

IT@Intel: Master Data – Managed!

Intel IT’s master data strategy goes beyond selecting a master data management (MDM) solution to include key elements that will drive business efficacy

Author

Candice Linnell
Director, Core Data Solutions
Intel IT

Table of Contents

Executive Summary	1
Background.....	2
MDM at Intel.....	2
Overview of Intel IT’s MDM Strategy .	3
Solution Architecture.....	6
Results	6
Next Steps.....	7
Conclusion	7
Related Content.....	7

Executive Summary

Master data is defined by Gartner as “the consistent and uniform set of identifiers and extended attributes that describes the core entities of the enterprise.”¹ Intel uses vast amounts of master data every day for many reasons, spanning transactions, reporting and advanced analytics. But often, Intel’s master data exists in silos, presenting a fractured view of reality between business units. This can result in decisions being made on bad data, lengthy attempts to clean up the data or can even negatively impact customer experience.

Intel IT is embarking on a journey that, when complete, will result in all data—including master data—being treated as a highly valuable corporate asset across the enterprise. Because it is Intel’s most highly shared data, we determined that Intel’s master data is the ideal place to start our journey.

One thing we have learned so far is that the road to trusted master data, made available to all employees, involves far more than simply choosing a data management solution. This paper describes our master data strategy and its key elements.

Background

Each of Intel’s business units depend on master data to make informed business decisions. However, we currently have a mix of clean and inaccurate master data, and it often exists in silos. Some of it is shared across a couple business units, but none of it is common across the entire company. Because business units need master data to perform their tasks, they take various approaches to dealing with the lack of consistency:

- Augment the available master data in an attempt to make it more accurate, but very few of these data fixes are ever incorporated into the actual enterprise sources of the data.
- Ignore the enterprise source and create their own source.
- Try to remain synchronized with the enterprise version of the data but struggle to find the data because it is in various places and it all looks slightly different, so projects take longer to execute.

To illustrate these scenarios, imagine a customer list being maintained by two separate business units, each unaware of the other’s updates. Or what if Finance and Manufacturing have differing views of supplier data? This siloed approach to data can hinder business decisions and can sometimes have catastrophic consequences. A customer may fall through the cracks, leaving a potential deal on the table. Or the wrong product could be shipped to a factory, delaying product delivery.

A better approach is to have a single version of truth that is available to all business units and employees. In this ideal scenario, data is managed as a valuable corporate asset, with all the strategy and governance that befits it—just like other assets such as intellectual property and financial assets. Intel IT is committed to creating that scenario, and is working with the business units to establish “uniformity, accuracy, stewardship, semantic consistency and accountability of the enterprise’s official shared master data assets.”²

Disadvantages of Data Silos

Everyone knows data silos are bad—but here are some tangible reasons to manage master data to avoid them:

- They give an incomplete view of the business.
- They create a less collaborative environment.
- They lead to poor customer experience.
- They slow Intel’s business pace.
- They waste storage space.
- They make data less accurate.
- They prevent the connection of data across business lines.
- They reduce automation opportunities.

Master Data Management (MDM) at Intel

Three common types of data exist at Intel:

- **Transactional data** is generated by various applications while running or supporting everyday business processes.
- **Analytical data** is created through calculations or analyses run on transactional and master data.
- **Master data** provides context to business transactions. It represents the actual, critical business objects upon which said transactions are performed, while also taking into account the parameters on which data analysis is conducted.

“We are successful when sources of master data are consistently creating accurate data and master data is democratized where every app, report and algorithm across Intel is getting its master data from these sources.”

–Candice Linnell
Director, Core Data Solutions, Intel IT

Master Data Characteristics

Master data, which is edited less often than other types of data, is structured with the enterprise in mind. Master data is intended to work for all business units, not just a few, which makes it a highly shared, reusable data asset. At Intel, we currently have eight master data domains that include Item/Bill of Materials, Customer, Supplier, Finance, Location, Calendar, Worker and Commodity.

