DATA
MILLENNIUM

ACCURIDS

A GUIDE TO CALCULATING THE ROI FOR REFERENCE AND MASTER DATA MANAGEMENT

Heiner Oberkampf
ACCURIDS

Malcolm Chisholm
DATA MILLENNIUM

accurids.com

datamillennium.com

EXECUTIVE SUMMARY

A Return on Investment (ROI) is often, but not always needed to make a business case for something, and Reference and Master Data (RMD) is no exception. There would be no need to develop an ROI analysis for RMD if it was mandated by regulators, but this is not usually the case, and executive management will need to be provided with an understanding of what it will cost to implement and run RMD and quantification of the benefits of this implementation. RMD will be just one of many budgetary priorities that executive management is asked to fund. The team proposing an RMD initiative may not understand, or even be aware of all of the other competing priorities. Likewise, the RMD team should not expect executive management to have an intuitive and detailed understanding of RMD. This means that not only must an ROI be developed for RMD, but the ROI must be presented in a way that can be consumed easily and effectively by executive management. Also, the ROI need to have the return on the investment in a reasonable timeframe.

In this report we show that the long-term gains from investments in RMD have a yearly ROI 16m$ for any organization with about 10 thousand employees. Today, there is general recognition that Reference Data and Master Data need to be managed well due to its enterprise-wide impact on data quality. Accordingly, RMD is a key part of the data strategies. The required efforts are, however, often under-estimated and we see a gap between recognition of the importance of RMD and the pace at which the maturity of RMD is increasing – often much too slow. While there are content and organizational reasons for the slow adoption, we additionally see a need to more explicitly calculate the ROI for RMD to unlock or re-allocate more resources. The main challenge here is to make long-term effects visible and recognizable in the short term of, say one year or so.

With ACCURIDS, we provide a data registry platform for managing distributed reference and master data across different parts of your organization. The registry is a light-weight approach to reaping benefits of RMD while keeping integration efforts and change in the existing data landscape low.

ROI ANALYSIS KEY COMPONENTS

Efficiency
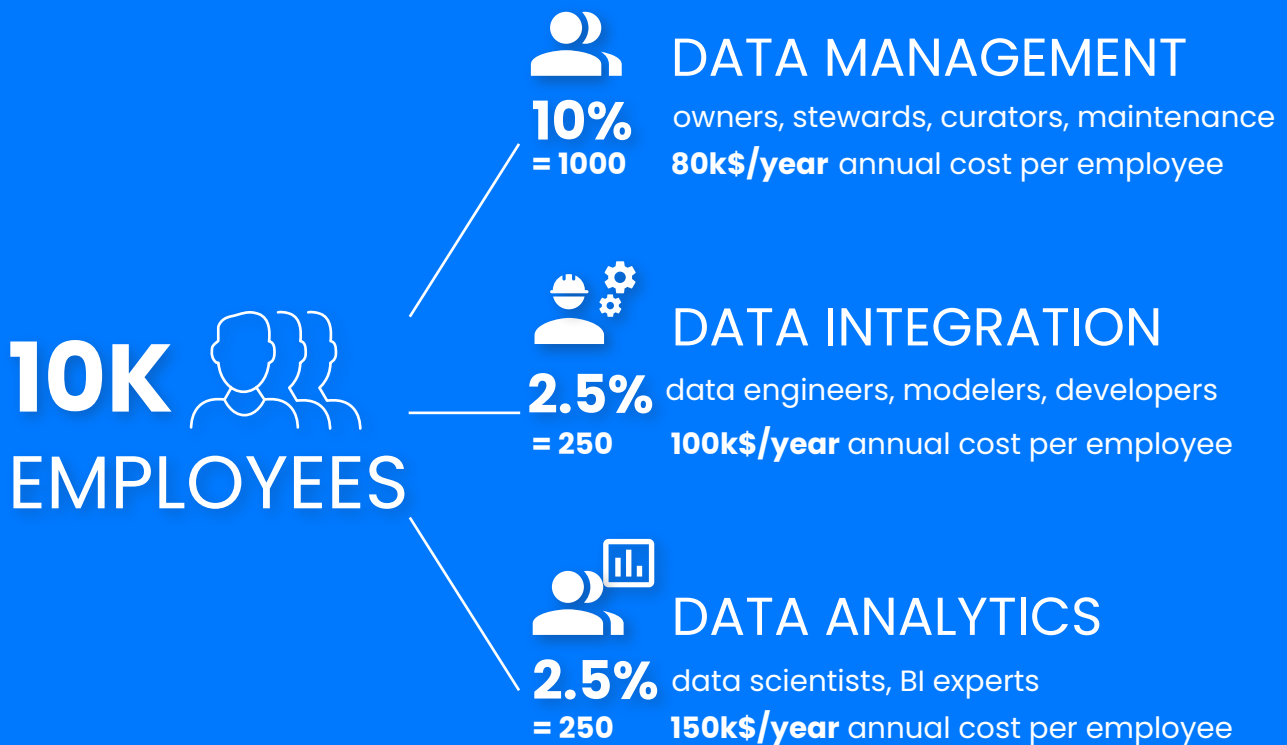The ability to reduce the required time and cost to perform a given task.

