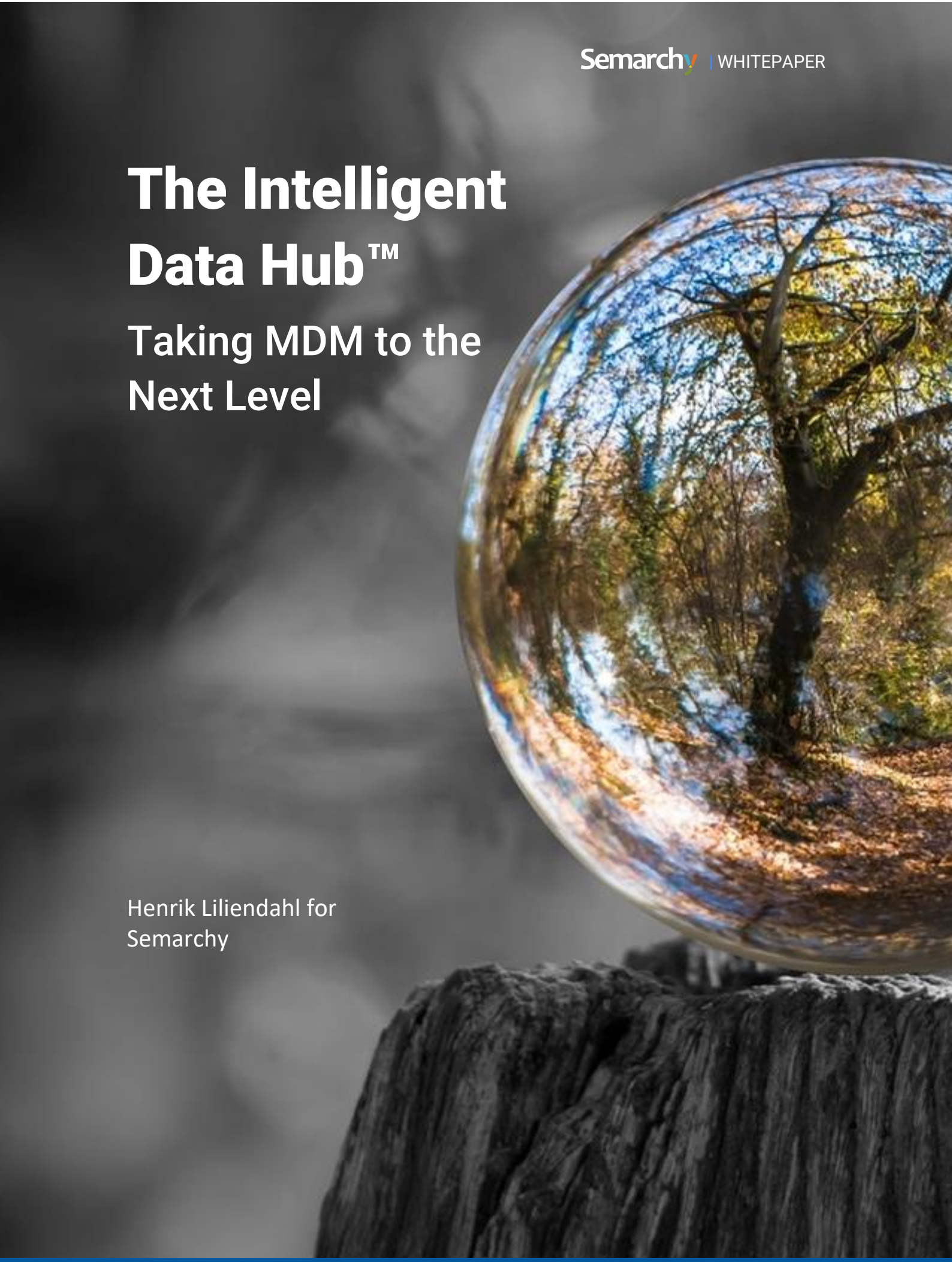
Semarchy | WHITEPAPER

The Intelligent Data Hub™

Taking MDM to the Next Level



Henrik Liliendahl for Semarchy

Abstract

Maturing your organization into a successful digital transformation requires taking your MDM initiative to the next level: The Intelligent Data Hub.