These terms are important to understanding our master data strategy:

- **Record of origin (ROO).** For master data, this is typically an MDM solution. This is where the data gets created with as much accuracy as possible. A ROO can also be referred to as an authorized source.
- **Record of reference (ROR).** A data warehouse, data hub or data lake that serves as a single source of truth for the master data domain. Creating and using a trusted ROR—one that pulls its data from the ROO—is key to establishing master data consistency across the enterprise.

Overview of Intel IT’s MDM Strategy

Our master data strategy is designed to provide the highest quality master data to the enterprise in a way that will flex and scale as the company evolves. All the pieces of this strategy mesh to make high-quality master data sustainable for the long term (see Figure 1):

1. Establishing Data Ownership
2. Defining a Master Data Policy
3. Creating a Master Data Glossary
4. Defining an LDM
5. Selecting an MDM
6. Integrating an MDM with Third-Party Data
7. Building a Trusted ROR
8. Deploying Reusable APIs
9. Governing Adoption of the Master Data Strategy

Step 1: Establishing Data Ownership

It is imperative to assign data ownership responsibilities for each master data domain. For example, Intel’s Supply Chain business unit “owns” the Supplier master data domain, and they are accountable for the overall health of this data for the entire company.

Data ownership formalizes the role of data owners to help manage data from creation to consumption. It puts people and processes in place to ensure the right people define usage guidelines, set quality standards and resolve data issues. Data owners ensure the data has the right controls in place and is effectively supporting the company.

Step 2: Defining a Master Data Policy

A vital component of our master data strategy is defining a corporate policy focused on master data. We intend to share this established policy with all employees so that they know what can and cannot be done with this corporate data asset.

The master data policy describes the rules that apply to the management and consumption (aka use) of master data. We have tailored the policy to conform with Intel’s company culture and made sure that it addresses the existing concerns about master data health. We plan to conduct annual updates so the policy can adapt to the evolution of both Intel as a whole and its data.

Our master data policy covers the following topics:

- Data Ownership, Stewardship and Custodianship
 - Data ownership and stewardship is the purview of a business unit
 - Data custodianship is the responsibility of Intel IT
- Rules for systems that create or alter master data
- Rules for systems that reference master data
- Oversight of systems that create, alter or reference master data
- Oversight of systems that consume master data, to help ensure they consume it from an authorized source

Step 3: Creating a Master Data Glossary

Once data ownership is established, the owning business unit assigns subject matter experts to create the terms and definitions for the data domain. Creating a master data glossary that includes terms and definitions for all defined master data attributes enables us to have agreed-upon terms that support the whole company, not just a few business units.

To help ensure enterprise-wide adoption, the owning business unit must verify the terms and definitions meet the needs of Intel’s major business units that depend on this data for their transactions, reports and/or analytics. This is often accomplished by representation of all major business units in the local data governing body for the specific master data domain.

Terms and definitions are bound to change as Intel evolves. We want to treat this kind of change as a production release, where all downstream uses of the old term are assessed for change impact and communicated and coordinated as such. The same process is to be used when introducing a new term.

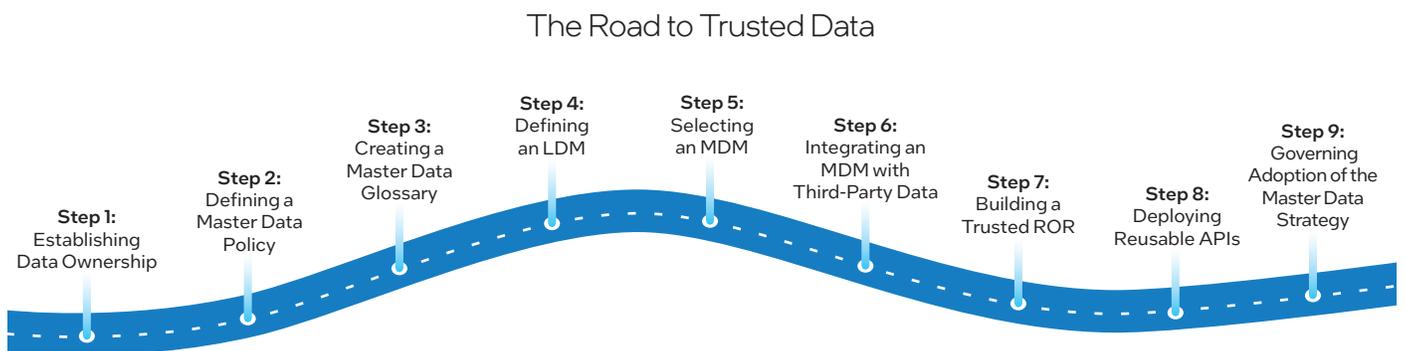


Figure 1. Our master data management (MDM) strategy.

