Reliable Context

RELIABLE CONTEXT / RESEARCH & PRACTICAL GUIDANCE

Master Data Management: What a Hub Holds and Cannot Prove

> Reliable Context is a kernel of truth: a system of record for organizations that fuse private and public information and must defend every value they act on. One canonical record per real thing, a cited source and an observation date behind every value, unknowns declared rather than left blank, and any past state of the record reproducible.

Published by Reliable Context · 2026-10-29 · master data management · what is mdm · mdm vs data governance · master data management examples · mdm for compliance

Original article: https://reliablecontext.com/articles/master-data-management-explained-what-a-hub-cannot-prove



In this article

1. Definition and scope
2. Master vs reference vs transactional data
3. Domains: party, product, location
4. Identifiers and matching
5. Stewardship and governance
6. Hub styles and where the record lives
7. What a hub rarely records: source, date, dissent
8. Regulatory reasons to want more
9. Selection questions
10. Sources

Definition and scope

Master data management (MDM) is the set of practices, roles and systems an organization uses to keep one consistent, shared description of the business entities that many processes depend on: customers, suppliers, products, sites and accounts. IBM describes it as "a comprehensive approach to managing an organization's critical data across the enterprise" (IBM, What is master data management). A peer-reviewed review in IJITEE describes MDM as a way of maintaining, integrating and harmonizing master data so that information stays consistent across applications and divisions (Hikmawati, Santosa and Hidayah, 2021).

The main reference for the discipline is DAMA-DMBOK, which DAMA International describes as a vendor-neutral, consensus-based view of generally accepted knowledge in data management. It sets out principles and practices but does not prescribe tools (DAMA International, DAMA-DMBOK). That distinction matters when people ask about "MDM vs data governance". Data governance decides who owns a definition, who may change it and how disputes are settled. MDM is one of the main places where those decisions are applied to records.

This article covers what an MDM hub (the central system that stores or indexes the mastered records) usually does well. It then looks at one question a hub is rarely built to answer: which source asserted the value now shown, as of when, and which sources disagreed. The article is educational, not legal advice.

Master vs reference vs transactional data

No freely readable international standard text was available for this article, so the working definitions below come from a vendor and an analyst firm and are labelled as such.

Category	Working definition	Examples	Typical change rate
Master data	"Core data describing key business entities, such as customers, products and locations" (IBM, vendor)	Customer, supplier, material, site, legal entity	Slow, but each change matters to many processes

Category	Working definition	Examples	Typical change rate
Reference data	"Data used to classify or categorize other data, such as country and currency codes" (IBM, vendor)	ISO country codes, currency codes, industry classifications	Rare, often set by external bodies
Transactional data	"Business events and transactions, such as sales, invoices and claims" (IBM, vendor)	Orders, invoices, payments, shipments	Continuous, high volume

Source for all three definitions: IBM. Gartner, an analyst firm, defines master data as the least number of consistent and uniform sets of identifiers and extended attributes that uniquely describe the core entities of an enterprise and are used across multiple business processes (Gartner, Master Data Management). The practical test follows from that wording. A field is master data when several processes must agree on it. An invoice line is transactional. The customer it is billed to is master data. The currency code on the line is reference data.

Domains: party, product, location

MDM programs are usually scoped by domain, meaning a family of entity types that share identifiers and stewards.

- **Party**: legal entities and people in their roles as customer, supplier, employee, counterparty or prescriber. The usual difficulty is corporate hierarchy (who owns whom) and the fact that the same organization trades under several names.
- **Product**: items, materials and services. Here MDM overlaps with product information management, and the hard parts are variants, packaging levels and regulated attributes.
- **Location**: sites, warehouses, plants and delivery points. Addresses look simple but change, and one legal entity may operate many sites.

Each domain has shared external identifiers that reduce matching effort. GS1 assigns the Global Trade Item Number (GTIN) to products and services and the Global Location Number (GLN) to parties and locations (GS1, Identification keys). For legal entities, the Legal Entity Identifier (LEI) is a 20-character alphanumeric code based on ISO 17442. Its data answers "who is who" (Level 1) and "who owns whom" (Level 2) (GLEIF, Introducing the LEI). The LEI standard page explains how those two levels map to party master data.

Identifiers and matching

When every source carries the same unique identifier, mastering is a join. In practice sources lack a shared key, or they carry several keys that conflict. The task then becomes entity resolution: deciding which records refer to the same real-world entity. Binette and Steorts define the related terms precisely. Deduplication removes duplicates within one database. Record linkage merges several databases and removes duplicates across them. Entity resolution does both at once. Canonicalization (also called merging or data fusion) builds "one representative record" after entity resolution, to resolve conflicting information (Binette and Steorts, (Almost) All of Entity Resolution).