Contents

Introduction

Maturing your organization into a successful digital transformation requires taking your Master Data Management (MDM) initiative to the next level.

While MDM solutions have been instrumental to solving core data quality issues in a traditional way - focusing on simple master data entities such as customer or product - organizations now face new challenges with broader and deeper data requirements to succeed in their digital transformation. The scenarios for consuming, authoring, verifying, enhancing and sharing data within the organization and within the business ecosystems have drastically evolved with an ever-growing complexity.

The Intelligent Data Hub™ is the concept that goes beyond MDM and encompasses an extended data landscape with a richer palette of capabilities.

The Intelligent Data Hub™ is the concept that goes beyond MDM and encompasses an extended data landscape with a richer palette of capabilities.

Semarchy | WHITEPAPER

The Evolution of MDM

There was a need for a place where the data cleansing and data matching results could be stored and maintained.

Master Data Management (MDM) became a discipline within data management back in the early 00's following a preceding awareness around Data Quality Management (DQM). Data quality tools were aimed at distinct data cleansing and data matching tasks around master data. However, it became obvious that there was a need for a place where the data cleansing and data matching results could be stored and maintained in order to avoid doing the cleansing and matching repeatedly.

Single Domain MDM

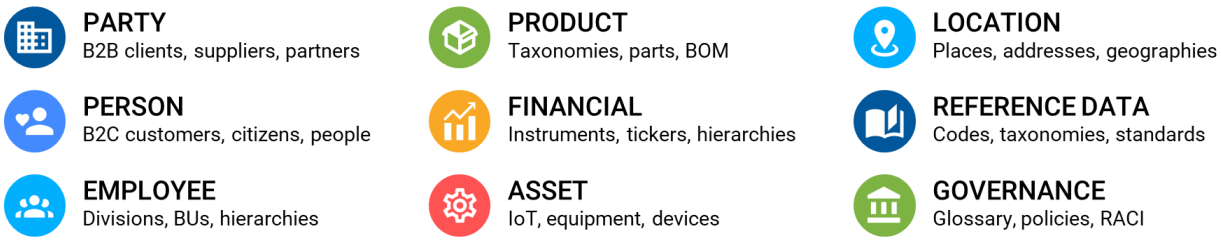
The first MDM solutions typically developed from either:

- Customer Data Integration (CDI) solutions that took care of federating customer records from various applications in the IT landscape and providing data matching and consolidation capabilities for building a duplicate free golden copy of customer records.
- Product Information Management (PIM) solutions used in marketing and sales operations and expanding this view to handling product master data across all functions in an organization.

Multidomain MDM

When implementing an MDM solution today it is common to be able to handle a relevant mix of master data domains. These domains fall into these buckets:

- **Parties:** Natural persons and legal entities that have customer, supplier/vendor, other business partner, employee/contractor and similar roles within your business.
- **Places:** Addresses and geographical positions belonging to the above parties, your plants and shops, delivery points and more.
- **Things:** Product models that your organization produce, sell and use in maintenance, repair and operations as well as each instance of a product, being an asset / IoT device, that needs to be owned, located and valuated. In addition, also intangible "things" such as services, financial securities and more must be covered.



Master data management domains

Master Data Governance

Data governance is the overarching data management discipline that concentrates on the people and matures the processes and capabilities delivered by data quality tools and MDM platforms.

Data governance includes handling organizational structures in data management, data policies and standards as well as business glossaries that define and align the terms around master data and other data used across your organization.

Modern MDM platforms encompass functionality that supports data governance activities required from the data owners, data stewards and data custodians. New platforms allow teams to collaborate in building the governance organization and in managing master data assets.

Master Data Workflow Management

Workflow management has become a core capability in most MDM solutions.

In the early days, MDM solutions were often integrated with a third-party workflow engine. Today, these solutions include workflow management as part of their built-in/native components.

The approval process for onboarding master data entities is the most frequent use case for building a master data workflow. However, we increasingly see other use cases addressed, especially data quality enhancements through collaborative or approval workflows.

For party master data (customer and other roles) some examples of workflows include:

- Duplicate management review and approval
- Golden data survivorship settlement
- Blacklist check
- Credit check and approval

For product master data some examples are:

- Completeness check for various stages (sellable, catalog ready, online ready)
- Compliance check and approval
- Pricing approval

Semarchy | WHITEPAPER

Entering Digital Transformation

The data hub concept is essential in bringing your organization through digital transformation taking the fastest and safest track.

