



Data Management Strategy Formulation

WHITE PAPER SERIES

Author: Edward Hoofnagle

For further information, please contact:
Edward Hoofnagle, CEO
ed@ebusinessware.com, +1 212 232 0285 Ext 101

Formulating a Data Management Strategy

This document defines the approach that ebusinessware (ebw) would follow to evaluate and define a reference data management strategy.

Scope of Reference Data

The subject matter of this whitepaper is focused on reference data (as opposed to transactional or operational data and statistics).

For purposes of this analysis, reference data should be understood as a defined set of core data elements that are used in the mission-critical computer applications and basic workflows of an enterprise – examples include customers, vendors, products, etc.

Definitions of reference data types/categories include:

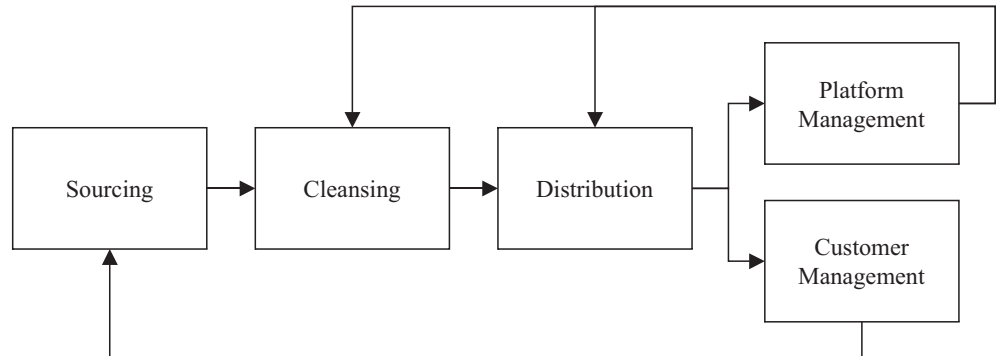
“Party” Reference Data – the name, address, government id, and related details about a company or “entity” that the company does business with or wants to know about (customers, vendors). Included in this area are industry identifiers, legal hierarchy information and other corporate data such as officers and directors that pertain to the organization or “party”.

Product Data – the details about the stocks, bonds and other financial instruments that support a firm's capital structure. This concept can be extended to include derivatives on the securities.

Market Data – the details about the transaction prices of stocks, yields for bonds, exchange rates for currencies, and other pricing indicators used in financial modelling. We also include credit ratings and other probabilities of default in this category.

Data Management Framework

We have developed a data management analytical framework (illustrated below) to enable our analysts to develop a “top-down” view of the strategic landscape.



The reference data management framework covers 5 key areas, with definitions as follows:

Sourcing	Sourcing means the process of gathering reference data and placing it into a database – whether from third parties or directly from the provider described by the data.
Cleansing	Cleansing means reviewing data to make sure that is correct. The primary responsibility of a research analyst is to perform this process. Sometimes people may refer to cleansing as “verification” – these terms (cleansing and verification) can loosely be considered synonyms.
Distribution	Distribution means the transmission of data from its database to the consumer of data. Companies distribute data in various formats – via an Internet Portal, in computer files and through web services. In general, the data is “no good” unless people can integrate it into their own systems.
Platform Management	Reference data sourcing and cleansing results in a large database. The cleansing tools and distribution engines require substantial hardware and software. Platform management refers to the activities associated with creating and managing this sophisticated technology that supports the reference data.
Customer Management	Even though reference data is critical to the success of a company, many people do not even know that it exists. As such, we recommend that our clients make sure they identify the internal users within their own company as well as the external customers of data.

Overview of Potential Engagement

1. Background and Objective

This document is intended for companies that intend to design a data strategy for the enterprise. The background research for preparing this strategy will include a review of the tools and processes which are used for managing data inside the firm “current state”. After the current state is inventoried, one must define the future state goals, taking into consideration best practices and available technologies to support the desired future state.

ebusinessware can provide assistance in defining the current and future state as part of our data consultancy. We can provide solutions including:

- Inventory of Existing State of Data Processes, Flows and Tools
- Elaborate a data strategy that is aligned with the enterprise goals
- Project Management
- Technology, Operations and Data Planning
- Implementation assistance and outsourcing

2. Our Analytical Approach

ebw will use the conceptual framework described above to create an inventory of repositories, data flows, operational processes and supporting infrastructure. In identifying suggestions for improvement, issues or “gaps”, our analysts would take into account industry-led best practices.

