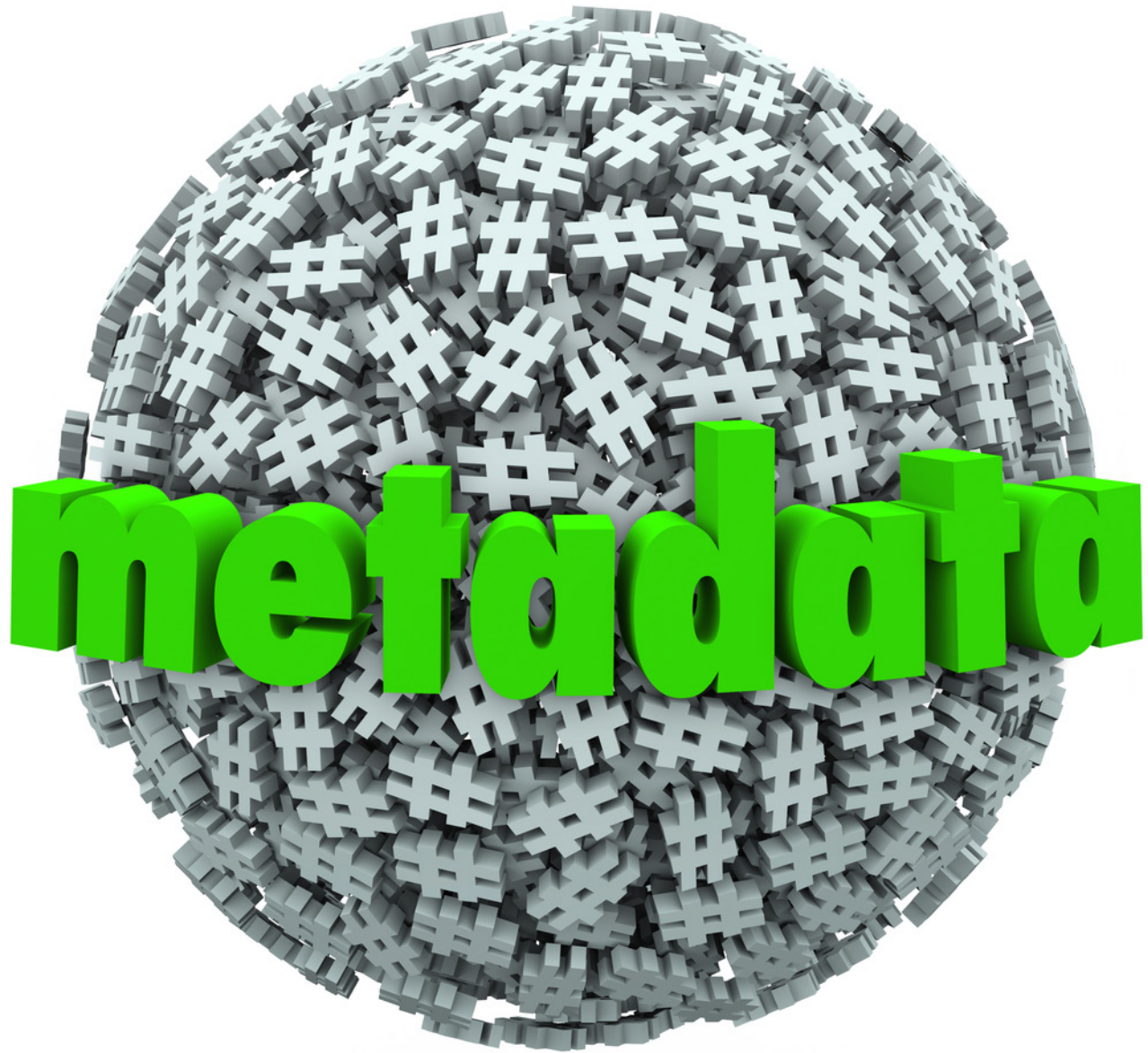


Emerging Trends in Metadata Management:

A DATAVERSITY® 2016 Report on the Top Business & Technical Drivers for Metadata

by Donna Burbank and Charles Roe



Contents

1. Executive Summary

- a. Objectives
- b. Findings

2. Research and Demographics

- a. Scope of Research
- b. Demographics

3. Introduction – A Contemporary Metadata Update

- a. Business and Technical Drivers – The Necessity of Metadata Management
- b. The Importance of Metadata to the Enterprise

4. Metadata Personnel and Training

- a. “How many specialists do you have in your organization who regularly works with your Metadata resources?”
- b. “Who are the consumers of Metadata in your organization?”
- c. “How do you get trained in Metadata?”
- d. Analysis of Results

5. Metadata Management and Usage within the Enterprise

- a. “For what types of information do you currently manage Metadata?”
- b. “For what types of information would you like to manage Metadata for in the future, but are currently unable to do so?”
- c. “What are your current main Use Cases for Metadata?”
- d. “What is the future planned Use Cases of Metadata within your organization (that you are NOT doing today)?”
- e. “What are your Use Cases for showing relationships between Metadata objects?”
- f. “Do you have a defined Metadata Strategy in your organization?”
- g. Analysis of Results

6. Metadata Tools, Functionalities, and Standards

- a. “Which of the following Metadata Solutions does your organization currently use?”
- b. “How do you feel about the quality of current tools for Metadata Management available in the market today?”
- c. “Which Metadata tool functionality is important to you?”
- d. “Are you using any Metadata Standards in your organization?”
- e. Analysis of Results

7. Conclusion

1. Executive Summary

The push to become a data-driven enterprise is a growing trend in the industry as more and more organizations see the value of data as a strategic asset. In order to effectively leverage this asset, data must be of high quality and clearly understood. Metadata Management is a foundational strategic practice that gives an enterprise the ability to manage its data in a reliable, authoritative, and timely fashion. Metadata is more important than ever before, with more than 80% of survey respondents stating that Metadata is as important, if not more important, than in the past. It is one of the most requested topics on the DATAVERSITY® website, and this interest was a primary impetus for the creation of this paper. Metadata is a key element to the success of the majority of Data Management practices that drive successful organizations, such as Data Governance and Data Stewardship, Data Warehousing, Master Data Management, Enterprise Resource Management, Customer Relationship Management, Business Intelligence and Analytics, and any other related activities within a data-driven enterprise.