Effectiveness
The ability to increase revenue and attain business goals.

Risk mitigation
The ability to eliminate or reduce risks or to better deal with the consequences if a risk becomes an issue.

SCENARIO FOR OUR CALCULATION

10K EMPLOYEES

DATA MANAGEMENT
10% = 1000 — owners, stewards, curators, maintenance
80k$/year annual cost per employee

DATA INTEGRATION
2.5% = 250 — data engineers, modelers, developers
100k$/year annual cost per employee

DATA ANALYTICS
2.5% = 250 — data scientists, BI experts
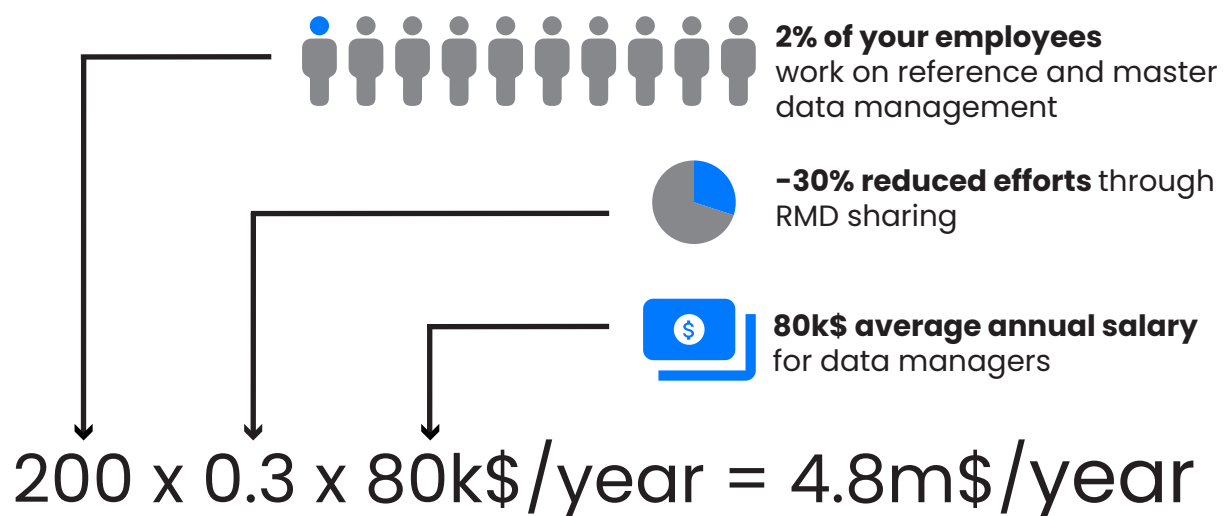**150k$/year** annual cost per employee

EFFICIENCY

Efficiency is realized through reduced redundancies in RMD maintenance through sharing and reduced data integration and analysis efforts. As we rationalize, proper RMD can save about 16m$ per year. This is a great sum for justifying corresponding significant investments.