Example Master Data Terms and Definitions

Definitions can be as simple or as complex as necessary.

Here are two example master data terms and their definitions from our current master data glossary:

Business Person Name. The text by which individuals who are Business Persons (external to Intel) are known. A Business Person is an individual that provides and/or receives goods and/or services to or from Intel. They may or may not be part of a Business Organization.

Generally, they are thought of as anyone who is not an Intel Worker. The names for a Business Person in the Western culture are first name, middle name and last name. In the Greater Asian Region (GAR), the names of a person are not ordered in the same manner as in Western culture (first, middle, last). Taiwanese, Chinese and Korean cultures are all examples of where the names are ordered differently than in the Western culture. Intel is pursuing an international approach to names where given name, middle name and family name will be used.

Company Code. A numeric code that uniquely identifies a specific legal entity for Intel.

We have found that the more closely the ROO for each master data domain matches the master data domain’s LDM, the less data model translation is required in the ROR. Sometimes, exact alignment of the data management solution with the LDM is not possible. For that reason, we established an enterprise LDM layer in our ROR. This layer is where any data coming from the physical data model in the ROO that does not align to the LDM is translated to align to the LDM.

Aligning an LDM with the master data glossary is the catalyst to remove and prevent the use of data silos across the company. Although we have defined LDMs for each master domain, they will change over time as Intel evolves its business.

Steps 5 and 6: Selecting and Configuring an MDM Solution then Integrating it with Third-Party Data

Over the last decade, the MDM industry has evolved and matured. A good MDM solution makes stewarding master data far easier than manual methods. Currently, there are several MDM solutions in use at Intel, all of which vary in capabilities and ease of use. It is our goal to consolidate and modernize all our MDM capabilities.

When selecting an MDM solution, ease of use is the first thing we consider. The business units need a tool that is intuitive as they steward their master data for accuracy for the enterprise. The second thing we look for is a robust business rules engine. The more business rules that can be automated, the less data errors occur due to human error. And finally, the last major thing we look for in an MDM solution is integration with industry third-party data providers.

Whichever MDM solution is in use, it is important to configure it so that it aligns to the enterprise LDM. This allows us to set up the master data records completely and accurately so that all downstream use of the data can trust it. Transactions fail when the master data is wrong. Teams cannot trust their reports and analytics if they are running on incorrect or incomplete master data.

Other best practices we follow when configuring an MDM solution include:

- Automating the business rules defined in the data model.
- Setting up the workflow so the master data has at least one approval step, which will maintain control over the data.

Step 4: Defining an LDM

With master data ownership, data glossary and policy in place, we can define an enterprise LDM for each master data domain.³ An LDM (see Figure 2) provides the specifications for data that describe the concepts, relationships and interpretation of the values of data. It is a logical data model because it does not define the physical structures in which the data may be stored or transmitted. We provide LDM documentation so that teams implementing trusted sources for master data can use it as a reference for their solution designs.

Many data management solutions are configurable. We encourage solution owners to align their data management solution as closely as possible to the appropriate LDM.

