

PRODUCT PAPER

Cordys Master Data Management

Understanding MDM in the SOA-BPM Context



EXECUTIVE SUMMARY

Rolling-out new Service-Oriented Architecture (SOA) based applications has become exceptionally challenging, due to the broad cross-functional scope of newer business applications, changing regulatory and security requirements, and the demanding real-time nature of today's business interactions. Web services protocols directly address component-based applications, but what about the data?

Consider the scenario of a process that requires no human intervention (straight-through processing) of an important aspect of a company's business.

As Diagram 1 illustrates, the executing process crosses multiple boundaries – such as departments, organization units, and geographic locations. The process exchanges business data as a part of the process execution flow. In the absence of an enterprise-wide consistent representation, understanding of key master data elements, the problems of data inaccuracy and incompleteness can hinder the smooth flow of process execution.

NATURE OF THE DATA PROBLEM

Typically, most corporate data is stored in relational databases. Though database schemas have been optimized for lines of business, many new applications require data that cuts across business lines, departments and geographies. Meeting these new information requirements requires not only streamlining access to data and ensuring efficient delivery of bits over the network, but also transforming and integrating data from multiple, diverse sources to create new consolidated views of information that are meaningful for today's applications and decision-makers.

- A Business Process exchanges business documents (Transactional data) in a defined flow
- Business documents refer to Master Data
- Business processes that need to be executed beyond system boundaries need to refer to the same master data representations through their flow