Following are some of the tasks that would be involved in creating a reference data strategy. Note that each customer's strategy project has its unique characteristics and requires a careful design of objectives and deliverables. Hence, this list is not intended to be exhaustive; rather, the goal is to provide indications of the types of tasks that you may ask us to perform as part of planning and launching the overall effort.

- Identify different types of reference data, their repositories, data sources and operational processes.

Objective

Analysis will be done to identify the different databases holding similar types of data and also to identify the set of internal/ external data sources feeding these databases. We also catalogue the people and processes that are involved in the value chain of capturing and storing reference data. This analysis is useful in measuring overlapping efforts going in supporting these databases. Minimizing duplicate data and/ or database redundancy (as well as data sourcing redundancy) will streamline operations and reduce cost.

Summary

There is more than one reason why organizations end up maintaining local copy of databases in their silos. This could be due to the rapid growth of business in different geographical regions and lack of corporate governance around IT infrastructure or it could be indecisiveness around data ownership.

When it comes to designing a data strategy for an organization we start by identifying the different products within the organization and their point of convergence. For each product, one should identify the internal and external sources of data – and preference should be assigned to each source. Internal source could be a ratings system or a new client database. External data sources could be an end of the day file from a stock exchange or a data feed from external data vendor.

Challenges and Solutions

As a global organization, coordination with business units becomes tedious and time consuming. It is always best to involve stakeholders in communications. In the beginning of the project, we help explain the project objective and benefits to the audience. We encourage information sharing about the initiative to motivate staff to work in a cordial and efficient manner.

- Review the list of required reference data, create a data dictionary

Objective

The overall data requirements are driven by your business offering and customer needs. Analysis is required make sure all necessary data attributes are captured and stored. A list of attributes (“data dictionary”) will be compared to current requirements and competitor offerings in the same space, gaps will be identified, and a strategy to fill gaps will be put in place.

Summary

Information that is captured, cleansed and maintained is the “soul” of a data business. A list of attributes and their coverage has to keep pace with customer's business requirements and competitors offerings. Moreover, customer data requirements may change with regulatory and compliance pronouncements.

A data dictionary is a critical document that is used from understanding the customer's requirements. In a properly implemented data strategy, the data dictionary is consistently applied from sourcing all the way to distribution. During a data strategy analysis, we investigate if attributes defined in the data dictionary are sourced, cleansed and utilized consistently across the enterprise.

Challenges and Solutions

Reference data is a key component of the business. It requires careful planning to make a comprehensive list of attributes required for each product type. Any missing information can cause a firm to lose a business opportunity and a decision to include an attribute at a later stage requires a lot of rework. The best practice is to share the effort with all stakeholders and get their approval. From a design perspective, we try to leave enough scope in the MDM platform to accommodate new attributes with minimal changes in code by using configuration-based solutions.

- Develop standard guidelines for data collection and identify approved data sources

Objective

The purpose of having data verification guidelines is to define company wide standards for the treatment of reference data. Guidelines vary for each reference data type. This is very useful documentation for the production teams while capturing information from approved sources and for data vendors who supply data. Since all teams and other sources follow standard guidelines it ensures consistency in data values across the board.

Summary

Data sourcing guidelines essentially define the best practices in the acquisition, cleansing and enriching of data. These guidelines not only manage the scope of research work, but also help customers to understand how data was created in the first place. In order to control the data collection process and to avoid any conflicts in decision making, it is essential to document assumptions and data sourcing rules in great detail.

Summary

If you are servicing more than one customer, it is inevitable that one or more customers will have a “unique “ set of requirements. Sometimes generic guidelines don't work for all. Best practice is to have a section which records exceptions to guidelines. Since this document keeps changing along with the business (it is a “working document”), version management is a critical task.