The same authors describe a four-stage pipeline: attribute alignment, blocking, record linkage and canonicalization. They also name a third difficulty, alongside scale and accuracy: "appropriately quantifying and propagating uncertainty" across the pipeline (Binette and Steorts). An MDM hub runs this pipeline in production. Its match rules perform record linkage. Its merge or survivorship rules perform canonicalization. The uncertainty they mention is exactly what a published golden record tends to drop.

Stewardship and governance

A data steward is the person accountable for the quality and definitions of a set of data in day-to-day operation. A data owner is the business role accountable for the data. The IJITEE review notes that an organization must define the data owner function and the data steward role to support governance (Hikmawati et al., 2021). In MDM, stewards review match candidates the algorithm could not settle, approve merges and unmerges, and maintain the rules that pick surviving values.

Stewardship relies on a shared idea of quality. Wang and Strong's survey-based framework groups data quality dimensions into four categories:

- intrinsic: accuracy, objectivity, believability, reputation;
- contextual: value-added, relevancy, timeliness, completeness, appropriate amount of data;
- representational: interpretability, ease of understanding, representational consistency, concise representation;
- accessibility: accessibility, access security.

(Wang and Strong, 1996). Two of their findings bear directly on MDM. Data consumers treated believability and reputation as part of intrinsic quality. In one application of the framework, users said they needed more information about data sources to judge reputation and believability (Wang and Strong). Source information is therefore a quality requirement, not an optional extra.

Hub styles and where the record lives

Practitioners describe four implementation styles. These are industry terms, not a standard, and published descriptions disagree on details.

Style (industry term)	Where authoring happens	What the hub holds	Write-back to sources
Registry	In source systems	A matched view of source records with global identifiers; source data unchanged	No
Consolidation	In source systems	A consolidated golden record, with steward review	Usually no (descriptions differ)
Coexistence	In sources and hub	Golden record, synchronized back	Yes
Centralized (also called transactional)	In the hub	The system of record for master data	Hub publishes to subscribers

Sources: Reltio, What are MDM implementation styles? (vendor) and Datamation, MDM implementation styles (trade press). The two disagree. Reltio describes consolidation as an upgraded registry with steward review and reserves write-back for coexistence. Datamation says consolidation changes are pushed to the original sources. Datamation also notes that the centralized style is "sometimes also known as the Transaction style." Ask any vendor or integrator to state which variant they mean.

For evidence, the important column is the third one. In a registry, the original values still sit in the sources, so they can be traced back if the cross-reference keeps source keys. In consolidation, coexistence and centralized styles, the hub publishes values that were selected or changed. Whether anyone can later reconstruct why a value was chosen depends on what the hub kept about the choice.

What a hub rarely records: source, date, dissent

A golden record is the single representative record the hub publishes for an entity. Survivorship is the set of rules that picks each attribute's value when matched sources disagree (a separate article covers survivorship rules in depth). Bleiholder and Naumann classify conflict-handling strategies into three families:

- conflict ignorance, which passes all values on;
- conflict avoidance, such as "trust your friends", meaning prefer one source decided in advance;
- conflict resolution, such as "cry with the wolves" (majority vote), "keep up to date" (most recent value) and "meet in the middle" (compute a new value).

(Bleiholder and Naumann, 2006). They draw out two consequences:

1. With conflict avoidance, the system "does not necessarily know" that a conflict existed.
2. Deciding strategies can keep where-lineage, the source a value came from. Lineage "is usually not attached" to values created by mediating strategies, which compute a value that no source supplied.

Conflicts are common, not an edge case. In Deep Web stock and flight data, which the authors expected to be fairly clean, Li et al. found more than one value for 70% of data items. They attributed nearly half of those conflicts to ambiguity, about 20% to out-of-date data and about 30% to mistakes. Even well-known authoritative sources were "not perfect" (Li et al., Truth Finding on the Deep Web, PVLDB 2013). Those measurements come from public web sources, not enterprise hubs. They show why one preferred source does not settle a value.

Worked example (fictional). Halden Therapeutics AG is a fictional pharmaceutical company. A fictional bank, Nordmark Bank, holds it as a counterparty in three systems and also keeps an extract from a public registry. All identifiers below are placeholders, not real codes.