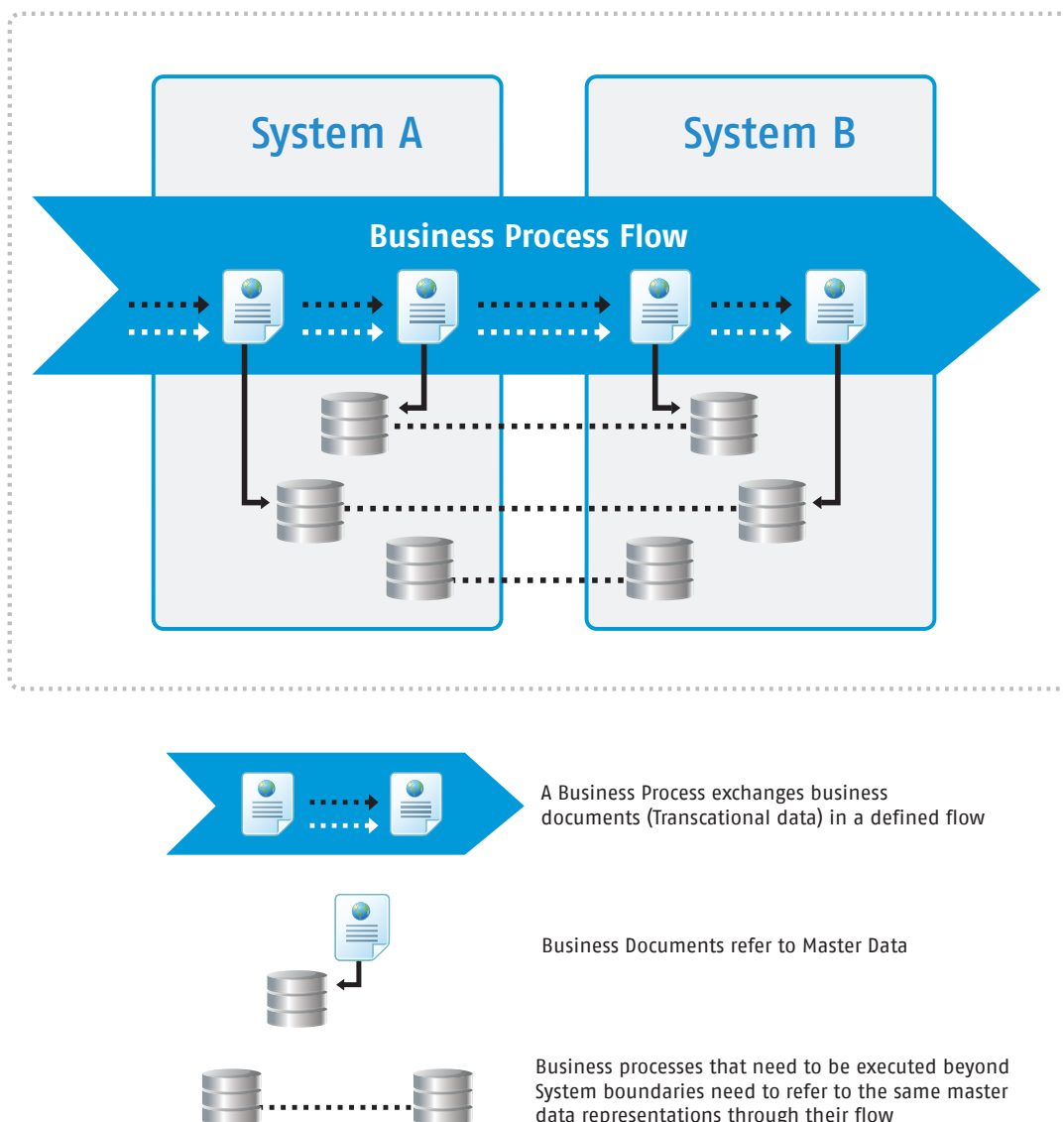


Diagram 1: Straight-through processing across multiple boundaries

“Seemingly small data quality issues are, in reality, important indications of broken business processes.”

Michael Hammer, Reengineering the Corporation

As Michael Hammer suggests in his seminal work referred /cited above, many organizations suffer inefficiencies in process workflows and decreased employee productivity as a result of costly and time-consuming manual interventions needed to overcome the challenges of disparate data definitions and inconsistent meanings.

To alleviate these issues and prevent them being compounded further, organizations must address data integration issues ahead of SOA-driven Business Process Management (BPM) initiatives, in order to reap the promised benefits of such programs. Enterprises that implement BPM without a coordinated strategy towards data integration activities will significantly lower the performance of their business process improvement activities, because each design point will have to verify data semantics, requiring more developer resources and time. Specifically, in the absence of a focus on data integration, the following data challenges become stumbling blocks during SOA-based application development:

- Finding trusted, authoritative information sources (master data stores)
- Knowing the underlying location, structure, context, quality and use of data assets
- Determining how to reconcile differences in meaning (semantic transformation)
- Understanding how to ensure the appropriate levels of quality of data elements.

MASTER DATA MANAGEMENT DEFINED

Master data is reference data about an organization’s core business entities. These business entities include:

- **People** – customers, employees, suppliers, partners, etc.
- **Things** – products, assets, etc.
- **Places** – locations and geographic classifications of interest
- **Other key entities**

Master data provides the business context for recording operational business transactions in an organization. To serve this purpose efficiently, master data should be trustworthy.

Master Data Management (MDM), viewed in the above context, can then be defined as a Business capability enabling an organization to:

- **Identify trusted master data** – MDM defines and/or derives the most trusted and unique “version” of important enterprise data
- **Leverage master data to improve business processes and decisions** – MDM incorporates the master version of data within functional business processes (sales, marketing, finance, support etc.) that will provide direct – contextual – benefit to employees, partners, customers or other relevant stakeholders within an organization

TRUSTED DATA POWERS YOUR BUSINESS

Trusted data is data used by business stakeholders to support their processes, decisions or regulatory requirements with no reservations as to the data’s relevance, accuracy, integrity and other previously agreed definitions of quality. Typically, the usefulness/utility of data can be measured against the following well-defined data quality dimensions:

- **Accuracy** – the degree to which data correctly represents the real-life objects they are intended to model, measured by how the values agree with an identified source of correct information
- **Completeness** – all relevant data attributes should be assigned values in a given data set
- **Consistency** – data values of a data element drawn from two different sets should not conflict with each other
- **Uniqueness** – no “real-life” entity exists more than once in a data set and there is a key that can be used to uniquely identify and access the entity
- **Timeliness** – refers to the time expectation for accessibility and availability of information

Thus, trusted master data serves as the fundamental enabler for enterprises to support some of their core business needs such as:

Single view of the business

When organizations wish to gain valuable insight into their operations in the context of its complete relationship with different business entities that it engages with, it becomes necessary to assemble operational data from a variety of relevant transactional systems. This results in the creation of the so-called “Single view of the business” or “360 degree view”. Without the capability to access master data from a single trusted source of information, this assembly, or consolidation process, becomes a futile exercise. In many instances, it could result in misguided decisions stemming out of incorrect data comparisons across inconsistent versions of the information base.

Automating recurring straight-through processes

As observed at the beginning of this paper, enabling automated execution without human worker involvement requires trusted data.

