



Unify Your Data with Master Data Management (MDM)

Steve Crosson Smith
SDG Group UK



Contents

Introduction	1
The Quest for a Single View – Unifying Your Data	2
The Main Functions of TIBCO EBX	3
Integrating TIBCO EBX Into Your Data Architecture	4
Integrating with a Data Warehouse using TIBCO Data Virtualization	4
Summary	5
About SDG Group	5



Introduction

The dream whereby all activities within a company, and hence all of its data, is centralised, governed and controlled by a single system has not, in the main, come to fruition. ERP vendors have been trying to achieve this for over 30 years but rarely does one come across a company completely orientated around any single system. It is not unusual to find one or more accounting, CRM, order management, warehouse management, product lifecycle management, logistics, HR, customer support management and a variety of other systems within mid to large-sized companies.

Generally speaking, each data source has its own set of IDs and naming conventions with no overarching means of relating entities across them, let alone spotting duplicates. The rise of cloud-based Software as a Service offerings only exacerbates the situation.

Alongside most systems (including ERP installations) is commonly found a plethora of Excel spreadsheets containing reference data, for example: territories, hierarchies, groupings, corporate calendars, foreign exchange rates, budgets, etc.

Controlling how entities are represented and created, gaining a centralised view of them, managing duplicates and applying common rules and conventions across such a variety of systems are daunting, if not impractical, tasks.

A further challenge is to ensure that all data users have a common understanding of the assets at their disposal: the nature and purpose of the data, a common view of calculations and what they mean, the rules governing how that data may and may not be used, the impact of that data, and its supply, upon the business, for example.



The Quest for a Single View – Unifying Your Data

It is the goal of most companies to have a single, trusted view of the various domains within their organisation, e.g. customer, supplier, product, employees etc. and, ideally, the inter-relationships between them. As mentioned, additional reference data is also often required to provide context.

Let's consider some of the challenges that prevent this:

- Different IDs within each system means that data can't easily be related, if at all
- Inconsistency of entity descriptions (e.g. customer names, product names, supplier names etc.) further prevents it being reconciled across systems, causes confusion and leads to duplication
- Siloed and ungoverned reference data (often in Excel spreadsheets) can lead to different perspectives, even on common data
- Poor and inconsistent data quality leads to a loss of trust in analysis derived from it
- Unidentified duplicate entities prevent a consolidated view
- Lack of control over who can create, read, update and delete key entities centrally leaves the company at risk
- Inability to find key information across multiple systems leads to delays and creation of duplicates
- Lack of a common understanding of the nature and authorised usage of the data prevents effective governance and leads to many conflicting versions of the "truth"

How then, is it possible to relate all of this data in order to achieve the much-coveted 360 degree view required to optimise the value of the data to the business? How can duplicate information, both within and across systems, be identified and how can a single, authorised, trusted and consolidated version be adopted, deployed universally and unambiguously understood?

The optimal answer to these three challenges is the deployment of Master Data Management (MDM), Reference Data Management (RDM) and Metadata Management (MM) respectively.