1 Reference master data sharing

Redundantly or (un-)managed RMD is a waste of resources. Discovering RMD, keeping it up to date, analyzing its semantics, checking its data quality, and answering questions about it are examples of these tasks. It is not too difficult to estimate the work-hours required for each task, multiply the result by the number of redundant instances, and determine the level of wasted effort in the enterprise.

A nice calculator for the ROI effect of data sharing is provided by CDQ, according to which about 30% of the costs can be saved through sharing. For your RDM sharing initiative, this means that you can save about 4.8m$/year.

2% of your employees work on reference and master data management

-30% reduced efforts through RMD sharing

80k$ average annual salary for data managers

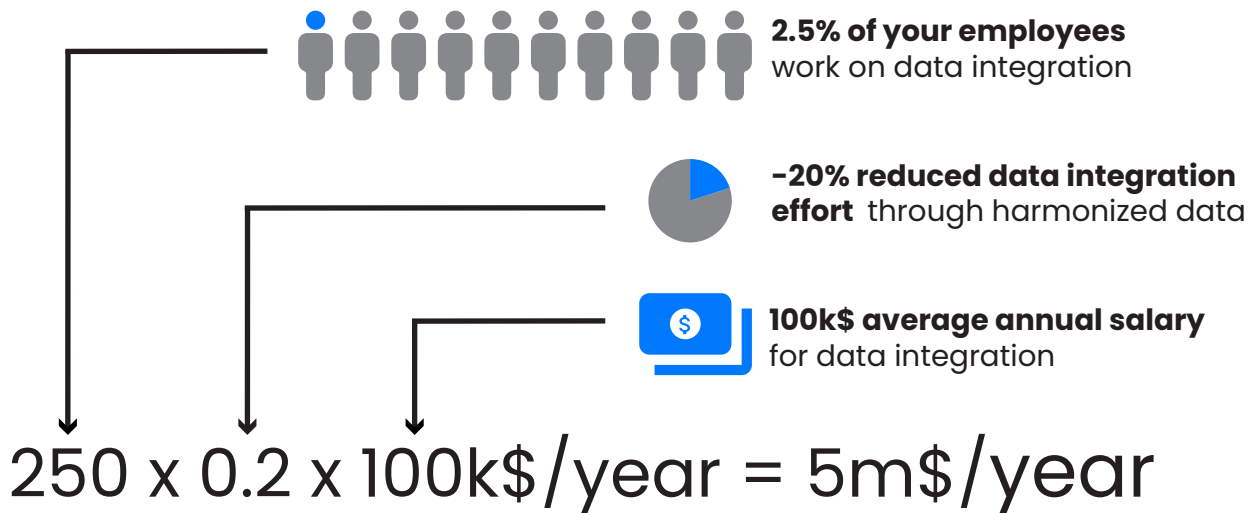$$200 \times 0.3 \times 80k\$/year = 4.8m\$/year$$

The ACCURIDS product is built to realize sharing benefits as it provides an RMD registry where RMD from different parts of your organization can be shared with others. Read more about this concept in our paper Managing Data as a Product with Distributed Reference & Master data.

2 Reduced integration efforts

When RMD is broadly adopted in your enterprise, data engineers can integrate data from different sources with less effort as the context of data is largely harmonized at the source or because mappings are centrally provided.

We assume a 20% reduced data integration effort for engineers, which will be resulting in savings of 5m$/year.