The burgeoning importance of new technologies such as Big Data, the Internet of Things, Advanced Analytics, Data Lakes, Artificial Intelligence/Machine Learning, and Semantic Technologies are creating a landscape of more complexity, higher volumes of data, and an even greater need to govern and understand that data with more certainty. What affect are these technologies having on Data Management in general and on Metadata Management specifically? How has Metadata Management evolved to support these new technologies and where does Metadata Management exist today within the data-driven landscape? This paper seeks to explore these questions through a survey of data professionals actively working in the industry.

a. Objectives

This report is a comprehensive evaluation of a DATAVERSITY 2016 Metadata Management Research Survey meant to provide greater clarity into the overall application of Metadata Management within the enterprise. We investigated the issue across multiple industries, in organizations of varying sizes, and from the perspective of many different job functions to gain the most thorough viewpoint possible. We asked a number of different questions, some open-ended, so the respondents could give their personal opinions on the issues, including (but not limited to):

- What is your definition of Metadata?
- Why is Metadata important to your organization?
- How do you receive training in Metadata Management?
- For what types of information do you currently manage Metadata?
- What types of information would you like to manage Metadata for in the future?
- What are your primary Metadata use cases? Future use cases?
- What Metadata tools and functionalities are most important to your organization?
- Are you using any Metadata Standards? Which ones?
- Do you have a defined Metadata Strategy?

b. Findings

The survey contained 23 separate questions of varying types. Some of the most remarkable findings were:

- Two-thirds of respondents said that Metadata is more important now than it was ten years ago.
- Business users are the largest group of Metadata consumers at almost 80%.
- Data Warehouses, Relational Databases, Data Models, and Business Glossaries are the top “currently managed” sources of Metadata.
- Big Data Platforms rank as the most wanted future Metadata asset for organizations.
- Data Governance is the most prevalent current Metadata use case, along with Data Quality Improvement, Data Warehousing, and BI Reporting.
- Master Data Management, Big Data Analytics, Data Science, Semantic Technologies, Regulation and Audit, IoT Management, and Efficiency and Agility all rank as significant future planned Metadata use cases.
- 53% of organizations are using tool-specific Metadata solutions in their organizations, while 50% are using Metadata Repositories, and 47% are using Business Glossaries.
- Metadata search, storage repository, and interchange with other tools ranked as the three most important tool functionalities.
- More than a third of respondents said their organizations have an “ad-hoc, loosely defined” Metadata Strategy, with only 14% saying they have a “clearly defined” strategy. Over 25% see a Metadata Strategy as part of a larger Enterprise Strategy.
- More than 35% of respondents said they get their Metadata training from blogs, screencasts, and various sources from the Web.

2. Research Demographics

a. Scope of Research

The central emphasis of the DATAVERSITY 2016 survey on Metadata Management was to get a more focused understanding of where Metadata Management exists within the modern enterprise while also evaluating future growth and planned implementation of Metadata Management resources. The survey asked respondents to answer a number of questions:

- General Demographics (4 questions)
- Metadata Overview Questions (4 questions)
- Metadata Personnel and Training (3 questions)
- Metadata Management and Usage within the Enterprise (7 questions)
- Metadata Tools and Functionalities (5 questions)

b. Demographics

The main demographics questions asked about Job Function, Respondent's Industry, and Number of Employees in Company.

Top job functions included the following:

- Nearly 30% of respondents identified their job function as Data and/or Information Architecture
- Information/Data Governance was in second place at 24.6%
- Business Intelligence and/or Analytics was in third place at 9%
- Executive Management was in fourth at 7.2%
- Other notable answers included IT Management and Other. The Other category allowed respondents to list their own choice, with answers including Data Strategy, Knowledge Management, Research and Teaching, Semantic Technology Consultant, Subject Matter Expert, Sales and Marketing, Training and Development, and Internal and External IT Audit, among many others [Figure 1]:

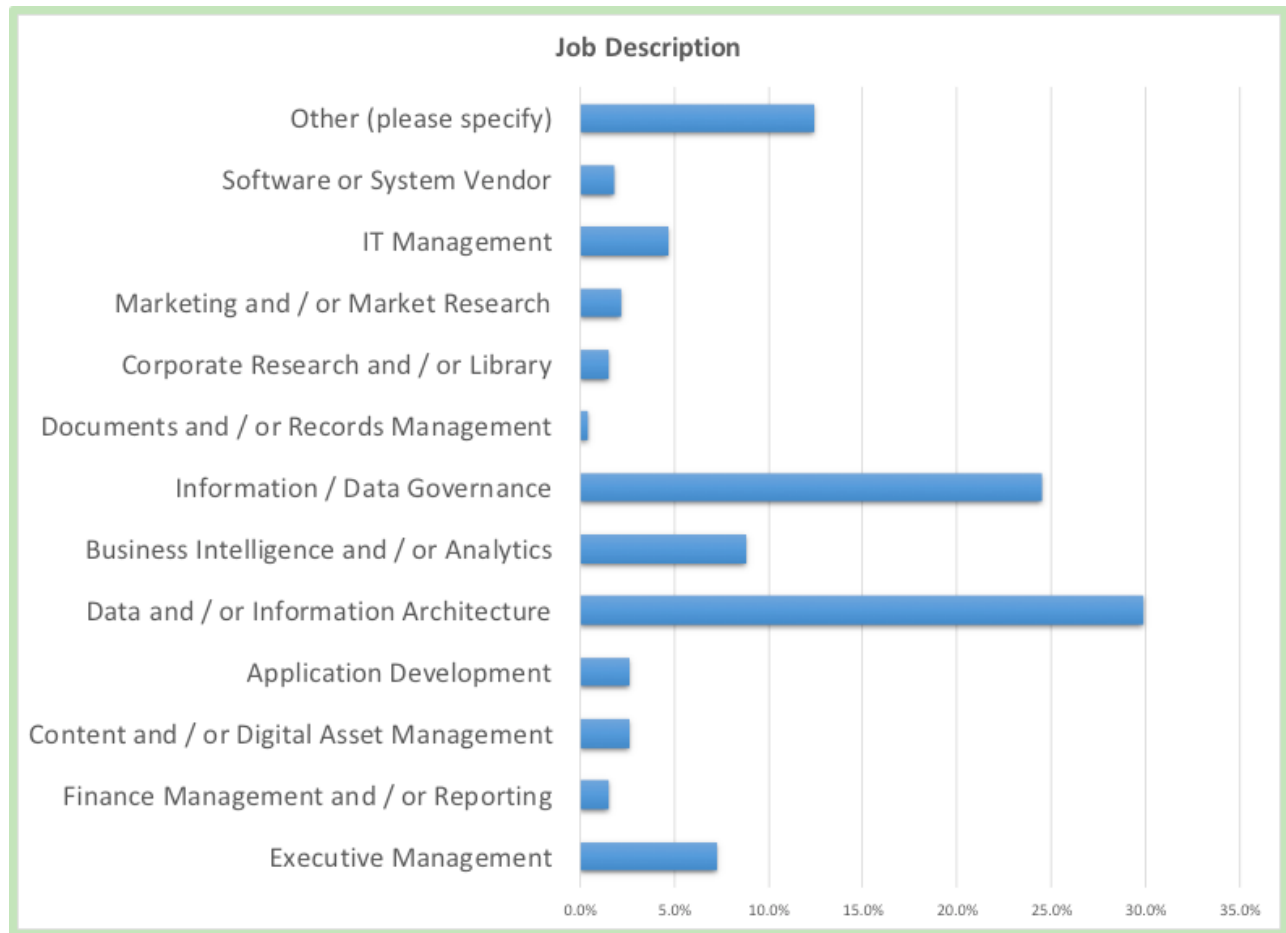


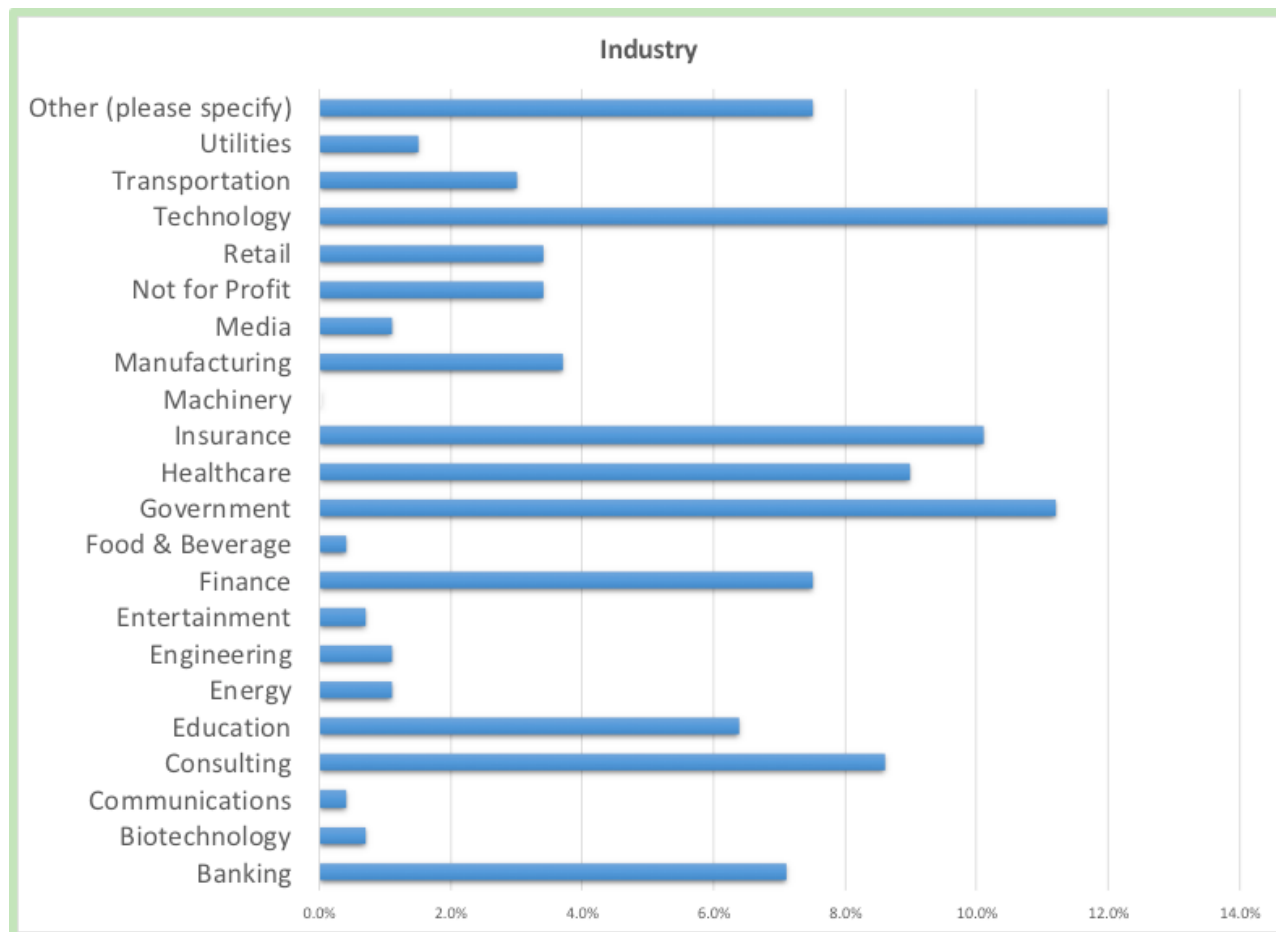
Figure 1: Job Function of Respondents

The most significant percentage of respondents said they worked in:

- Technology (11.85%)
- Government (11.11%)
- Insurance (10.37%)
- Healthcare (8.89%)
- Consulting (8.52%)
- Finance (7.41%)
- Banking (7.04%)

Other well-represented answers included Education, Manufacturing, Retail, and Not for Profit. The Other category garnered 7.41% and had diverse responses such as Pharmaceutical, HR Services, Wholesale/Distribution, Aerospace, and Publishing. [Figure 2]:

Figure 2: Industry of Respondents



The last demographics question asked about Number of Employees in the respondent's company, and the range was quite uniform across-the-board [Figure 3]:

- 1,001 – 5000: 21.93%
- 10,001 – 50,000: 19.70%
- 101 – 1000: 15.61%

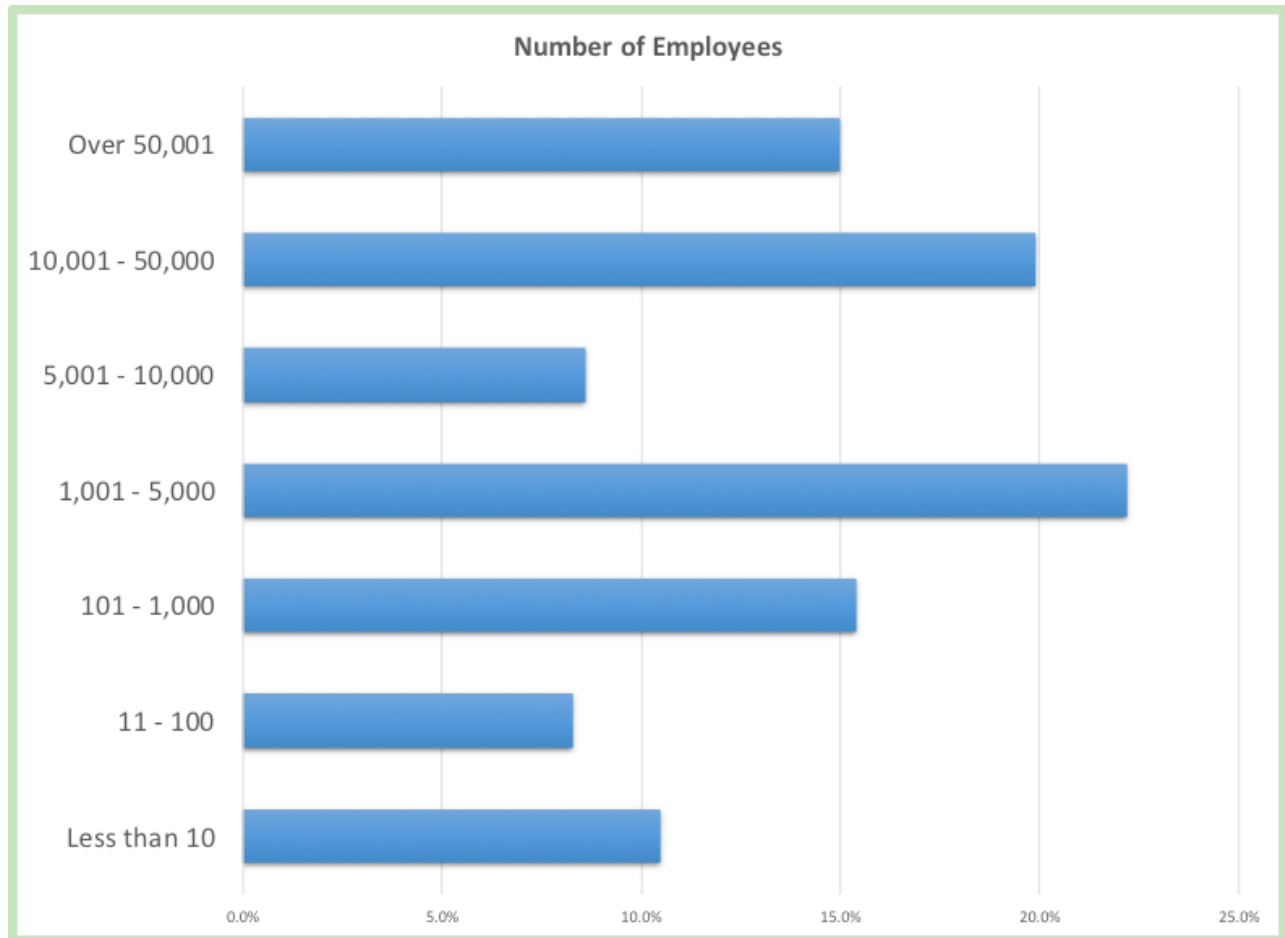


Figure 3: Number of Employees in Organization

3. Introduction – A Contemporary Metadata Update

There are numerous definitions for Metadata. For the sake of simplicity this paper will use one definition: Metadata is data in context. To expand upon that definition a bit further, Metadata is the “who, what, where, why, when, and how” of data.

To get further clarity on the current working definition of Metadata in the industry, the survey asked respondents an open-ended question: “What is your working definition of Metadata, or the working definition within your organization?” We asked that question as an initial jumping off point from which to define the rest of the survey, allowing the respondent’s

“Metadata helps both IT and business users understand the data they are working with. Without Metadata, the organization is at risk for making decisions based on the wrong data.”

to refine their focus and expand upon where Metadata exists within the modern enterprise – whether as an IT system, a tool for the management of data assets, a focus of business users, or within the consciousness of an organization’s employees.

The main definition provided by respondents, in some form or another, was “data about data.” That is the most quoted and most referenced definition of Metadata in any number of blogs, white papers, webinars, and articles on the Internet. But what exactly does “data about data” mean? It’s rather ambiguous. It may make sense to people involved within the data profession, as it brings up images of business rules, data models, physical table structures, and file attributes. But what does it mean to others? To add to our simple working

definition of “data in context,” a few of the many clarifying responses from the survey were:

- “Metadata is information that describes characteristics about data within our company. This allows the reader to understand the business, technical, and operational contexts of the data.”
- “Metadata is the information that describes things. The description of a thing may include information such as its name, its purpose, its main characteristics or features, its ownership, its location, its life stage, and the other things it relates to, such as the rules to use it.”
- “Metadata is a set of characteristics that describe the origin, usage, and brief technical aspects of each object used to store or process business data, as in: tables (and fields), stored procedures, functions, and the like.”
- “Data about data; including name, definition, Data Steward, data type, valid values, location, security classification, source of record, volumetrics, lifecycle and retention, and lineage.”
- “Data that describes the attributes and judgement of an Object during its lifecycle. Some Metadata is considered core to the Identification of the Object and should accompany it at all times e.g. Title. Some Metadata is specific to functions of or applied to that Object and can be linked by association, e.g. Taxonomies.”
- “The context within which a given piece of data exists. Where it came from, where it is headed, what it is used for, why it is used that way, when it was created / last altered, etc.”