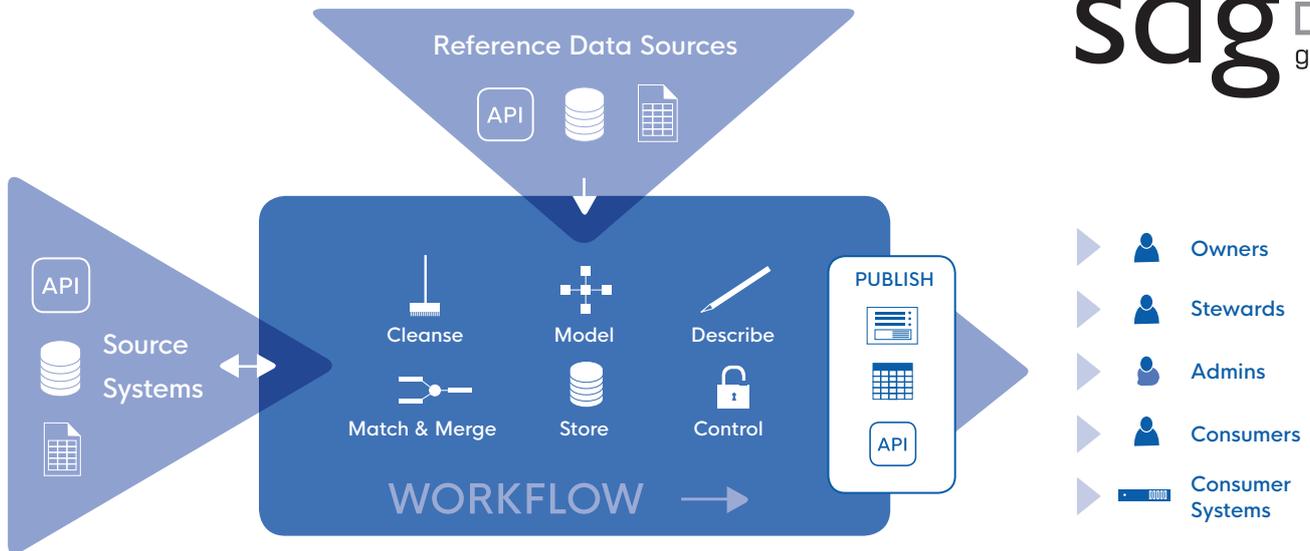
For sake of clarity, let's explain what we mean by these three terms:

Master Data – this is the data describing your key domains, e.g. customers, suppliers, products or staff.

Reference Data – this is data that provides context to your domain data e.g. hierarchies, groupings and supporting information such as exchange rates, corporate calendars, etc.

Metadata – this is "data about data" i.e. descriptive information that enables your data to be understood and utilised correctly. Typical examples are business definitions of data fields, data types, names of the systems where the data can be found, rules around its use, etc.

Most vendors treat these separately and two or three solutions are required to implement MDM, RDM and MM. SDG Group UK works with the TIBCO EBX solution, which handles all three, so for the rest of this paper will use this for illustration.



The Main Functions of TIBCO EBX

The diagram summarises the main functions of TIBCO EBX, which are:

- Master data, reference data or metadata is loaded, or entered, into the system (and updates are loaded on an ongoing basis), either via its own data loading mechanism or often using an ETL or other integration tool
- Data cleansing is performed, either by the integration tool or it may be implemented within TIBCO EBX
- Data is fuzzy matched within TIBCO EBX to detect suspected duplicates and associated records
- An automated merge or insert process is usually executed where the degree of confidence is high, otherwise exception records are routed to Data Stewards for action
- Key attributes of the data are modelled within TIBCO EBX (e.g. data types, IDs, field names, relationships, validation, constraints etc.), independently of the source systems, and a master ID is generated by it which can be used to tie the individual source system record IDs together via a “Golden Record”
- Metadata can be modelled and maintained within TIBCO EBX, adding descriptive context to enable governance
- The data is stored in a centralised repository
- Create Read Update Delete (CRUD) permissions within the MDM system are applied
- Workflow ensures that processes are followed and that appropriate tasks may be automated
- The data is made available (published), along with descriptive elements, to users and external systems (see below)

For many organisations, the MDM system is the only one in which master data may be created and these new records, having been validated and approved, are then made available to all of the relevant systems around the enterprise. This approach brings the greatest level of consistency and control.



Integrating TIBCO EBX Into Your Data Architecture

Having created a centralised, trusted and governed source of master data, reference data and metadata within the organisation it is important that it can be integrated into the overall data architecture.