2.5% of your employees work on data integration

-20% reduced data integration effort through harmonized data

100k$ average annual salary for data integration
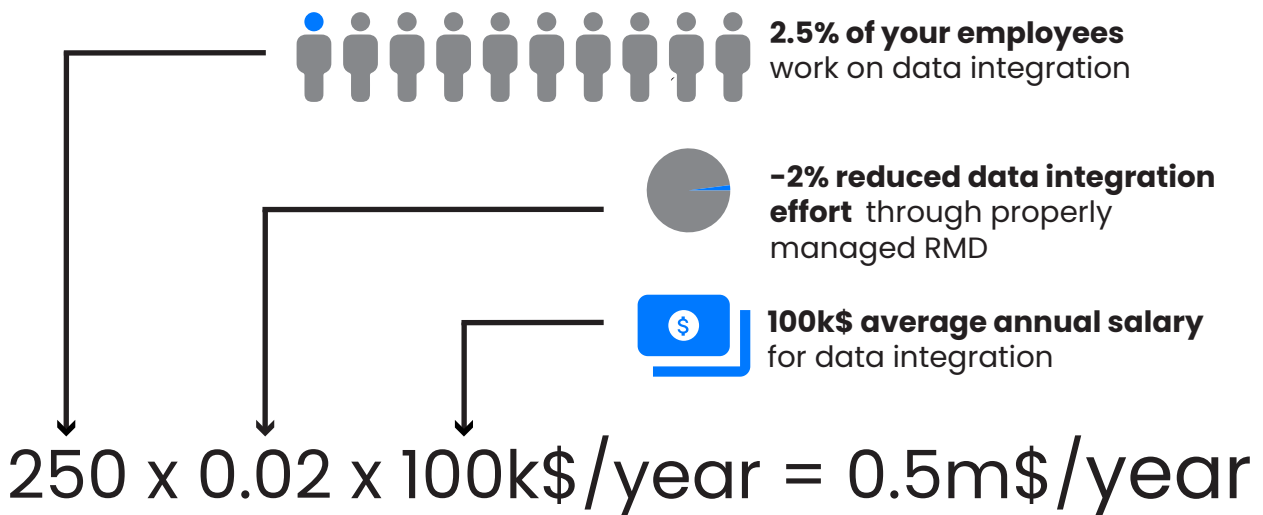
$$250 \times 0.2 \times 100k\$/year = 5m\$/year$$
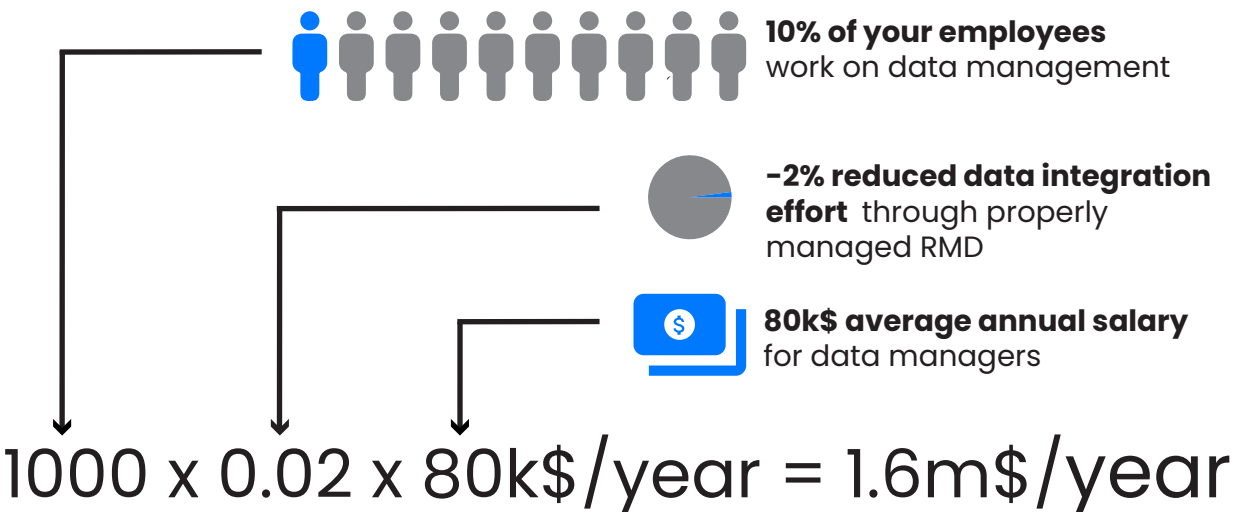
3 Improved data quality

A report can break because an upstream reference data table has had a new entry added to it, which the report does not recognize. Similarly, an Extract-Transform-and-Load (ETL) job can break when the logic in it that depends on specific reference data values encounters a new entry from an upstream source. Gathering statistics on such issues, and estimating the losses, can help build an ROI case.