Data is becoming the core corporate asset that determines the success of your business. Digital transformation is on the agenda everywhere. You can only exploit your data assets and do a successful digital transformation if you are able to govern, manage and measure your data.

Digital transformation is built on an expanding data landscape where the Internet of Things (IoT) adds to the opportunities while increasing complexity.

Machine learning and Artificial Intelligence (AI) brings new ways of optimizing data but also increases data quality requirements. The data lake concept has many advantages, but it is still perceived as another data silo.

The data hub concept is essential in bringing your organization through digital transformation taking the fastest and safest track.

The Expanding Data Landscape and the Data Hub Concept

Master data exists in symbiosis with other kinds of data. These are in general:

Reference data

Smaller lists of values that give context to master data and/or transactional data and ensures that all stakeholders use the same (or linkable) codes for describing data records. For example:

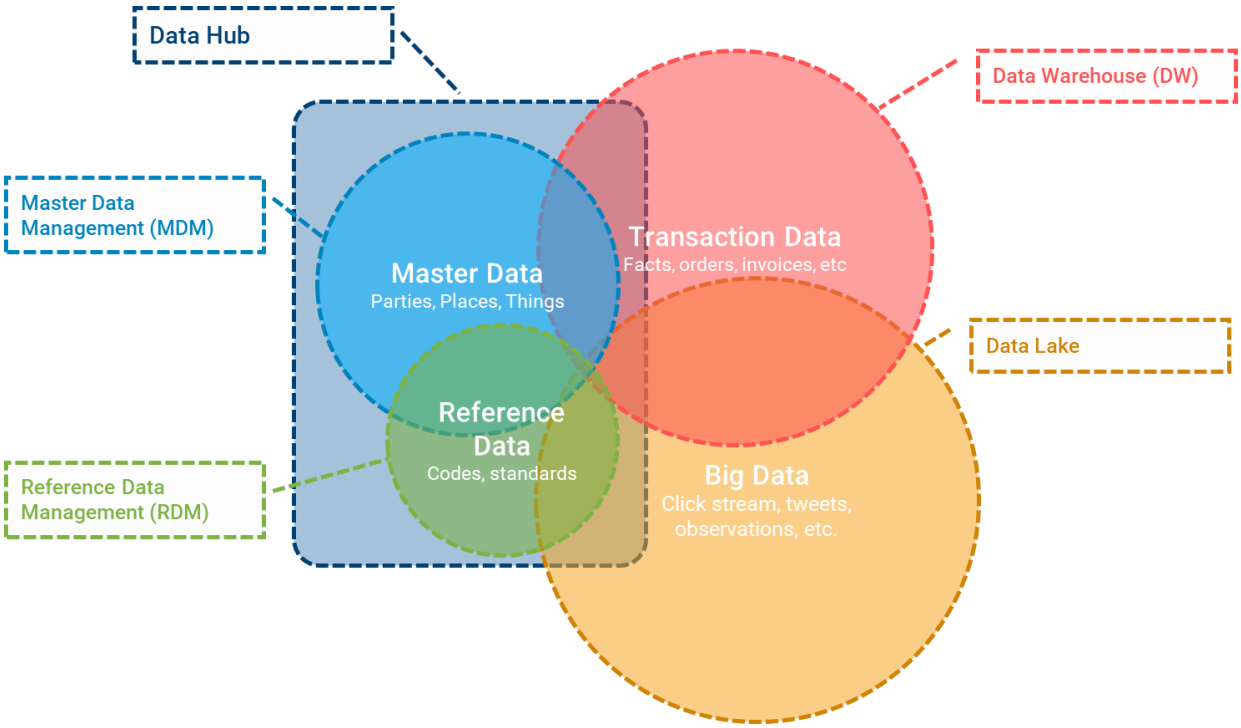
- A list of countries, a list of states, provinces or postal codes in given countries
- A list of industry sector codes
- A list of product classifications

Transaction data

Data about events such as ordering, invoicing, payments, deliveries and more. Transactions refer to parties, places and things described as master data.