- Review the list of required reference data, create a data dictionary

Objective

Once an initial evaluation of data sourcing and cleansing is performed, the next step is to select Master Data Management (MDM) approach for each reference database. MDM model selection largely depends on data you want to host in the system and the business operations around it. In this phase of evaluation, we will examine the current state of MDM systems and suggest platform enhancements where required.

Analytical

- Synchronizing master data for data warehouses and their downstream clients for improved business reporting
 - This is a registry style database which is mostly used for referring meta-data residing on other databases
 - It is not the most suitable model if your objective is to improve performance of daily enterprise operations

Collaborative

- Collaborative MDM mode provides ability to maintain information at one place using a single master process that ensure validity and completeness of data
- It requires services to support workflow and check-in, check-out services to control the creation, management, and quality of master data
- After the information is complete and validated it is integrated and synchronized with legacy systems, enterprise applications, and data repositories within the enterprise, and the exchange and synchronization of information with business partners

Operational

- These modes support a true ability to manage uniquely identifiable data about each master data entity across the operational application infrastructure

- Review the data platform and platform management implementation

During the analysis phase, ebw can examine the existing systems and compare them to the recommended operating model. Our analysis will cover the following areas:

- To check if the systems support the mode of MDM implementation
- To check if optimal methods are being followed for data loading and distribution

- Develop data consolidation strategy

Objective

Fragmented databases complicate data maintenance and distribution tasks. One of the challenges with distributed databases is to keep them synchronized. We assist clients in defining a strategy for consolidating the data fragments identified in the inventory of data.

Summary

The most common practice for consolidating data is by using unique common identifiers between each database; like unique external identifiers such as DUNS, or RIC CODE. In the absence of external ID some of the industry standard identifiers (ISIN, CUSIP, SEDOL) can also be used or an internal numbering process can be put in place. If one is working with a customer database, then in the absence of any of these enrichments one can use name and other available attributes like hierarchy or address.

Challenges and Solutions

Successfully consolidating databases is crossing one of the biggest humps in implementing a data management strategy. If the database has inadequate enrichment and is missing common identifiers, then data consolidation becomes a challenging task. Sophisticated normalization and matching rules are useful techniques for data matching. We can provide technology and a knowledgeable team of analysts to help in resolving exceptions during data consolidation.

- Benchmark the data quality

Objective

To measure data quality and produce attribute wise data quality reports. Measuring data quality lets you know in advance if the data is good enough to meet an SLA signed with customers. If the quality is unacceptable, you can be proactive and take necessary steps to fix issues before customers complain.

Summary

Expansion of business largely depends on the quality of data being offered to customers. In today's world of severe competition between data vendors, data quality is one factor which can either make or break leadership status. In order to avoid operational risk, customers demand for strict service levels on data quality. Data quality assessment should be an ongoing activity.

Challenges and Solution

A large challenge in measuring data quality is to select adequate sample set of reference data. Sample sets should be statistically robust to form an opinion on the underlying data population, and sampling should be freed from biases. Data quality should be an independent task and should not be part of the data remediation handled by the production team.

- Evaluate the data distribution “channels”

Objective

To evaluate current modes and time-frames of data distributions to data subscribers.

Summary

Ebw will evaluate different distribution channels that a firm offer to its customers. In conjunction with your product managers, we will evaluate timeliness and availability of data distribution solutions. The objective of this evaluation is to propose possible improvements that would make it easier for internal and third party customers to get access to your data and improve profitability of data assets.

Challenges and Solution

This process would require workshops with IT staff and suggested changes may or may not be supported by existing platforms (in such case a gap analysis would be performed to evaluate the investment required to close the gaps). A team of data management architects will work with your IT team and could assist in the design and building of extensions to the existing infrastructure.

3. Potential Deliverables

As described in the above section, the output of a data strategy project can include a variety of analytical artefacts. Following is a summary of the types of deliverables that may be agreed-upon at project inception:

- Current state analysis for reference and market data (data flow diagrams, data attribute maps, systems schematics)
- Gap analysis for business processes and technology infrastructure designed for data capture and data distribution
- List of Data-related “issues” (e.g., definitions, quality, duplication, gaps, etc.)
- Data Dictionary, Definitions, and Process Summaries
- Data quality reports
- Draft Designs and Data Flows
- Strategy Documentation

4. Dependencies

Sponsors in key areas of the organisation will be required to commit time and resources during the analysis (and also during any subsequent implementation). Sponsors will be sought in all areas of the enterprise that are associated with the data “value chain”.