We assume that at least 10% of the time of data management and data integration efforts are attributed to correcting processing issues and that 20% of those (i.e. 2% overall) can be eliminated with properly managed RMD. This leads to total savings of 2.1m$/year.
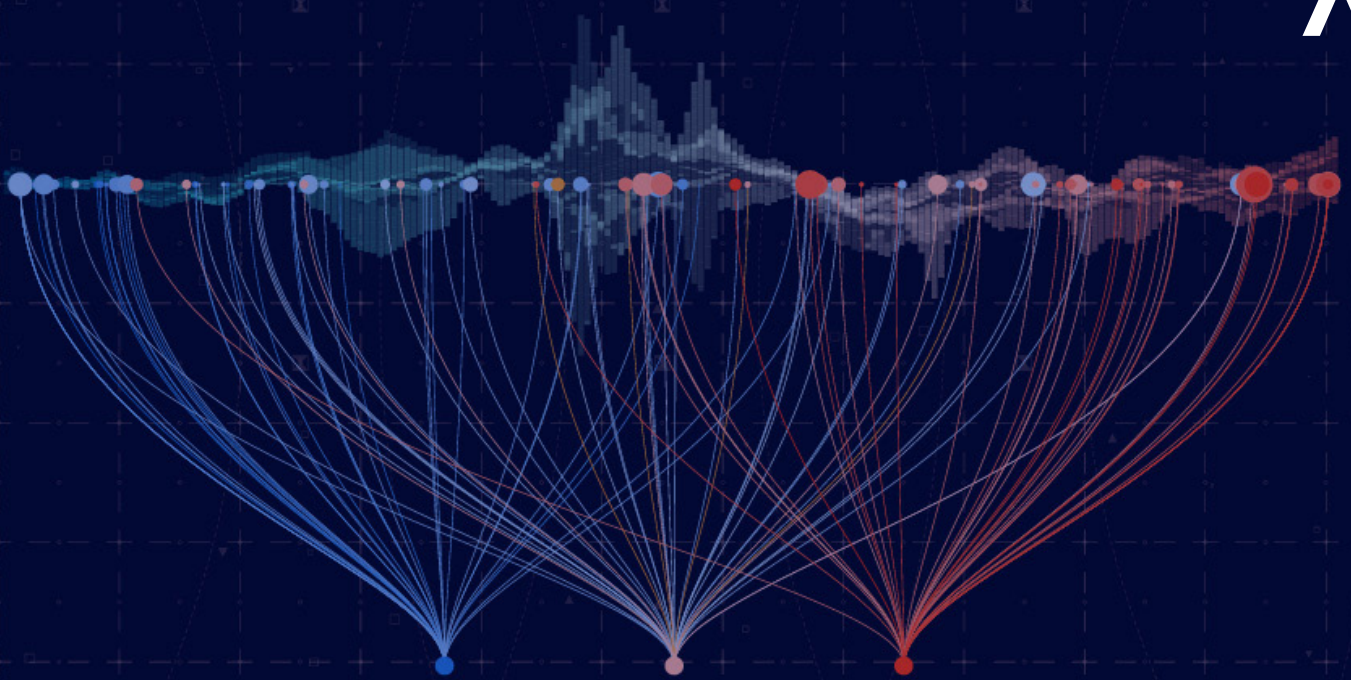
3 A   Data integration aspects