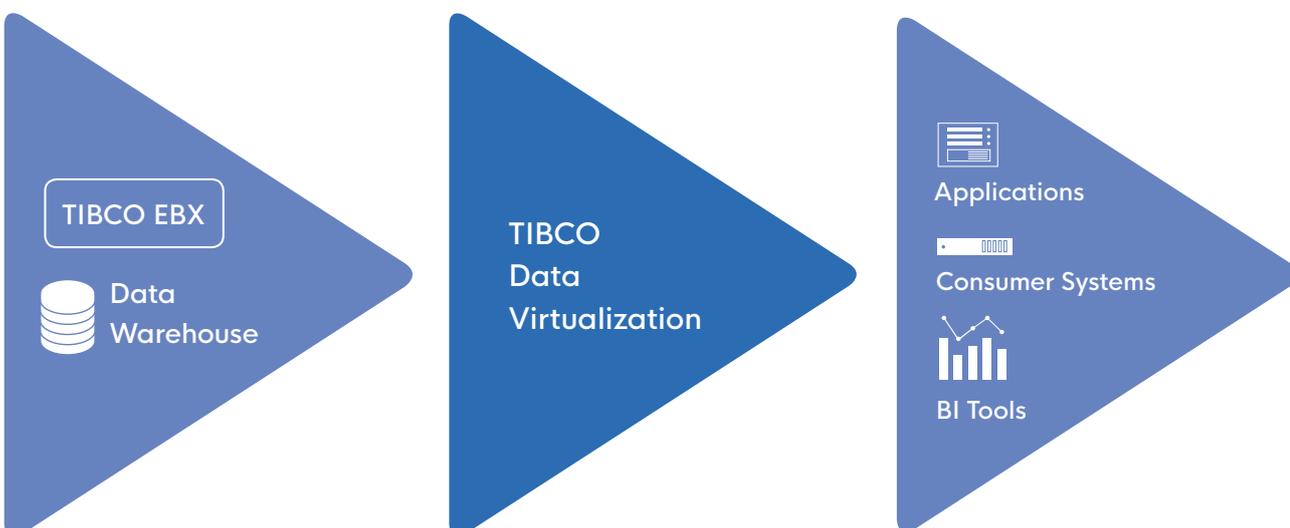
TIBCO EBX provides a number of options for this and it is commonly used in a variety of ways:

- As a manual reference source via the provided customisable user interfaces (including mobile)
- To update host systems when changes in master or reference data occur, usually via integration software
- To provide real-time access via API calls using SOAP, REST or Java standards
- Data can be published into database tables in the EBX repository which are accessible via ETL and BI tools. These can be copied into, say, a data warehouse to provide context to transactional data or, better still, integrated in real-time using TIBCO Data Virtualization (see below)

Integrating with a Data Warehouse using TIBCO Data Virtualization

TIBCO Data Virtualization, another product in the TIBCO Unify stack, is an abstraction layer which sits over data sources and, amongst other things, can provide query federation. This means that physically separate data sources are effectively joined and presented to consumers as a single source, as far as a SQL query is concerned.

This is an ideal way to maintain master, reference and metadata in the MDM system and join it to transactional data in data warehouses or other systems, without needing to copy it.





Summary

Master Data Management, Reference Data Management and Metadata Management permit data from disparate sources to be unified, centrally managed and governed.

TIBCO EBX is unique in that it supports all three in a single system and data models can be completely user-defined. It is aimed at business users and has a configurable user-interface that enables application views customised for the role concerned. It has a number of integration options and provides a truly enterprise-class solution.

Its powerful features and flexible workflow engine enable a single, holistic view of the key data distributed around your organisation and applies governance to its creation, management and utilisation. It provides a system-independent, graphical mechanism for modelling the important entities in your business and a means of defining relationships between them, as well as the metadata attributes that describe them.

About SDG Group

SDG Group is an international consultancy, solutions and services provider specialising in data management, business intelligence, advanced analytics and integrated planning.

Headquartered in Milan, it has almost 1,000 professionals worldwide with over 20 offices. Data management includes migration of data warehousing into the cloud, data pipelines, ETL and MDM.

SDG Group is an Elite TIBCO partner, working with their Unify and Predict product sets (which include TIBCO EBX, TIBCO Data Virtualization, TIBCO Spotfire, TIBCO Streaming and TIBCO Data Science). Other key SDG Group partnerships include Snowflake, Fivetran, Matillion, Tableau, Looker, Anaplan and Board.

If you would like to discuss any of the topics raised in this paper email infouk@sdggroup.com quoting mdmwhitepaper.

www.sdgggroup.com/en