Attribute	Lending system (updated 2024-06-03)	KYC onboarding (reviewed 2025-11-12)	Trade finance (updated 2026-03-20)	Registry extract (retrieved 2026-02-01)
Legal name	Halden Therapeutics AG	Halden Therapeutics AG	Halden Therapeutics	Halden Therapeutics AG
Registered address	Seestrasse 10, Basel	Seestrasse 10, Basel	Rheinweg 4, Basel	Rheinweg 4, Basel

Attribute	Lending system (updated 2024-06-03)	KYC onboarding (reviewed 2025-11-12)	Trade finance (updated 2026-03-20)	Registry extract (retrieved 2026-02-01)
LEI	(blank)	LEI-H1 (placeholder)	LEI-H1 (placeholder)	LEI-H1 (placeholder)
Direct parent	(blank)	Halden Holding SE (fictional)	(blank)	none reported

The hub's survivorship rules are typical: KYC wins for name and parent ("trust your friends"), the most recent update wins for address ("keep up to date"), and any non-blank value wins for LEI. The published golden record is:

Attribute	Golden value	What the golden record does not show
Legal name	Halden Therapeutics AG	Nothing material (sources agree up to a missing suffix)
Registered address	Rheinweg 4, Basel	That it came from trade finance, dated 2026-03-20, and that KYC still had Seestrasse 10 on 2025-11-12
LEI	LEI-H1	Which sources carried it, and that lending never did
Direct parent	Halden Holding SE	That the registry extract of 2026-02-01 reported no parent, which contradicts KYC

Every golden value is defensible under its rule. Still, a risk officer reading "parent: Halden Holding SE" cannot tell that a later public source disagreed. An auditor cannot tell whether the address change was ever verified, or only won on recency. Many products keep cross-reference and history tables in the hub. The test is whether the published value carries its source, its as-of date and the competing values, or whether a steward must reconstruct them from logs after the question is asked. The comparison of golden records with evidence-carrying records sets out the two designs side by side.

Regulatory reasons to want more

BCBS 239 (Basel Committee on Banking Supervision, international standard implemented by national supervisors). The Principles for effective risk data aggregation and risk reporting applied first to global systemically important banks (G-SIBs), which were to meet them by January 2016 (BCBS, 2013, paragraph 14). Several principles read like MDM requirements:

- Principle 2 (data architecture and IT infrastructure), paragraph 33: banks should have integrated data taxonomies and "single identifiers and/or unified naming conventions" for legal entities, counterparties, customers and accounts.
- Principle 3 (accuracy and integrity): banks should reconcile risk data with their sources and strive towards "a single authoritative source for risk data per each type of risk". A footnote defines reconciliation as comparing items or outcomes and "explaining the differences".
- Principle 3, paragraph 39: supervisors expect banks to document and explain all aggregation processes, manual or automated.

(BCBS 239). A golden record helps with identifiers. Explaining differences also requires the losing values and their sources.

EU AI Act Article 10 (European Union). Article 10 of Regulation (EU) 2024/1689 requires that training, validation and testing data sets for high-risk AI systems be subject to data governance and management practices. Those practices concern, among other things:

- data collection processes and "the origin of data" (Article 10(2)(b));
- data-preparation operations such as "cleaning, updating, enrichment and aggregation" (Article 10(2)(c));
- the identification of data gaps or shortcomings (Article 10(2)(h)).

Annex IV, point 2(d), asks technical documentation to describe training data sets, including "information about their provenance" (AI Act). When master data feeds such data sets, survivorship is a data-preparation operation. A team that cannot say which source supplied a value will struggle to document the origin of the data.

The Digital Omnibus on AI, Regulation (EU) 2026/1744, published in the Official Journal on 24 July 2026 and in force since 27 July 2026, amended Article 10 (Digital Omnibus on AI):

- It replaced paragraph 1, so the quality criteria now refer to paragraphs 2 to 4 and to a new Article 4a(1).
- It deleted paragraph 5, the special-category data exception for bias detection, which now sits in the new Article 4a.
- It replaced paragraph 6.
- Under the amended Article 113, Chapter III Sections 1 to 3, which include Article 10, apply from 2 December 2027 for high-risk systems under Article 6(2) and Annex III, and from 2 August 2028 for those under Article 6(1) and Annex I.

Selection questions

The questions below are useful when choosing or reviewing an MDM approach. They avoid product comparison and test whether the design can answer audit questions:

1. For each published attribute, can the hub return the source system, the source record key and the as-of date, without a steward rebuilding it?
2. When survivorship drops a value, is the losing value kept and queryable, or only visible in change logs?
3. Which conflict-handling family does each rule belong to, and do any rules compute values that no source supplied?
4. Can a merge be undone with every original value recovered, and is the unmerge itself recorded?
5. Are external identifiers (LEI, GLN, GTIN) stored with the date and registry they were checked against?
6. Can the hub answer "what did we believe on date X" as well as "what do we believe now"?
7. Who owns each survivorship rule, and where is the rule's change history?

The master data management glossary entry gives a short definition for readers who need one. Reliable Context is a system of record built around source-backed values.