Delivering operational insight

Organizations need a view into operational processes to understand and react to unforeseen situations, arising during process execution, in order to respond in a timely

fashion. This would require the assembly of information from different sources including the Process Engine, Line Of Business (LOB) systems, etc. The presence of a trusted data layer with services to consolidate and present an aggregated view of related information is absolutely essential to satisfy this business requirement.

In summary, MDM is not viewed simply as a technology space; it is a business capability enabled through the integration of multiple technologies and business processes. The Cordys Business Operations Platform (BOP) recognizes the critical role that data has to play and hence delivers a set of cutting-edge tools and frameworks in its MDM module to address the challenges of data in organizations.

CONCEPTS UNDERPINNING CORDYS APPROACH TO PROCESS-DRIVEN MDM

The MDM capabilities of the Cordys BOP are supported by a deep understanding and insight gained into behaviors and principles governing data management practices in organizations. The following core concepts are at the heart of the process-driven MDM framework of Cordys BOP.

Process orientation

The core essence of managing master data, or shared data, effectively involves a continuous process comprising complex activities involving a multitude of organizational resources, including interaction amongst automated systems, amongst human knowledge workers and between systems and humans.

Within this collaborative effort, the following activities might be involved in a typical lifecycle of a master data entity:

- Automated sub-processes to perform data integration tasks such as data extraction, transformations, data quality routines to cleanse, standardize and parse data, find duplicates and manage potential matching candidates, validations and checks
- Accessing third party external web services for additional information to support data enrichment
- Manual intervention of knowledge workers or data stewards to handle data exceptions, knowledge-driven classification, manage data hierarchies, and to enforce data governance policies such as business rules to enforce on master data

This enables collaboration in many ad-hoc, dynamic ways, not necessarily pre-defined and configured in a rigid and structured application flow.

Most other vendor offerings advocate conventional approaches to MDM, an approach emphasizing structured workflows requiring costly and time-consuming customization efforts. To make any changes needed as part of data governance practices need in-depth technical knowledge of the product.

Thus, to support collaboration across these multiple participants, there is a clear need to provide a robust process-oriented infrastructure which can be used as part of a platform. This platform can be used by business-aware data stewards and information architects, as well as IT users such as Web service developers and data administrators,

to both model and execute the process steps in a single, collaborative environment.

Cordys MDM is designed to satisfy these compelling needs surrounding process-oriented lifecycle management of data.

Event-driven data processes

Complementing the process-centric nature of master data management is the knowledge, that the lifecycle activities on master data are undertaken in response to well-known and identified business data events – occurrences of actions that are of business relevance to the organization.

These events may occur within the immediate scope of the organization's own business or may in some cases occur outside its boundaries (such as a general market impact occurrence) whilst still carrying the potential to significantly impact the organization's business.

Thus, it is important to understand that MDM, as envisaged with Cordys BOP, represents the convergence of processes and events – master data is subjected to various data processes which are invoked in response to relevant data events. This reinforces the use of Cordys BOP as the underlying infrastructure to power MDM actions, since both processes (and their multiple patterns) and events can be modeled, developed, executed and monitored in a single environment.

Diagram 2 shows an example of a process-based, event-driven lifecycle of a master data entity.

The master data object is typically following a life cycle process right from the instance of being created until it is persisted in a physical data repository. Such a lifecycle behavior is typically described by the following characteristics:

- Many possible variations of data operations, ranging from data cleansing actions, automated data validations, human intervention enabled approval workflows, to complex data integration actions associated with publishing the object(s) in a controlled manner
- Sequence of orchestration of actions is not determined at design time (when designing the MDM framework) and is dependent upon actual incoming business object attribute values and existing business object records in a repository – a runtime determinant
- An object instance undergoes several "transformations" in an orchestrated flow of state transitions triggered by different types of events. For example, there could be a first transformation step to a format suitable for matching, and further during the process, another transformation to a format suitable for Marketing Service Provider (MSP) processing for data enrichment.
- Data management actions are initiated in response to specific lifecycle events that could occur in any random fashion, defying the notions of predictable sequenced-driven business workflows. Thus, there are almost limitless data integration scenarios possible in several combinations of variables. These cannot be handled by hard-coded framework logic.