- “From an IT view I think Metadata refers to attributes of data elements within our systems. However, I view Metadata as the information needed to truly understand your Analytics. So it includes a business element that is sometimes missed in IT projects in pulling the ‘information’ through to the end users to enable true understanding of the meaning of the Analytics available from the various databases and data marts that support our reporting.”
- “Standardized XML tags that relate the semantic ID of various fields.”
- “Metadata is certainly ‘data about data.’ To be more succinct, Metadata encompasses a realm that includes: Define meaning, context, requirement, and purpose to a data element depending on use, viewpoint, or perspective. Business rules, taxonomies, classification schemes, and allowed values supporting consistent use and reuse across all domains. Provides lineage, derivation, traceability, auditability, and history. Records the transformations and participations that transpired during the lifecycle of a data element. Allows a data element to be found through search, query, or direct access using specific coordinates, relationships, or inferences.”

Metadata provides perspective and answers questions about the data assets within an organization, such as:

- “Who created this data?”
- “What are the business drivers for using this data?”
- “Where did this data come from?”
- “Why are we storing this data?”
- “When does this data need to be purged or deleted?”
- “What is the security or privacy level for this data element?”
- “How is Net Sales calculated within this

database and other databases?”

The list of questions is potentially endless. Asking the right questions at the right time is key, and Metadata allows for appropriate answers.

Metadata is everywhere. In providing context to data, it exists in all of the innumerable platforms, applications, systems, and tools used by organizations in their daily operations – as well as in the minds of their employees. Some of those many sources include relational databases, data models, XML, spreadsheets and text documents, application code, ETL tools, BI tools, Big Data platforms, Internet of Things sources, Data Quality tools, and many more that will be discussed later in this paper. All of these sources are important for understanding the collection, movement, transformation, and analysis of data assets within the enterprise. Metadata Management provides the capability for authoritative governance of those assets. Without quality Metadata all other tools and processes are suspect.

a. Business and Technical Drivers – The Necessity of Metadata Management

The world of business and Data Management has changed. The rise of Big Data, with all of its associated technologies, has altered the data landscape in ways that are only beginning to be understood. Predictive and Prescriptive Analytics are allowing enterprises to use forecasting, sentiment analysis, and other techniques to provide more clarity into their current and future business needs in ways not previously possible. NoSQL systems allow for scalable storage of large volumes of data that can then be analyzed through Data Science techniques for even greater organizational clarity. Machine Learning and Artificial Intelligence are developing methods through advanced algorithms that allow computer systems to learn without extended human interference, process data relationships, and discover hitherto unknown facets of business operations and efficiencies in near real time.

Such advances are important and have the ability to alter the way business is enacted far into the future, but none of them will function without the operative management of the data at its most foundational levels. Two key trends offer challenges for the data-driven business moving forward:

- **Growing Technical Complexity:** In technical terms, the growing number of disparate technologies and sources makes Metadata Management more complicated than the past, when the data sources were more consistent (e.g. flavors of relational databases). Those old sources are not going away -- they are just being mixed in with a growing number of new sources. The technical landscape is only going to grow more complex.
- **Increase in Business User Involvement with Data:** In business terms, demand is growing as data becomes a more critical part of the business. Drivers such as Data Governance and regulation/compliance no longer make Data Management an option, but rather a requirement. In addition, more business users are actively interacting with data, in areas such as Self-Service BI and even self-service data preparation. That makes them more aware of the data issues and closer to the Metadata. Such trends are going to increase moving forward as more businesses see data as a key driver of new business opportunities and a differentiator in the market.

b. The Importance of Metadata to the Enterprise

In completing this survey, we wanted to find out about the relative importance of Metadata within the organizations of our respondents. It appears that business users in a variety of organizations have taken control of the Metadata assets under their authority; they interact with those resources on a daily basis and they know the value of Metadata, even if they weren't accustomed to calling those resources by such terms. We know that data professionals have understood the

value of Metadata for decades. The main results of the survey have shown that organizations see Metadata as more important than ever and many have understood its importance for years. More than 80% of survey respondents state that Metadata is as important, if not more important, now than in the past.

Metadata is recognized as a crucial enterprise asset and the numerous tools and techniques for Metadata Management have become more commonplace. To this end we asked the open-ended question: "Why is Metadata important (or not important) to your organization?" It was one of the initial questions to help set the stage for the rest of the survey. For this particular question, only 3.5% of the respondents said

More than 80% of survey respondents state that Metadata is as important, if not more important, now than in the past.

that it either wasn't important, that they don't have a Metadata Management program in place, or that while it has some importance, there are other aspects of Data Management that are at this time considerably more important to their organization. The rest of the answers garnered a range of answers focusing on why it's important. Some of the key themes that stood out are how Metadata helps business and IT work together, the importance of Metadata Management for new technologies such as Big Data, Data Lakes, and Semantics, and how Data Governance continues to be a crucial need. A small sampling of those that stood out are:

BUSINESS ALIGNMENT & GOVERNANCE

- "Metadata is important to our organization because it allows us to add information to the data and enables:
 - Search: find what is needed in the right context
 - Improved decision making: Insights can be created from data and used to empower decision makers and create value

- Powerful governance: With the right governance information applied to data, rules can be enforced to ensure that the right data is released to the right people in the right context.”

- “We have a variety of constituencies within our organization, and Metadata is necessary for us to bridge the gaps between them. We are also a research facility, so accurate definition of terms is crucial to our scientific mission.”

STRATEGIC BUSINESS DECISION MAKING

- “Metadata helps both IT and business users understand the data they are working with. Without Metadata, the organization is at risk for making decisions based on the wrong data.”
- “Metadata is very important to our organization because it helps the business make better strategic and operational decisions by understanding our data.”
- “We are a heavily regulated and analyzed industry; data is a major resource for our decisions and ultimate financial health.”

BIG DATA ANALYTICS & DATA SCIENCE

- “We are implementing a Data Lake and know that a searchable Data Catalog of Data Lake assets along with contextual information on each asset regarding usage guidelines, Data Quality, Data Management, and data lineage will be necessary for success.”
- “We rely on Metadata as part of Analytics planning where we expect the prepared dataset, input to Analytics, to publish its quality and quantitative metrics and its readiness to use in the upstream analytical applications.”
- “Our Data Scientists and Analysts can understand data only if Metadata and definitions for Metadata are available.”