Big data

Another kind of data that comes in a larger volume, in an increased variety and in an accelerated velocity. Such data is produced by sources such as social media data streams, sensor data and weblogs, but also good old transactional data stored in new scalable data stores.



The Data Landscape

We now see the data warehouse concept being supplemented by the data lake concept, where data warehouses are mainly used to analyze transaction data and data lakes are used when adding big data in lesser structured forms to the list of data sources.

The role of Reference Data Management (RDM) has increased with the introduction of more data sources that have to be combined in the data lake. In the same time Application Data management (ADM) has become an additional concept to MDM, as sharing of classic master data entities is extended to other critical data elements held in specific applications such as ERP, CRM, SCM (Supply Chain Management) and more.

The data hub concept brings all the operational data use cases above together with the aim of governing, managing and measuring data as a corporate asset. It can also consume aggregated metrics from the analytical stores – the data warehouse and the data lake – but does not necessarily store or manage transactional and observational data.

The Rise of the Internet of Things (IoT)

The potential in the Internet of Things (IoT) is enormous. We will increasingly use smart devices that are connectable in our daily life. Perhaps even more significant manufacturers will use smart machines that are connectable within the Industrial Internet of Things (IIoT)– sometimes also referred to as Industry 4.0.

Data management practitioners have traditionally handled a product model like a certain model of a refrigerator and a certain model of a drilling machine. When each produced instance of that product model becomes intelligent, we will see requirements and opportunities for handling each instance – each thing or asset - as a master data entity.

In a traditional master data initiative, one must handle relations between customers and products. These can for example include what product models a given customer bought, what locations a customer (and other party roles) visited and in which locations a given product model is available.

With IoT, the number of relations will increase drastically. A “thing” will have relations to many party roles as who is the manufacturer, who is the operator, who is the maintainer and who is the owner. If the thing is a face recognition device, it will also generate new relations to every single human – or animal – it has recognized! Things will shift locations; therefore, enterprises need to be more precise about each location. Products will have many things with varying configurations produced leading to increased potential relations and data streams.

You need a data hub, that goes beyond traditional MDM, to encompass this complexity.

With IoT, the number of relations will increase drastically.

Utilizing Machine Learning and Artificial Intelligence (AI)

Machine Learning (ML) is the discipline used to ignite Artificial Intelligence (AI). While you support the machine with training datasets for ML it may be tempting to do a little data cleansing as the training data will not be part of the continuous future AI processing.

The drawback of this approach is that you will probably cleanse each training dataset slightly differently. This means, that when more AI supported business processes start to interlink, they will have a different “way of thinking”. You will then have the same situation as when humans do not work very well together.

Solving this challenge relies on training datasets that are derived from rationalized and trusted production data having the same master data, reference data and application data foundation.

Having your AI supported business processes running on top of master data, reference data and application data that is unique, accurate, consistent and timely will make a huge positive difference for the business outcomes achieved from applying AI. The results will be reliable. The processes will be repeatable over time. The concept will be reusable in other scenarios.

Coexistence of the Data Hub and the Data Lake

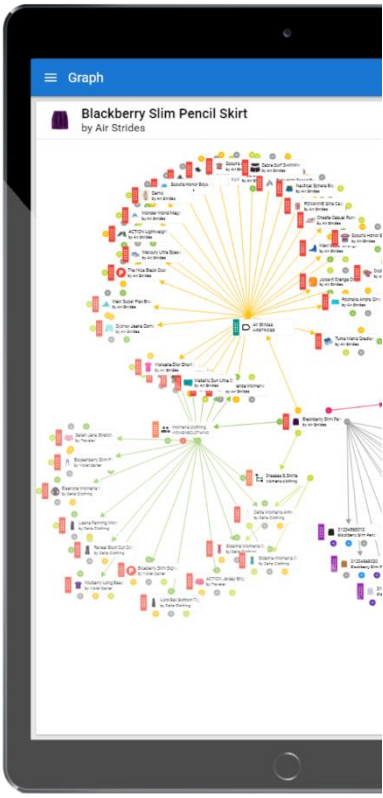
The data hub covering master data, reference data, critical application data and armed with intelligent capabilities will amplify the advantages of the data lake concept and counterwork the trend that have marred the data lake idea in its early stages, comparing it to a data swamp.