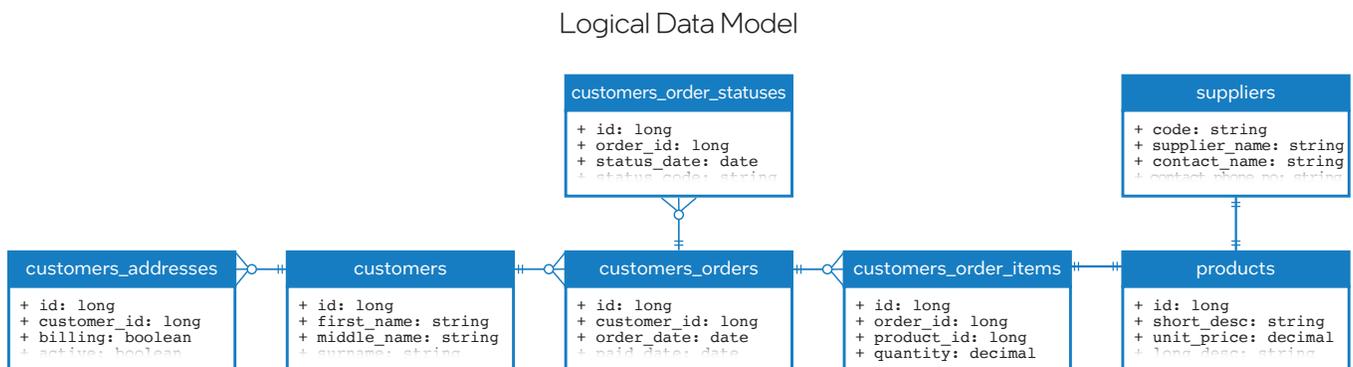


Figure 2. A Logical Data Model (LDM) provides the specifications for data that describe the concepts, relationships and interpretation of the values of data.

- Integrating the MDM solution with industry third-party data, which helps ensure accuracy of the master data and helps prevent “dirty data.” By using the services of third parties, whose business it is to maintain accurate data on a global scale, we bring that level of precision into our organization. Examples of third parties that we use include the International Organization for Standardization (ISO) for country data and global languages, as well as an industry leader that provides us with the legal name, address and entities for Intel’s global business partners.
- Archiving capabilities, which keep the dataset sizes manageable.

Step 7: Building a Trusted ROR

A ROR earns the title of “trusted” when it aligns to the enterprise LDM, follows best practice data rules (such as using standard naming conventions), has consistent support in production and uses master data that comes only from an authorized source, such as the MDM solution.

Aligning a ROR to the enterprise LDM model is crucial. Various sources of data may have different physical data models. For example, two sources may label the same value differently. By translating both sources into labels that align to the enterprise LDM within the ROR, we align the company to business-defined terms and definitions (see Figure 3). Imagine if half the company pulled master data from source 1 in the figure below and the other half from source 2. How would the two halves of the company connect their data?

A Lesson Learned

For many years, Intel did not establish business unit stewardship of customer master data. Instead, everyone (Intel IT and the business units) assumed our expensive, best-in-class master data management (MDM) solution would magically cleanse, complete and fix our customer records. We avoided implementing comprehensive business rules on the MDM solution, because we wanted it to accept the data as-is, much of which had very little structure.

Despite our best intentions, unforeseen chaos occurred. No MDM solution can match and merge records without some amount of structure. As a result, this bad dataset grew to huge proportions (more than 13 million records), while less than 20% of the records were able to be matched to industry third-party legal name and addresses.

Applying master data principles to that dataset is enabling us to reduce the dataset to a more accurate number of 200,000 records, with far greater accuracy.

The abstraction of the physical data models in the sources makes the replacement of those sources less disruptive to downstream consumers, and helps prevent data silos within the master data. We apply the same best practices to each enterprise data warehouse that we create for master data domains:

- To create a high-quality ROR, we work to ensure that our enterprise data containers (data warehouses, data hubs and data lakes) pull data only from the best source(s) possible—the authorized sources, or ROOs—as well as align the data to the appropriate LDM.
- We establish controls so that data is never changed in the ROR, but rather it is fixed or updated at the authorized source(s).
- One ROR never pulls master data from another ROR, which would introduce an extra data hop and cause data to get out of sync.

These best practices help us to keep the data flow clean and maintain control of our master data.