Cordys MDM employs the formal method of state charts for the specification of processes associated with data

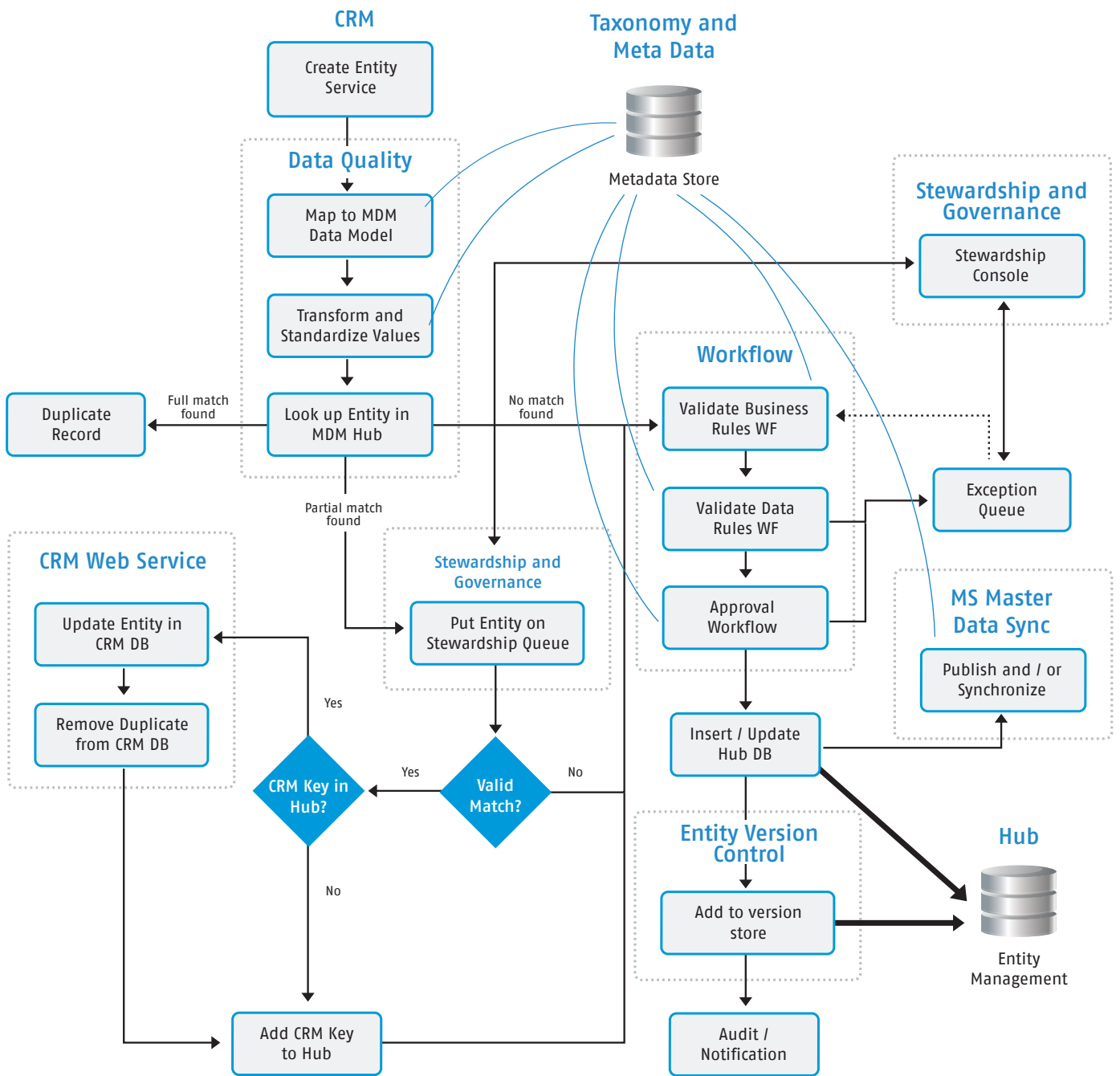


Diagram 2: Process-based, event-driven lifecycle of a master data entity lifecycle management based upon data states

integration. A state model is a graphical representation of the sequence of states of an object. The life cycle of master business entities or objects can be defined in terms of a State Model. This model contains states and transitions.

Events trigger the transitions and activities (any type of data management tasks, either modeled as automated web service, a process flow, a human worker-driven data form with logic, a decision case set of rules, etc) are performed as part of the transition flows.

By using state charts as the specification method, users are able to model data tasks and processes which can be automatically executed by a State Machine Engine.

Pluggable data quality tasks to support data governance and stewardship

Data governance involves establishing data ownership along with data stewardship roles and responsibilities to proactively manage information risk, regulatory compliance and ensure enterprise shared data is "fit for purpose".