2.5% of your employees work on data integration

-2% reduced data integration effort through properly managed RMD

100k$ average annual salary for data integration

$$250 \times 0.02 \times 100k\$/year = 0.5m\$/year$$

3 B   Data management aspects

10% of your employees work on data management

-2% reduced data integration effort through properly managed RMD

80k$ average annual salary for data managers

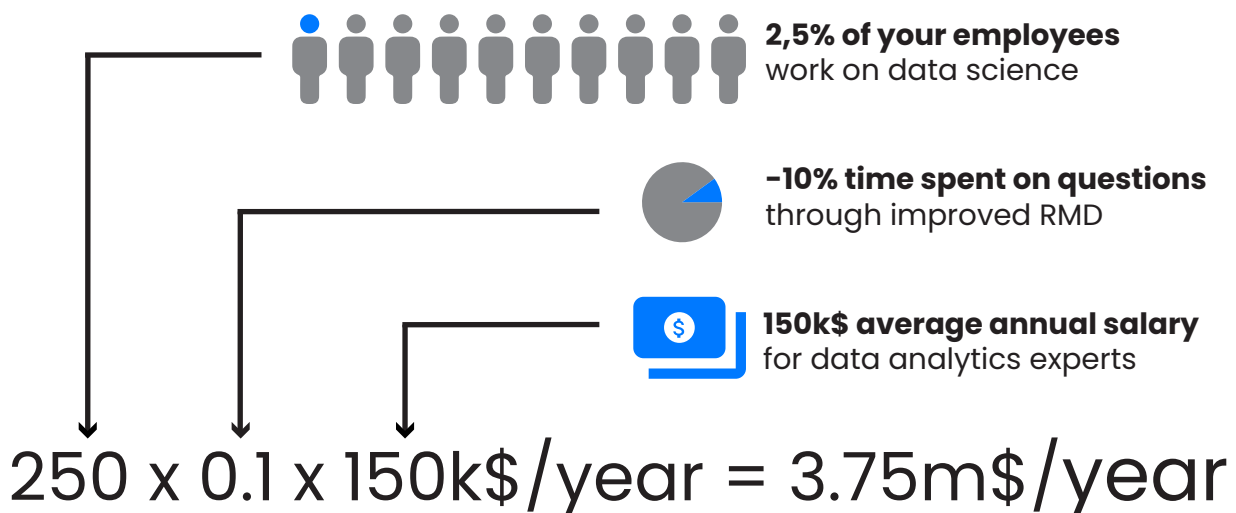$$1000 \times 0.02 \times 80k\$/year = 1.6m\$/year$$

(4) Reduced analysis efforts

Adoption of RMD makes the data more self-explanatory because the context is explicit and more standardized. Thus, data consumers (e.g., data scientists) have a reduced need to ask the people who manage the data (e.g., data owners, stewards, or other key experts) to make sense of the data.

Currently, people active in data management activities spend at least 10% of their time answering questions. We assume that at least 50% of this time can be eliminated with better RMD. This amounts to a total savings of 4m$/year.

2,5% of your employees work on data science

–10% time spent on questions through improved RMD

150k$ average annual salary for data analytics experts

$$250 \times 0.1 \times 150k\$/year = 3.75m\$/year$$

EFFECTIVENESS

As part of the digital transformation, data is now recognized as a product itself - and RMD is what primarily gives business context to other data products. When data needs to be used in an analysis that ultimately is used in making business decisions, it is critical that results are of high quality. Thus, the quality of RMD has a direct impact on several aspects listed below. The ROI of this category heavily depends on the concrete business. The below-mentioned criteria influence, e.g., time-to-market, innovation speed, increased product pipeline, better decisions, and knowledge advantage.

Increased Discovery

Enable all people in your organization to find the data they need for their tasks. This requires a data registry and index.

More accurate Interpretation

Data meaning needs to be explicit so that data consumers can use it correctly. This means that the reference and master data context must be available to everyone.

Faster Integration

Effective value generation often requires a combination of data from different sources. This is faster if you have a high adoption rate of RMD standards across systems.

Improved Analytics

Analytics can be semi-automated if the data context uses liked reference and master data knowledge graphs that provide rich input to algorithms.

More reliable Results

Data products are reliable only if the data foundation, i.e., RMD is properly governed.

RISK REDUCTION

Lack of Governance of RMD can lead to processing failures and compliance issues.

1 Processing failures

As motivated above, data processing can break or be wrong because data context changes or is not correctly recognized. This leads to inacurrate data and correspondingly to erroneous decisions. Absent or unplanned governance of RMD lead to substitute or "organic" processes on the lowest levels with local and intransparent reference data. This hinders the informattion fluency leading to compliance issues because data cannot be understood within the required time and slowing down of the decision making process in general.

The impact of wrong data on business decisions and corresponding outcomes are heavily depending on your business and the type of errors. For instance, NASA lost a spacecraft because of a mistake with metric units. Though most of your colleagues may be reluctant to talk about similar issues in your organization, it is important to bring these to the surface to avoid them in the future. Additionally, these issues are very illustrative also for management presentations.