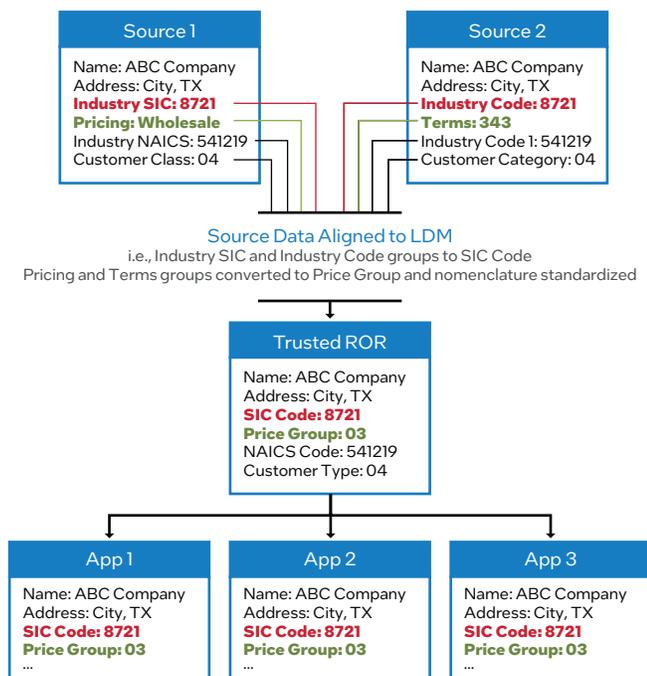


Figure 3. A trusted ROR is where we align various sources of data to the enterprise LDM; apps pull data from the ROR, which acts as the single source of truth.

Step 8: Deploying Reusable APIs

Application programming interfaces (APIs) are pieces of code that enable applications to communicate with each other, or enable end users to control the behavior of an application.

In the context of managing master data, APIs are how downstream applications access master data. Because there are many applications used at Intel, having each application build its own APIs would result in massive API sprawl. In contrast, we have found that deploying reusable APIs for highly shared data, like master data, is most efficient; for example, we can use the same API for twenty applications that all need the same master data as opposed to each application using its own interface. We also use a tool that catalogs all the master data-related APIs and labels them as “trusted” so users know that the data provided by the API will be accurate. Using such a cataloging tool makes APIs easy to find and use. To date, we have built more than 300 reusable APIs for our master data.

It is our experience that APIs should always pull data from a trusted ROR, not from the authorized ROO. This approach avoids potentially overwhelming the ROO, which may not be built to support thousands of data requests at once. We build our RORs to be resilient, so that they can serve up data whenever anyone at Intel requires it. Also, we manage the master data APIs like any production software at Intel. Any time the LDM changes or a trusted ROR is enhanced, we test and version each API.

Step 9: Governing Adoption of the Master Data Strategy

Simply creating a master data policy and some training materials, then hoping everyone will comply is not enough to achieve success. Our goal is to reap the best business value from our efforts to manage master data so that all of Intel has access to comprehensive and accurate master data. To accomplish this, we are establishing project governance, which can help ensure everybody is using master data correctly and consuming master data from an authorized source. Enterprise-wide adoption of master data unleashes automation and improved insights because now all of the data surrounding master data can be connected across disparate systems and business processes.

Solution Architecture

Large companies typically have more than one enterprise data warehouse and/or data lake, and Intel is no different. As mentioned previously, we avoid having one ROR feeding another ROR—primarily because it introduces an extra data hop where mis-translation can easily occur. The more hops data takes downstream from the source, the more potential there is for that data to become out of sync with the source. This is why we apply the same architectural principle to all enterprise data warehouses, which is to have the ROR obtain the master data directly from the authorized ROO. This provides us with the most control of our master data, helping to ensure that the same data is made available to

all corners of the company while continuing to reap the benefits of some data warehouses, such as those with in-memory database capabilities that provide near-real-time access to the data. The farther that in-memory database is from the source, the longer it takes to get data updates.

Our ROOs and trusted RORs are built using Intel® processors, solid state drives and Ethernet products. As shown in Figure 4, we make master data available everywhere the company needs it by having trusted RORs both on-premises and in the cloud.