5. Logistics to Consider

Current state analysis

The analysis is normally broken into three phases.

Information Gathering

Stakeholders and contributors will be identified and agreed.

Information will be gathered through interviews with key personnel in all areas of the firm.

Information requirements will be mapped out in advance, where possible. A series of questionnaires tailored to the operational area of the firm, will be used as a rough guide to each interview. As a matter of course, interviews will be written up and distributed to contributors for verification.

The preliminary analysis will also include a proposed deliverable structure appropriate to the organisation and agreed with stakeholders. Depending on the organisational structure, we anticipate conducting from 40-50 primary interviews at this stage.

Analysis

The issues identified by contributors will be subjected to the following checks to ensure balanced and consistent findings:

- Completeness of information in terms of product, process and organisation
- Reasonableness checks based on experience
- Validation against current practices, including best practice, in other financial markets organisations

Follow-on interviews will be required to complete the picture. Independent sources of information may be consulted to validate controversial findings, which may include, where appropriate and available, systems extracts, costs and headcount numbers.

Delivery

Draft versions of current state documentation will be developed during this stage for informal review with stakeholders. Final documentation will be produced and formally reviewed with stakeholders. Documentation will include sections relevant to individual businesses (i.e. products) and individual functional areas from front to back office, as agreed in the preliminary analysis, as well as an overall summary. If required, eBusinessware can make a presentation to stakeholders and other interested parties.

5. Timelines and Schedules

Depending on the complexity of the assignment, strategy efforts usually have a duration of approximately 8-16 weeks. This elapsed time is dependent on a number of factors, the most important of which is the availability of key personnel.

About eBusinessware

eBusinessware is a specialist, practitioner-led financial markets solutions firm. The company is based in New York with offices in UK, Singapore, India and Tokyo. We work globally for a wide range of firms. We bring to projects deep knowledge and experience of financial markets, their instruments and products, and business processes spanning the organization. We also have a set of technology solutions and a team of outsourcing experts who can create a sustainable, long-term solution in a cost-effective manner. Our data related solutions include;

CreditDimensions – a reference data management platform and outsourcing utility.

EZ Data Manager – a software solution for integrating non-standard documents into a data platform using “smart parsing” templates, workflow automation and OCR technology for faxes/document images.

We have successfully assisted financial market organisations in the delivery of IT strategy and leading edge system implementations and apply pragmatic, cost-benefits approach to their projects. One of our areas of specialization is reference data management. We help our clients to identify and implement strategies that meet cross functional requirements and which work across the entire organisation. We understand the importance of reference data in an organization and its impact on process efficiency, compliance, risk management, and overall client service.

Case Studies

Reference data usage review and technology audit

A leading Investment Bank had a large number of applications that used reference data, although actual patterns of usage were inadequately documented and therefore poorly understood. The bank needed an independent audit to help it to form a strategy to move to a new, scalable solution.

We were engaged to analyze the consumption of instrument and client static data across the bank, from front to back and across all asset classes and businesses. For each application, we investigated and documented the data population, the attribute coverage and the delivery method from both strategic and supplementary data sources.

Our report gave the client a comprehensive understanding of how different applications within the bank were using reference data, and made a number of recommendations to improve the data population, the coverage and the delivery mechanisms. Our advice was presented in terms of industry best practice for data management, based on our observations over years of involvement in this specialist area.

Reference data strategy

A leading European Investment Bank needed a detailed specification of requirements for all aspects of its current reference data management. Because of our expertise in this area the bank engaged us to help define a new global strategy. Two data experts reviewed global usage of all key data types, including party (client and issuer), product, person and all related details - e.g. markets, calendars and mandates.

Their report gave detailed specifications for all aspects of reference data management, including data ownership, data models, operational maintenance processes and maintenance screens. Our analysis and specifications helped to mitigate the risk in the company's data management process by providing a detailed definition of what had to be implemented. The client subsequently asked us to deploy a team of experts to manage the delivery of this significant global initiative.