② Compliance issues

Since RMD is at the core of enterprise information, audits will very likely touch it - and lack of governance or the inability to answer basic questions about, e.g., products, clinical studies, or other regulated information will have significant consequences. Also, GDPR is related to master data, e.g., about customers or suppliers.

As compliance is a 'non-negotiable', the ROI can be determined best by checking past audits and audit issues regarding the role of RMD. This helps to justify existing RMD maintenance efforts and motivate further investments.

Currently, under GDPR, the fines may sit at 10 million Euros or 2% of the previous year's revenue (whichever is highest). While GDPR focuses on Personal Identifiable Information (PII), it is good practice to have a proper data management structure for all data within a business.

An example would be 1&1 Telecom in Germany getting a 9.55 Million Euro fine for failing to have the proper infrastructure to manage and secure their data. While many of the GDPR Data Breach headlines tend to focus on the unauthorized collection of data and inadequate data security, it can also cover RMDM. Having a proper management system can mitigate this risk, including storage, access and decommissioning of data. See, e.g., Morgan Stanley´s $60 Million fine for failure to properly decommission.

Compliance is a 'non-negotiable' and it is worth making the role of RMD transparent when discussing with senior stakeholders.

Compliance Components

Law   Policies   Regulations   Rules   Standards   Transparency

AND WHAT ABOUT ACCURIDS?

ACCURIDS is a registry for distributed reference and master data. It serves as a discovery solution and provides reliable access for anyone in your organization to reference data managed in different applications. The key functionalities include

1. Global lookup service that allows searching for any term or code used within the enterprise. This allows data stewards to find the preferred terminologies for a given domain and encourages re-use instead of re-creation.

2. Persistent identifiers can be generated to provide long-term stable and resolvable IDs on which users and applications can rely – even when the place where the reference data is managed changes. Persistent IDs are required for FAIR data.

3. Matching reference data is important to aligning existing terminologies that are already used. ACCURIDS automatically generates matching proposals. Data stewards need to review and approve suggestions before mappings become available.

4. Public standard terminologies such as the Gene Ontology, NCBI taxonomy or Legal Entity Identifiers can be imported into ACCURIDS so that all internal consumers are using the same, latest and, validated version.
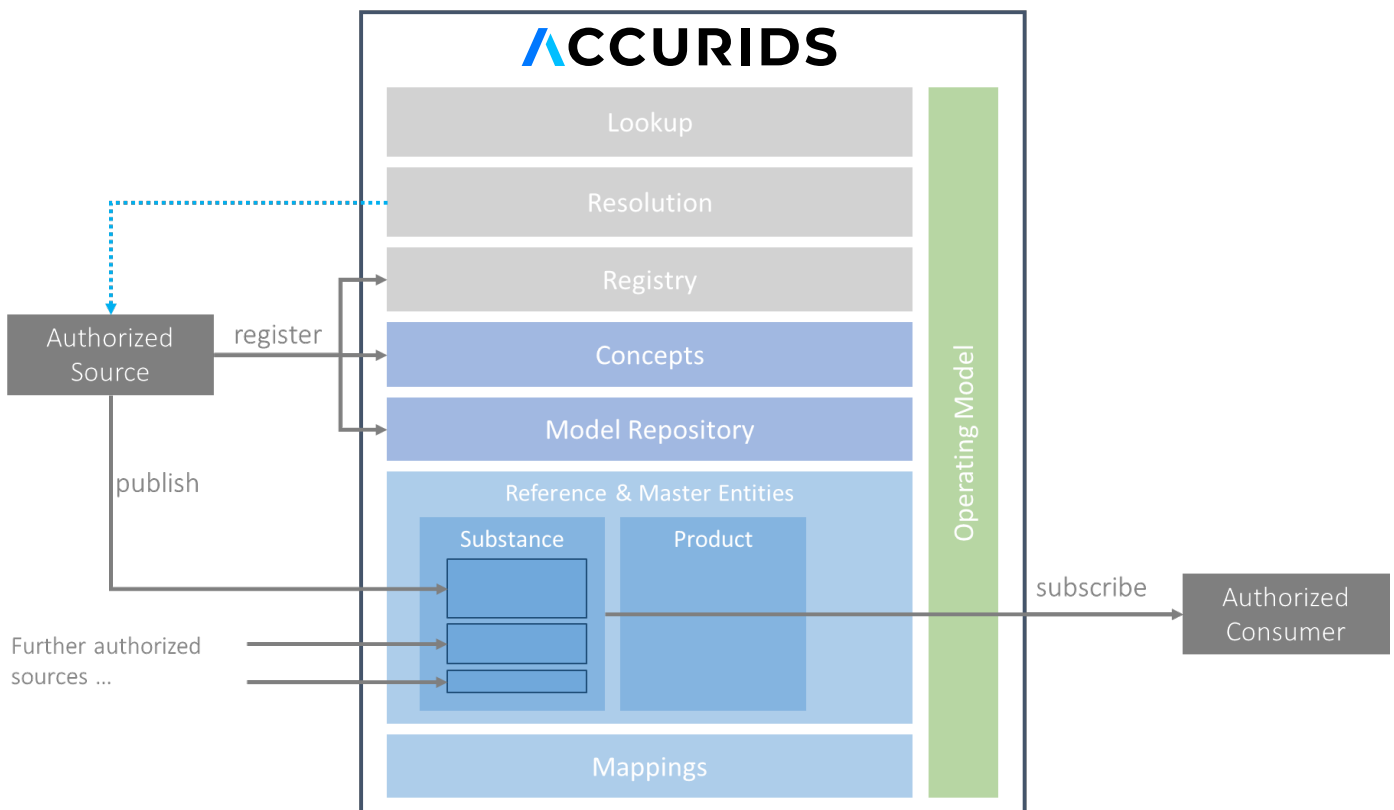
CONCLUSION