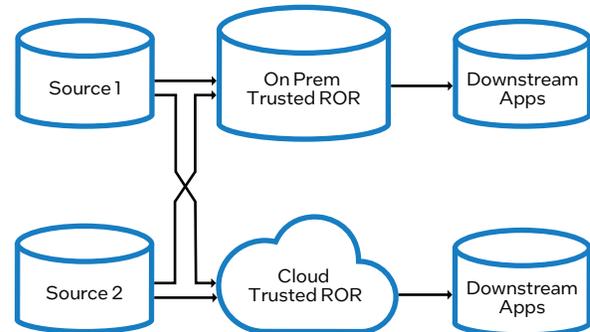


Figure 4. To increase our reach of trusted master data, we are creating trusted records of reference (RORs) on-premises as well as in the cloud.

Results

Properly mastered data is important to Intel’s future success, which relies on data to fuel growth and transformation. Bad master data can disrupt reporting, analytics, dashboards and even machine learning model accuracy. We expect that our efforts to create clean, accurate master data across Intel will result in higher business efficiency and effectiveness, as shown in Figure 5.

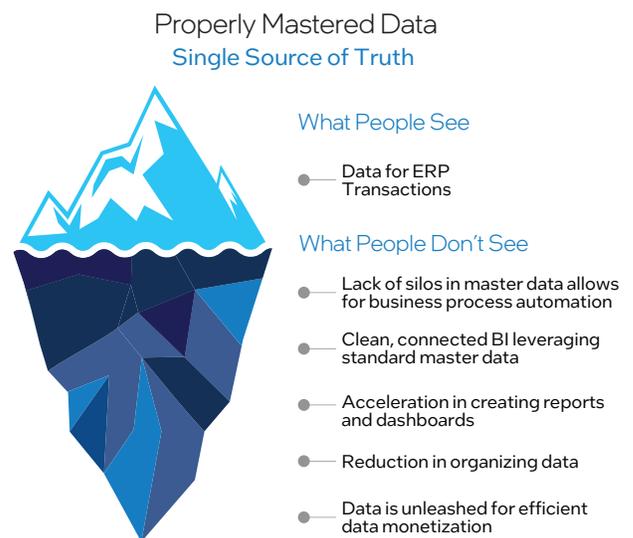


Figure 5. The benefits of properly mastered data go far beyond just getting data for enterprise resource planning.

Next Steps

Moving forward, we plan to complete the modernization and consolidation of our master data ROOs along with carrying out the setup of trusted RORs for our eight master data domains. Our enterprise LDMs are in good shape and will continue to change as Intel's business evolves. We will mature the oversight and governance of master data across Intel to help ensure that we keep this data asset accurate and consistent.

In addition to the above activities, we are broadening our scope to focus on making our trusted master data easy to find. It is already easy to consume by reusing APIs and trusted RORs; now we want to make data easy to find by tagging it as trusted in our enterprise data glossary.

Conclusion

We have made great strides toward providing one version of the truth, making accurate and consistent master data available to all of Intel. This master data connects Intel across transactions, reports, analytics and advanced analytics. There is more work to do, and we will undoubtedly revise our LDMs, modify APIs, get new data sources and hone our governance policies. But we are well on our way to creating reliable sources of master data and democratizing this corporate asset so every app, report and algorithm across Intel is getting its master data from authorized ROOs.

Related Content

If you liked this paper, you may also be interested in these related stories:

- [2021-2022 Intel IT Annual Performance Report](#)
- [Enterprise Technical Debt Strategy and Framework white paper](#)
- [Enterprise Architecture: Enabling Digital Transformation at Intel white paper](#)

For more information on Intel IT best practices, visit intel.com/IT.

IT@Intel

We connect IT professionals with their IT peers inside Intel. Our IT department solves some of today's most demanding and complex technology issues, and we want to share these lessons directly with our fellow IT professionals in an open peer-to-peer forum.

Our goal is simple: improve efficiency throughout the organization and enhance the business value of IT investments.

Follow us and join the conversation on [Twitter](#) or [LinkedIn](#). Visit us today at intel.com/IT if you would like to learn more.

Acronyms

API	application programming interface
LDM	logical data model
MDM	master data management
ROO	record of origin
ROR	record of reference



¹ Gartner, "Master Data Management (MDM)," gartner.com/en/information-technology/glossary/master-data-management-mdm

² Ibid.

³ ScienceDirect, "Logical Data Model" sciencedirect.com/topics/computer-science/logical-data-model