Reference data strategy

We were asked by a large European Investment Bank to conduct a complete audit of its reference data. The objective was to improve process and control efficiency and to recommend how to meet users' needs more completely.

Our consultants mapped out the existing processes for reference data capture and maintenance. We then analysed the process flows and dependencies by system and assessed the extent to which business requirements were met.

Information was obtained through a review of system documentation, as well as by interviewing key personnel. We conducted 20 interviews with system owners and data managers, and a further 10 with key business representatives. We used the information gathered to produce a set of data process flow diagrams and some clear recommendations, both tactical 'quick wins' and strategic solutions. This three month engagement was completed on time and to budget.

Product requirements specification

We were asked by a major financial markets data and service provider to develop a detailed business case for a new enterprise level data service. The objective was to develop a new product, leveraging the organisation's existing data assets, but in an area not part of the organization's core products.

Information was gathered through close collaboration with client personnel and using market contacts to assess market demand. Our consultants reviewed the range, coverage and quality of existing data assets and processes for data capture and maintenance. The team also conducted external market review across four major groups of potential customers and assessed the appetite for the proposed product against cost and other factors. To supplement gaps in the data, internal sourcing and outsourcing options were evaluated and short-listed.

Models and conclusions were validated with the client and with select potential customer contacts all the while maintaining confidentiality of the commercially sensitive project. A detailed business case was delivered to the client comprising: a definition of the product in terms of data range, coverage and quality and specifying the interfaces required for customer delivery of the service; a commercial model; product roadmap and plans for the implementation; and a high-level sales and marketing outline. The business case was successfully used to secure launch funding for the product and to drive through acquisition of data for implementation.

MDM Implementation

Assisted the firm in defining and implementing a Master Data Management (MDM) solution for their counterparties. We also integrated this MDM solution with other systems inside the bank.

In the beginning of the project our team of senior analyst spent time with stakeholders to understand their vision. The project started with the analysis of the current state of business at that time. It was identified that customer data was created and managed in multiple disjointed silos and thus creating a huge overlap in information and effort to maintain information. As first step a plan was made to consolidate data from multiple databases into a database and assign unique identifiers to each entity. Matching and exception handling technology was built by ebw IT staff using the Credit Dimensions platform.

As a next step, a data dictionary was made to record list of attributes that the client would need to host in MDM solution for each entity type (corporate/ bank/ trust/ funds etc.). The Data Dictionary was made keeping Basel II requirements in mind. For every single attribute of data dictionary data cleansing guidelines and data sourcing rules were made. All this guidelines were specially tailored to the client's nomenclature of reference data.

Ebw IT staff worked with the client's IT department in the development of web based application for data governance and supporting data enrichment and maintenance process. An Ebw team of data analysts currently use this application for cleansing and enriching reference data as per the guidelines on an outsourced basis.

Market Data Product Launch

Worked for a market data company to help them in expanding their product offering by adding counterparty reference data to it. Our subject matter experts created detailed data sourcing guidelines and business rules for benchmarking data quality. Entire initiative was supported by a team of eBusinessware data analyst and technologists. Our production teams enriched approximately 170,000 counterparties and at the same time proactively monitored them for any changes. Transitioned responsibilities to the internal teams over a 3 year period.

Risk Modelling Data Automation

Worked with a money center bank to define and set up a process for automating the data gathering and cleansing for ratings models. We implemented a solution using the Credit Dimensions platform where our system accepts, from the bank, monthly feeds of external rating agency identifiers and of the firm's internal identifiers. Our outsourced team of analysts perform automated & manual matches of the External Credit Agency IDs to the internal ID using such criteria as legal entity name and domicile country at an Obligor or Issuer level.

The resulting cross-reference is then made available daily via reports and a feed back to their counterparty data management team. The client Users to access the External Mapping information where available and feed a Global Risk Ratings process.

We have also helped the client to design MIS for to compare internal ratings with external ratings providers and identify shifts (changes). Once shifts are identified our analyst performs research and validates shift in information.