Data stewardship is the executive role whose primary responsibility is to ensure that governance rules and policies are actually applied in practice.

Data quality (DQ) tasks support the investigation and analysis of the state of quality within data repositories as well as facilitate cleansing of data records to support data

stewardship and governance processes. Most commonly known instances of applying data quality are in cleansing of names and address data.

Data quality services are highly domain dependent, geography specific and knowledge-driven tasks. The logic applied to clean customer names and addresses and identify potential duplicates in a data set cannot be reused with product / item data descriptions. Thus, data quality services have to be chosen and applied based upon specific data domain context.

Cordys MDM recognizes this domain-specific nature of DQ and hence advocates an architecture that facilitates easy plugging of best-in-class data quality services from vendors or even allows consumption of customized cleansing logic developed in-house. This ensures that the framework is open and at the same time can leverage investments in existing assets.

To demonstrate the capability of the pluggable DQ framework, Cordys has proven the integration with two market-leading DQ vendors – Human Inference and Trillium. These implementations have interesting characteristics:

- DQ services can be consumed either as external web services or called “in process” via API calls
- Contrary to the traditional use of DQ services in

batch mode, cleansing and matching functions of the DQ vendors can be consumed in an “in line” mode, interactively DQ tasks can be invoked in response to specific events modeled in state transitions. Depending upon the response of the DQ tasks for a given input dataset, data stewardship could be enabled by letting human knowledge workers oversee and manage the quality of data being generated as part of the data integration process.

Let’s now turn our attention to consolidate our understanding on how Cordys MDM, as a key component of Cordys BOP, can help organizations succeed in creating a single, trustworthy view of their business operations – the single version of truth.

Building the single view with data services

The single view of the business lies at the intersection of the concepts and capabilities described above. When you combine these different classes of services to invoke event-based workflows that apply tasks to manage data, the result would be the creation of a ‘Single Version of Truth’ – a representation of the object that is accurate, unique, consolidated, consistent and up-to-date, which can then be consumed in business process work flows of the enterprise.