By ensuring the quality of the operational data held by your organization, the data lake and the data warehouse can utilize well defined references for transactional and behavioral data streams used in advanced analysis, predictions and prescriptions.

Semarchy | WHITEPAPER

Components of the Intelligent Data Hub™

A comprehensive set of key features are required to ensure flexibility in the implementation styles, and to support a variety of business models.



At the core the Intelligent Data Hub has a comprehensive and flexible data model with rules, policies, and workflows specific to your data domains, implementation style and business model. The data managed in the hub can be authored and governed by the business stakeholders or delivered by and consumed by applications in the IT landscape through a range of data integration capabilities. The data can be visualized in hierarchies and graph models.

Map the Data Landscape as It Is with Data Discovery

Data discovery serves as the weapon used when exploring the current data landscape at your organization with the aim of building a data hub that reflects your data model and data portfolio as it is.

One of the key capabilities in data discovery is data profiling. The frequency and distribution of data values is counted on relevant structural levels. Data profiling can also be used to discover the keys that relate data entities across different databases.

Furthermore you can use data discovery to map data lineage, find potential data relationships where data matching, data cleansing and/or data stewardship might help with ensuring data quality and business process improvement and explore where the same data have different labels (metadata) attached or the same labels are used for different data types.

With the increasing awareness of data security, data protection and data privacy – and the regulatory compliance enforced in this space – it is also crucial for your organization to know what kind of data flows between applications and where instances of this data are stored within the organization.

With the help of data discovery, you will build the data catalog that serves as the cornerstone in the data hub and adjacent data lake.

Define and Enforce the To Be State Through Collaborative Data Governance

Data champions, in the roles of data owners / sponsors, data stewards and data custodians / data operators must have ways to define the data policies, data standards, business rules and business terms in the Intelligent Data Hub throughout the course of the digital transformation.

Data champions must also be able to involve all business users in setting up the processes, constraints and permissions required in a continuous improvement of master data,

reference data and critical application data and hereby enforcing the desired direction towards the to be state of the data portfolio with the optimal data quality.

In collaborative data governance you will be able to strike the right balance between rigidity and agility. This includes democratizing core data concepts such as having a 360-degree view of parties, places, things and reference data.

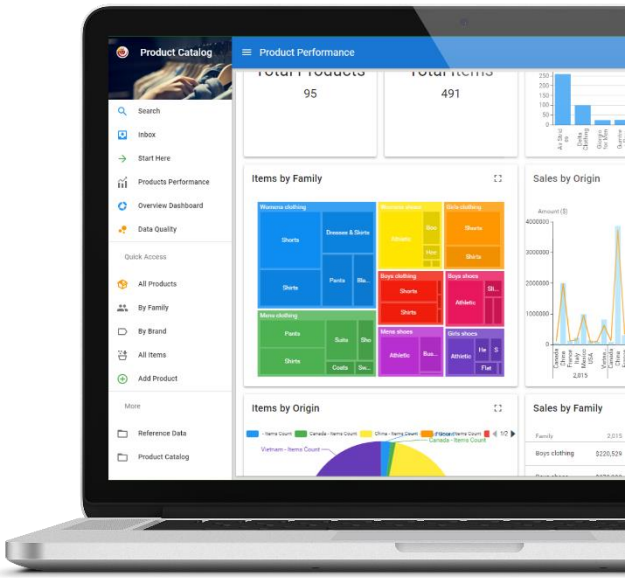
In collaborative data governance you will be able to strike the right balance between rigidity and agility.

Measure Results with Data Dashboards

Data analytics, business intelligence and data visualization are disciplines that have often been thought of as measuring corporate performance "after the facts". In a global collaborative data governance initiative, data champions often have requirements for measuring key performance indicators that can influence how they properly manage their core data assets in their daily operations.

As such they need to be able to perform real-time analysis and data visualization of contextualized subsets of data reflecting both the data managed in the hub and outside of the hub such as the data lake or the data warehouse.

The typical requirements to enhance the data quality and usage of the data vary. Champions can choose to design charts, dashboards and applications that measure the quality of the data and efficiency of the processes managed in the data hub. They can also decide to combine these metrics with data from operational or analytical sources.



