systems in the first place. This is ironic considering that well-implemented MDM is probably the single most critical component of any successful and comprehensive enterprise application and/or analytics solution.

Two other conventional approaches that achieve an improved integration of master data with other data are (2) a pure registry and federated data access, and (3) a hybrid centralized repository, registry and federated data access. The latter approach (3) overcomes many of the problems facing the other two approaches, however, problems remain, including the ultimate goal of a simple and seamless integration of master data with other data without having to write or modify applications, and master data security and privacy, which is becoming an increasing concern.

WhamTech takes the latter, (3) hybrid approach to the next level using the unconventional hybrid approach of distributing and maintaining the master data repository as part of unique index-based federated data access adapters. In doing so, WhamTech MDM/MPI overcomes most

problems associated with any and all of the conventional approaches, and realizes the ultimate goal of a simple and seamless integration of master data with other data without having to write or modify applications, and addresses master data security and privacy.

WhamTech offers Virtual MDM/MPI as an integral and optional module for its distributed data virtualization platform, along with other essential data processes and optional modules.

About WhamTech, Inc.

WhamTech, Inc. (WhamTech) is a privately-held US-owned Delaware corporation established in October 2000 and based in Dallas, Texas. WhamTech's mission is to develop indexed adapter-based data virtualization, federation, integration and interoperability software products. WhamTech develops these products to anticipate, meet and exceed the demands of customers seeking an alternative to the conventional approaches of data warehouses, federated data access with conventional adapters and enterprise search. Our goal is to provide a better and more seamless way to work with data by changing the way data is accessed, analyzed, integrated, shared and interoperated. WhamTech adapter products are based on independent, cleansed indexes that execute both structured and unstructured queries for data warehouse capabilities, including results when data sources are unavailable and/or for archive.

Information on WhamTech solutions, sales and services, and partnership and investment opportunities can be obtained through whamtech.com.