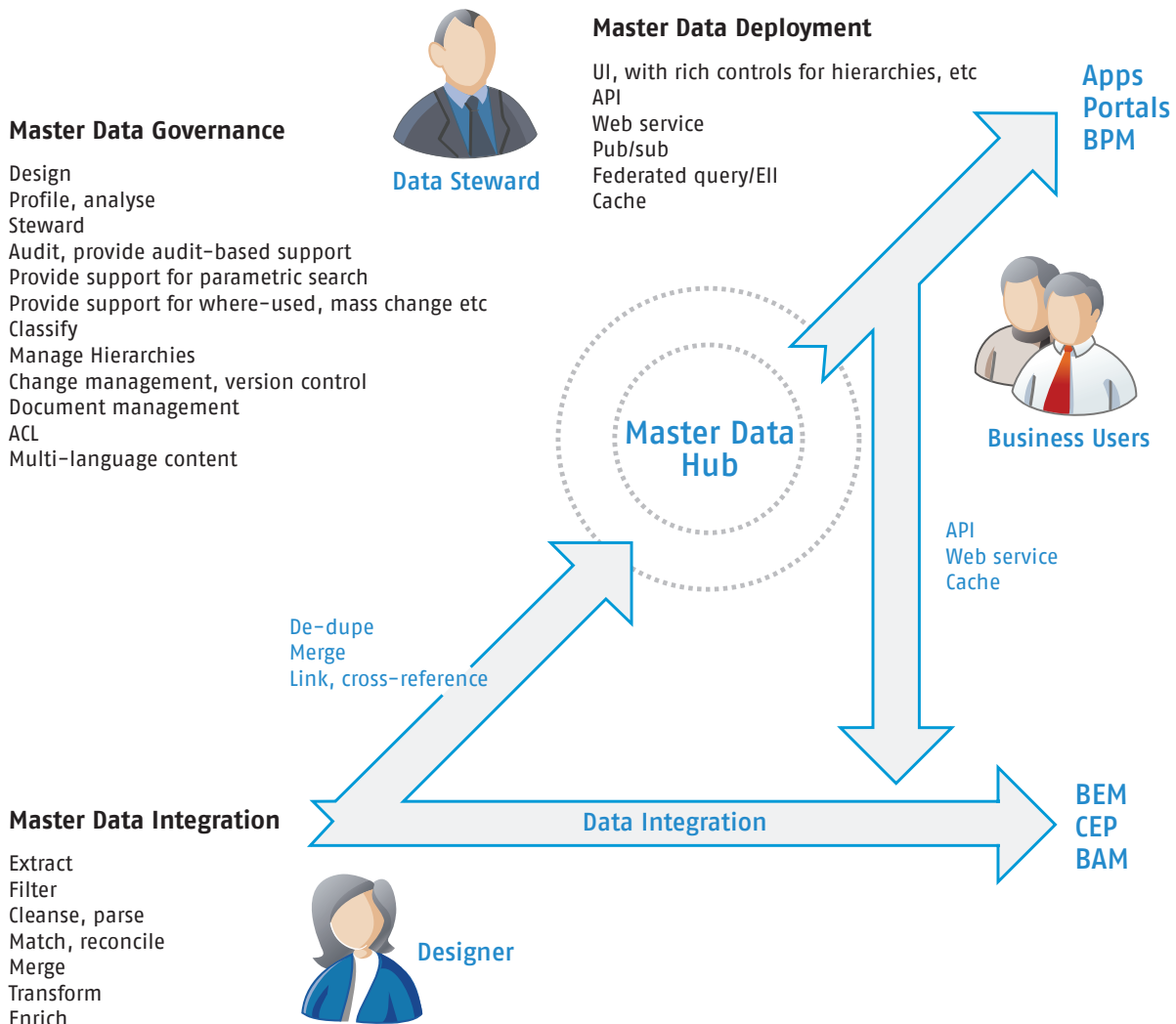


Diagram 3: Building the single view with data services

Different resources of the enterprise (automated systems, knowledge workers with specialist skills) can creatively collaborate in a single, user-friendly environment provided by Cordys BOP to model the tasks and activities required to assemble the single view of data for the business user community.

OTHER BUSINESS APPLICATIONS OF CORDYS MDM

The Cordys MDM framework, built on the key concepts described in previous sections, can be applied in multiple data integration contexts to support SOA-based composite business applications. The following is an indicative list of scenarios where Cordys MDM can be used:

- 1. Simple data synchronization across shared data instances** – in near real-time, scheduled, in batches or for every instance of change
- 2. Complex data synchronization across multiple data sources** – in situations which require a combination of data integration, data validation, data governance, human-intervention enabled data approval work flows etc., need to be applied
- 3. Data consolidation to create a trusted data source** – involving data reconciliation, aggregation, transformation, data stewardship, etc., to create a high-quality source of data. Depending upon the business requirements, different architectures can be applied for the creation of the trusted data hub. These are:
 - I. Cross-reference registry** – in this scenario, the data hub contains enough data elements/attributes to be able to recognize similar instances of data records across multiple data sources and link them together

through the generation of a new enterprise-wide unique identifier

- II. Co-existence hub** – here, the consolidated hub contains the most accurate, high quality trusted data for shared data elements that are used or referenced throughout the enterprise in multiple business processes. This hub also forms the backbone for maintaining consistency of these data values across all systems through data synchronization services
- III. Transactional hub** – the hub becomes the central and single master for the creation and maintenance of all enterprise master data. Applications which need to consume master data in their work flows reference this hub on demand

- 4. Data acquisition for Business Intelligence (BI) and analytical purposes** – involves data aggregation from multiple sources, transforming and merging data based on business rules, and publishing data to reporting servers, data warehouses etc, only when the data is “fit” to be published.

SUMMARY

Business processes and their management, governed by the discipline of BPM, have rightly emerged as the key levers for achieving competitive advantage in today's rapidly changing market place. However, the promised benefits of BPM can be seriously compromised if organizations do not pay adequate attention to the state of their data assets that run these workflows and processes.

Cordys MDM, powered by Cordys BOP, is built to support process-based, event-driven lifecycle management of shared master data – the vital ingredient for automating business processes and to build the elusive “Single View of Business”.

CORDYS

The Enterprise Cloud Platform

Cordys is a global cloud platform software provider that helps businesses move to the cloud quickly and with minimal risk. The Cordys platform combines Platform as a Service, enterprise application development, a complete Business Process Management Suite (BPMS) and the next generation of high productivity middleware. Global 2000 enterprises choose this analyst-recognised platform to improve their business operations, deliver better customer service and release additional value from existing IT. Service providers implement the platform to leverage cloud business opportunities, create new high value products and services in the cloud and reduce cost while driving innovation. Headquartered in the Netherlands, Cordys has offices in the Americas, EMEA and Asia-Pacific.

For more information on Cordys please visit <http://www.cordys.com>.