The business impact of changing rules or even re-classifying some data elements within the data hub can be significant. Measuring such an impact with pre-built KPIs and visualizing these KPIs in real-time within the data management applications can drastically reduce non-compliance risks while optimizing the operational costs due to misinterpretation of data. To put it in context, a data champion will probably take more caution in editing a product description if he or she knows that this product is consumed by 30 downstream applications, that it is in the top 5 selling products and that it has been recently modified by the marketing team for Search Engine Optimization (SEO). The entire context of the product about to be edited can only be obtained through efficient

dashboards, blending together data from the hub and from a significant amount of other external data sources.

Data health monitoring, real-time data insights, data compliance checks, and operational processes optimization are ideal use cases justifying the use of solid real-time dashboarding features within the Intelligent Data Hub.

Semarchy | WHITEPAPER

How xDM Underpins your Digital Transformation

Using a single software platform for centralizing core critical data and processes is the accelerator of the Intelligent Data Hub™ initiative.

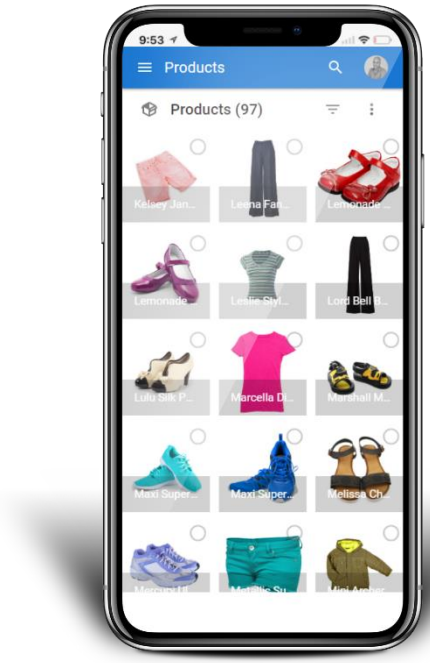
While everyone agrees that mastering master data is instrumental in digital transformation, there are other kinds of data too. Building a 360-degree view of customers, suppliers, business partners, locations, product models, devices and other assets takes well managed master data seen in conjunction with reference data and other critical application data as well as interactions and observations held in transactions and big data streams.

xDM from Semarchy enables this quest by offering the Intelligent Data Hub™.

Value Driven Data Management

Many data management and data governance initiatives have been challenged with delivering tangible business outcomes that justify the spending. They often take too long and require large budgets with questionable results. With the agile nature of xDM, you can implement your data management and governance maturity step by step, harvesting measurable results from the start.

Additionally, you will be able to adjust and prioritize the efforts as you progress on the learning curve and as your business environment evolves. xDM will help you cope with the rapidly changing ecosystem while your customers, suppliers and other business partners are also undergoing a digital transformation.



Reference Data Management (RDM)

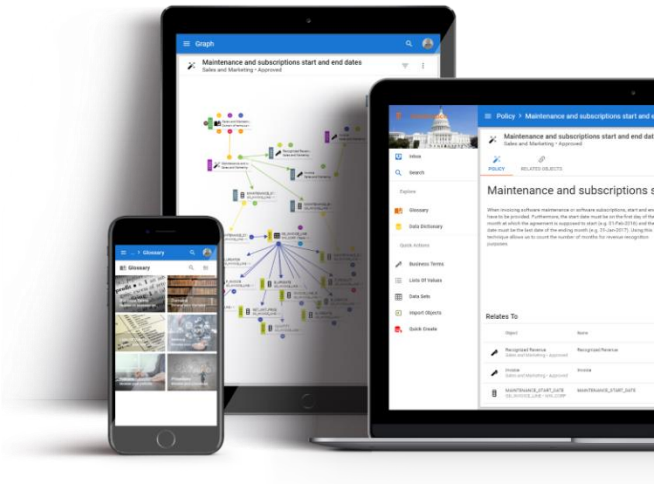
Discovering, governing and measuring your reference data is a perfect way to start a data hub initiative. Though one should not underestimate the challenges implied by the small size of reference datasets, they are usually easier to understand across the organization and the technical hindering within data integration can be solved in simple ways.