Sources

1. What is master data management? IBM Think (Tim Mucci), undated. https://www.ibm.com/think/topics/master-data-management
2. Master Data Management: Build a Strong Process, Framework and Solution (FAQ: What is master data?). Gartner. https://www.gartner.com/en/information-technology/glossary/master-data-management-mdm
3. DAMA-DMBOK: Data Management Body of Knowledge. DAMA International. https://www.dama.org/cpages/body-of-knowledge
4. Hikmawati S., Santosa P.I., Hidayah I. Improving Data Quality and Data Governance Using Master Data Management: A Review. IJITEE 5(3), September 2021. https://journal.ugm.ac.id/ijitee/article/download/66307/32204
5. GS1 Identification Keys. GS1. https://www.gs1.org/standards/id-keys
6. Introducing the Legal Entity Identifier (LEI). GLEIF. https://www.gleif.org/en/about-lei/introducing-the-legal-entity-identifier-lei
7. Binette O., Steorts R.C. (Almost) All of Entity Resolution. arXiv:2008.04443, v3 January 2022. https://arxiv.org/abs/2008.04443
8. Wang R.Y., Strong D.M. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12(4), 1996 (MIT TDQM full-text copy). https://web.mit.edu/tdqm/www/tdqmpub/beyondaccuracy_files/beyondaccuracy.html
9. What are MDM Implementation Styles? Reltio (vendor glossary). https://www.reltio.com/glossary/master-data-management/what-are-mdm-implementation-styles/
10. Robb D. 4 Popular Master Data Management Implementation Styles. Datamation, 22 August 2023. https://www.datamation.com/big-data/master-data-management-implementation-styles/
11. Bleiholder J., Naumann F. Conflict Handling Strategies in an Integrated Information System. Informatik-Berichte 197, Humboldt-Universität zu Berlin, 2006. https://edoc.hu-berlin.de/18452/3112
12. Li X., Dong X.L., Lyons K., Meng W., Srivastava D. Truth Finding on the Deep Web: Is the Problem Solved? PVLDB 6(2), 2013 (arXiv:1503.00303). https://arxiv.org/abs/1503.00303
13. Principles for effective risk data aggregation and risk reporting (BCBS 239). Basel Committee on Banking Supervision, January 2013. https://www.bis.org/publications/201301-guidelines-principles-effective-risk-data-aggregation-and-risk-reporting.pdf
14. Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union, 12 July 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
15. Regulation (EU) 2026/1744 (Digital Omnibus on AI). Official Journal of the European Union, 24 July 2026. https://eur-lex.europa.eu/eli/reg/2026/1744/oj

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.ibm.com/think/topics/master-data-management
2. https://journal.ugm.ac.id/ijitee/article/download/66307/32204
3. https://www.dama.org/cpages/body-of-knowledge

4. https://www.gartner.com/en/information-technology/glossary/master-data-management-mdm
5. https://www.gs1.org/standards/id-keys
6. https://www.gleif.org/en/about-lei/introducing-the-legal-entity-identifier-lei
7. https://arxiv.org/abs/2008.04443
8. https://web.mit.edu/tdqm/www/tdqmpub/beyondaccuracy_files/beyondaccuracy.html
9. https://www.reltio.com/glossary/master-data-management/what-are-mdm-implementation-styles
10. https://www.datamation.com/big-data/master-data-management-implementation-styles
11. https://edoc.hu-berlin.de/18452/3112
12. https://arxiv.org/abs/1503.00303
13. https://www.bis.org/publications/201301-guidelines-principles-effective-risk-data-aggregation-and-risk-reporting.pdf
14. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
15. https://eur-lex.europa.eu/eli/reg/2026/1744/oj

THE PUBLISHER

About Reliable Context

Reliable Context is a system of record for organizations that fuse private and public information and must defend every value they act on.

The guarantees

One canonical record per real thing: duplicates resolve by rule, ambiguity is shown rather than hidden, and related organizations relate without ever being merged. A verbatim, cited source and an observation date sit behind every value. Unknowns are declared as not found, blocked or not applicable with a reason, never left as an empty cell. Any past state of the record is reproducible, together with what changed since.

Where it runs

It runs inside the customer's perimeter, one database per organization. Permitted sources, model routing, reader clearance and retention are set as the customer's own policies, and every value is exportable with its chain of custody for an auditor or regulator.

What it is not

It is not a vector database or a RAG framework, not a data catalog, MDM or data-quality suite, not a CRM or data-enrichment vendor, and not a chatbot or AI assistant. Where a reader needs one of those, the honest answer is to say so and point them there.

Visit

See Reliable Context.

Reliable Context is in production use in two sectors; more than six thousand of its identity decisions were made by rule and none has been reversed. Articles published here explain general practice, cite their primary and academic sources, and name the jurisdiction of any regulatory rule they describe. They are not legal, regulatory or compliance advice for a specific organization, and requirements change.

Website: https://reliablecontext.com

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.