Sometimes an organization has got to the point where it recognizes that RMD must be addressed, and in such cases, a well-documented ROI analysis is more of a precursor to a strategy. However, in most organizations, the ROI will need to advocate for an investment in RMD that may not be intuitive to executive management. Careful attention to all categories of return, with studies from the organization, as well as explicit linkage to business strategy, are needed.
Academic studies can sometimes add needed support. The estimation of the investment needed must be based on a very detailed set of requirements, which inevitably require a somewhat clear picture of the future state and a link to business goals.

The estimation of the return must be credible, and not vague, and be able to stand up to tough questioning. None of this is easy, but proposals for resources rarely are. Good preparation will ensure the success of an RMD ROI proposal.

15.6m$ savings annually
- 4.8m$ through RMD sharing
- 5m$ for reduced integration efforts
- 2.1m$ by virtue of increased data quality
- 3.75m$ through reduced analysis efforts

More effective data products
- increased discovery
- more accurate interpretation
- faster integration
- improved analytics
- more reliable results

Significant risk reduction
- reduced data processing errors
- higher compliance standards

RELATED WORK

Though frequently overlooked, academic studies can provide additional support to an ROI analysis. This is particularly effective in organizations that are heavily regulated or which have a solid orientation to science.

One example is [Operational Risk and Reference Data: Exploring Costs, Capital Requirements and Risk Mitigation](#) by Grody et al. This paper explores RMD in the financial services industry, estimating the level of RMD in securities trades, and gathering reasons for failures of securities trades from practitioners. It lays out how industry initiatives and the regulatory framework have been impacted by RMD problems and presents an in-depth risk analysis.

Another paper, again by Grody et al., is [Final Report on Global Identification Standards for Counterparties and other Financial Market Participants](#). This contains an analysis of costs associated with the lack of RMD standards in global financial markets and could be a good source of ideas for any RMD ROI analysis, even outside of financial services.

The [CDQ Business Case Calculator for data sharing](#) is a nice example of how one can get to concrete numbers related to your organization.

In our paper on "[Managing Data as a Product with Distributed Reference & Master Data](#)", we describe which components are critical to realizing the benefits effectively.

There is no doubt that additional credibility is provided to an RMD ROI analysis by the addition of careful, independent research that can provide reliable facts about RMD and share well-grounded recommendations.

ACCURIDS

Learn more about how ACCURIDS can help
you govern and publish reference and
master data from distributed sources.
Visit accurids.com