By utilizing the flexible data modelling capabilities in xDM you will get quick wins by having cross organizational reference data managed in the Intelligent Data Hub. The wins will cover uniform reporting leading to better business decisions and more efficient interaction with customers, suppliers and business partners in the business ecosystems where your organization operates.

Flexible data modeling of xDM allows for uniform reporting of cross organizational reference data.

Application Data Management (ADM)

ADM shares a lot of requirements with MDM and RDM around data governance and workflow management. However, within ADM you need to mirror the data models and chosen configuration for the ERP, CRM, SCM and other applications that are the authoritative source of this data.

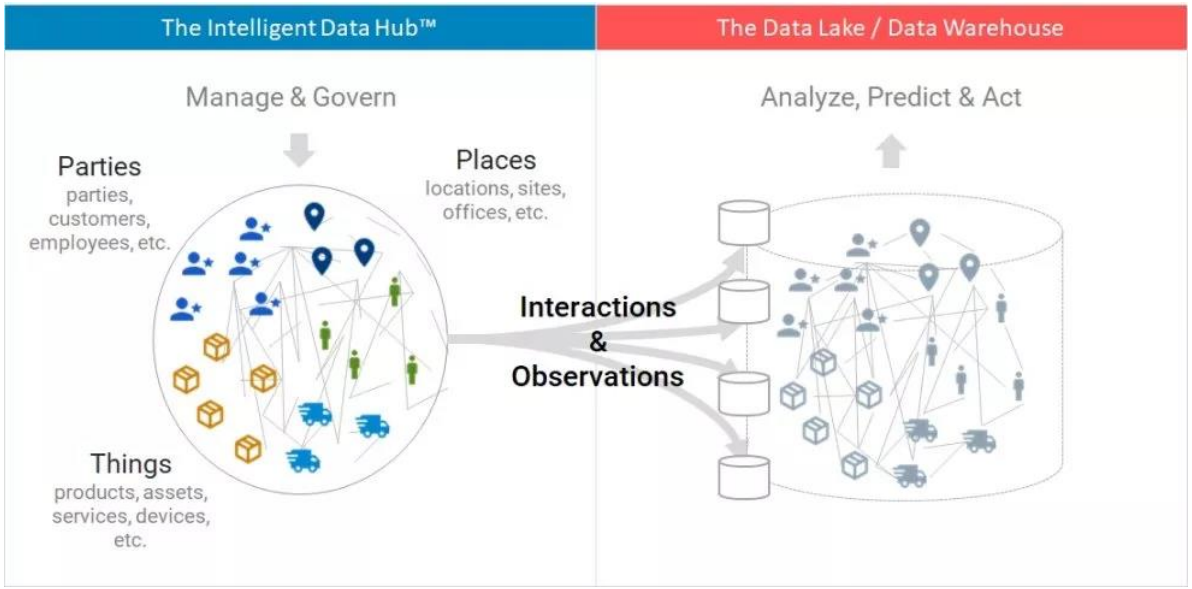


With xDM you do not need to have separate data governance and workflow management in place as the hub nature of xDM allows for the same capabilities to be applied to master data, reference data and the critical application data in whatever data model they are.

A Non-Invasive and Intelligent Data Hub

xDM integrates into your IT landscape in a non-invasive way meaning that you do not have to change your existing application portfolio to measure the incremental benefits of the Intelligent Data Hub.

With the Intelligent Data Hub, operational analytics, predictive and prescriptive analytics as well as traditional analytics share the same master data, reference data, business glossary and data catalogs to transform the interactions and observations into measurable business value.



The Intelligent Data Hub connected to the Data Lake

About

Henrik Liliendahl

Henrik Liliendahl is a seasoned data management consultant with a long track record from both the vendor and user side within data quality management, multidomain master data management and data governance. Henrik is also very active in the data management community on social media including running a popular blog. Henrik's latest venture is a cloud service for sharing product information between parties in business ecosystems.

Semarchy

Semarchy is the Intelligent Data Hub™ company. Its xDM platform empowers organizations of any size to build data applications quickly, with fast time to value. A single software platform for governance, master, reference and application data management as well as data quality, enrichment, and workflows means that trusted information projects can be delivered without the wasted time of cobbling together bits of software, interfaces, and integration layers.

Published in September 2019.