APPLICATION DEVELOPMENT AND AGILE BEST PRACTICES

- “Agile development requires strong Metadata Management to keep track of changes.”
- “It allows impact analysis, what if we need to change attribute A, it will impact applications X, Y, and Z.”

DATA-DRIVEN BUSINESS MODELS

- “Our organization is about sourcing and selling information (on Local and International Standards, Government Legislation, and Industry Regulations) to customers, and the Metadata associated with these documents is critical for storing, searching, and archiving them.”

COMMON BUSINESS LANGUAGE

- “Metadata is critical as it provides the common lexicon by which the entire organization can converse. Without it, there can be only rudimentary knowledge sharing.”
- “It allows sharing of definitions within the enterprise so we have one version of the truth and not 234 versions.”
- “Metadata is the foundation of our knowledge content of the data and information created from it. Our Metadata environment provides the impact analysis across our business, technical, and information assets generated.”

Metadata Management is clearly a core element to the Data Management systems in many organizations. Metadata Management provides tools, techniques, and can act as a lens that both guides and reveals the flow of data assets throughout the enterprise. It is a necessary component for impact analysis, it allows the creation of shared business vocabularies between applications, helps manage accountability and compliance, allows for more successful data migration and integration processes, is a central element in Data Governance and Data Quality, and can improve enterprise-wide decision making and Business Process Management.

Metadata is data in context. It is the who, what, where, why, when, and how of data.

To expand upon the open-ended question of why Metadata is important to their particular organization, we asked our respondents: “How would you rate the relative importance of and interest in Metadata Management now compared to 10 years ago?” The overwhelming answer was “More Important” at 65.89%, with only 0.47% saying it was “Less Important,” and 13.55% saying it was “The Same.” A total of 20.09% answered the question, “If you’ve seen a change, what is the reason for this change? (Please specify).”

The main reasons for the growth of Metadata Management generally fell into the following main categories [Figure 4]:

- Increasing regulatory and compliance requirements
- The importance of Data Governance and Data Quality
- Increased system complexity
- The need for better Analytics
- The recognition of data as an important enterprise asset

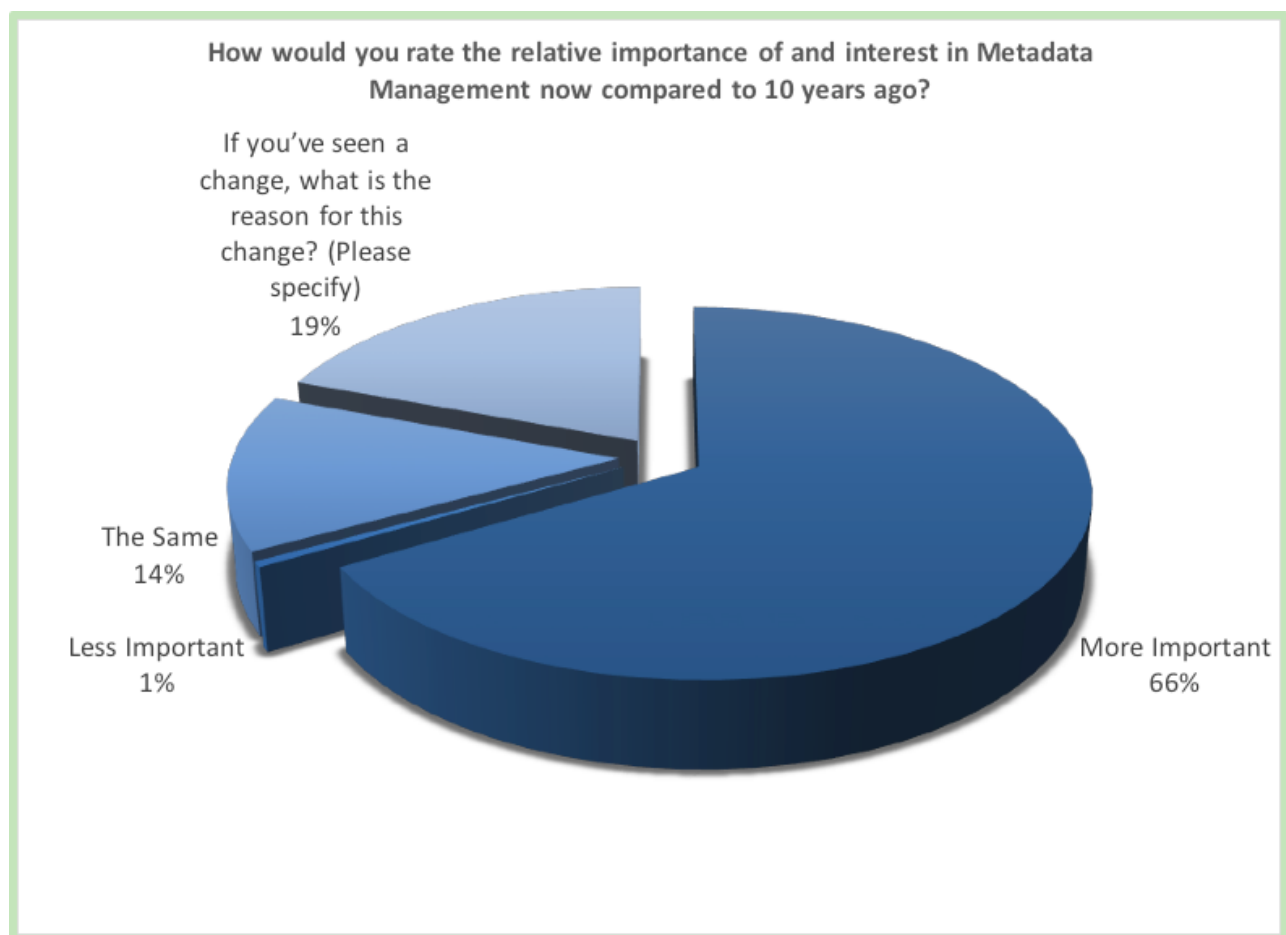


Figure 4: Importance of Metadata

To conclude the introductory section of the survey, we asked how long our respondent's organizations have been using Metadata Management. The results are as follows [Figure 5]:

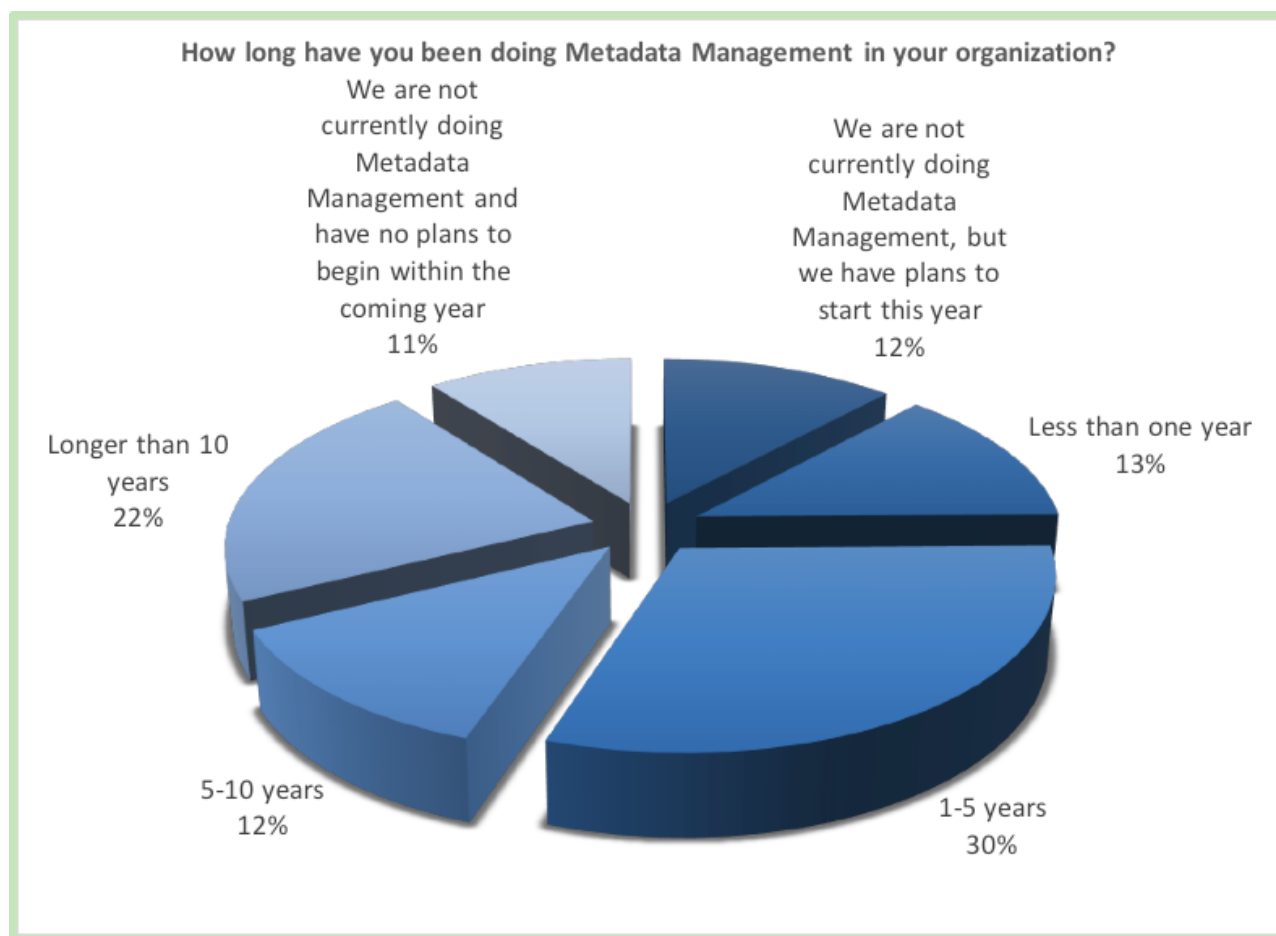


Figure 5: How Long Practicing Metadata

In all, 34.43% have been practicing Metadata Management for five years or more, with 24.53% either planning to start a new program soon or having a program less than one year old. The numbers clearly favor the increasing importance of Metadata across the board, with the smallest percentage saying they are not practicing and don't plan to in the future. As is clear from the answers of our respondents, many organizations are employing Metadata Management in a variety of ways. The rest of this report will drill down further into how such practices and tools are being used.

4. Metadata Personnel and Training

To understand how staffing supports Metadata Management in the organization, the survey contained questions regarding Metadata, the number of specialists and users, and the process by which people are getting trained in Metadata Management.

“Our Data Scientists and Analysts can understand data only if Metadata and definitions for Metadata are available.”

a. “How many specialists do you have in your organization who regularly works with your Metadata resources?”

While it is clear that Metadata is an important piece of the Enterprise Data Management puzzle, we wanted to know how many people were directly involved in managing the Metadata tools and resources. Recalling the previously stated demographics, a total of 65.79% of respondents’ organizations have over 1,001 employees and 14.87% have over 50,000 employees. 14.01% claimed to have “20+” Metadata specialists, with the number one choice being “2-3” Metadata specialists at 30.43%. So it seems that in many organizations -- of all sizes -- a small, targeted Metadata team is used [Figure 6]:

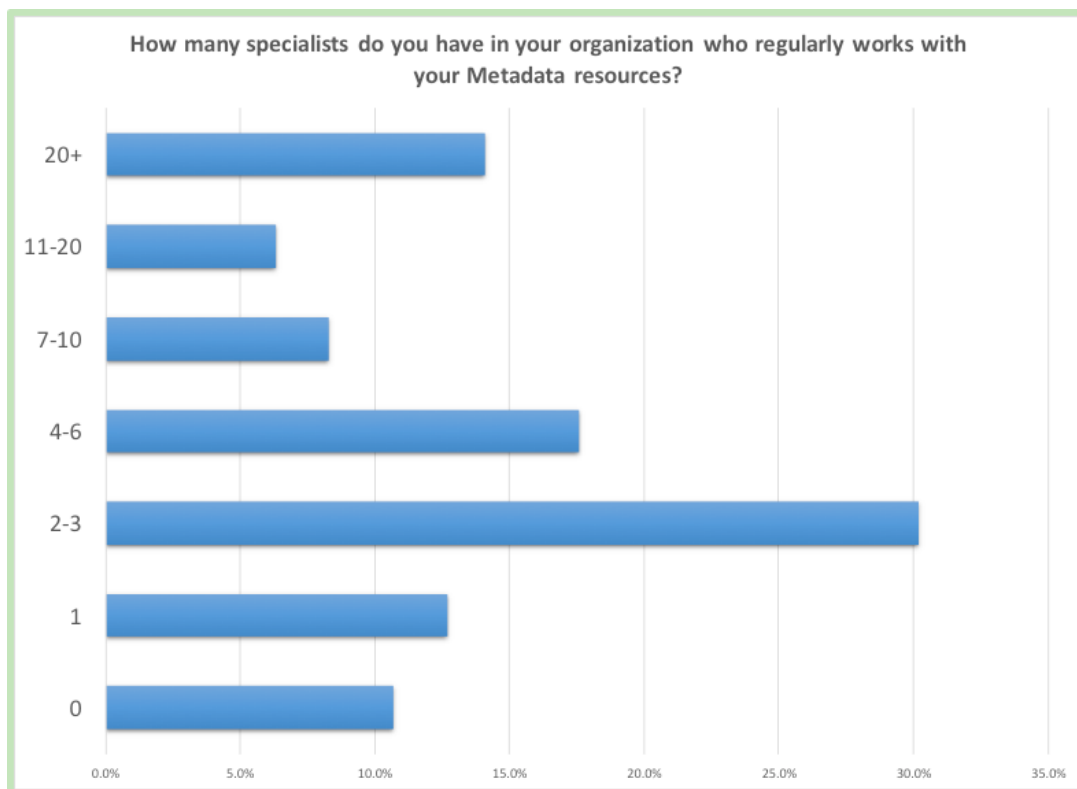


Figure 6: How Many Specialists

b. “Who are the consumers of Metadata in your organization? [Select all that apply]”

Business Users are the largest percentage of Metadata users in most organizations at more than 79%.

The results of this question signify the internal importance of Metadata within organizations. It also helps add context to the results above. The pervading assumption for including this question in the survey was that business users are the primary users of Metadata, along with various data professionals engaged in specific tasks that rely upon Metadata as a primary aspect of their jobs, such as Relational Data Modeling and/or Analytics. We allowed respondents to “Select all that apply” since most organizations have multiple Metadata consumers. The survey results strengthened this conclusion, with the top four choices being [Figure 7]:

- Business Users: 79.13%
- Data Architects/Data Modelers: 69.42%
- BI Reporting Team: 64.56%
- Developers: 62.14%

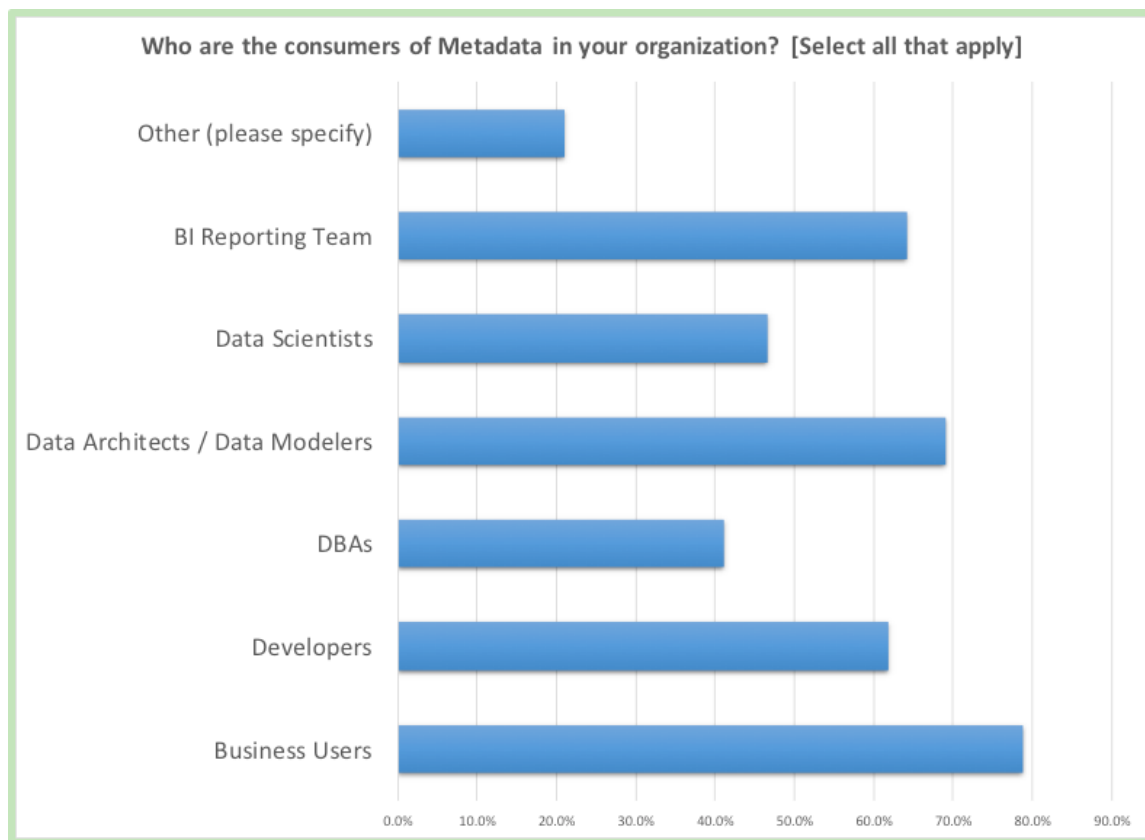


Figure 7: Consumers of Metadata

c. “How do you get trained in Metadata? [Select all that apply]”

Is it necessary to be trained in Metadata Management? How do the various consumers and practitioners learn how to better manage the Metadata resources under their control? Clearly, for any systematic program of Metadata Management to be successful, some sort of training on acceptable processes needs to be completed. How is data duplication dealt with? What are the processes for tracing lineage and data transformation? Who has control of various data assets? What are the escalation procedures? This is where Data Governance and Data Stewardship meets Metadata Management. Both are important elements that help to provide clarity and precision.

With some surprise, the top answer to this question was “No training, just started doing” at 46.38%. Respondents were allowed to “Select all that apply,” so a range of answers were welcome. The other top choices were [Figure 8]:

- Books on Metadata: 40.58%
- Blogs, screencasts, various sources from the Web, not from a tool vendor: 35.27%
- Conferences on Metadata: 30.43%

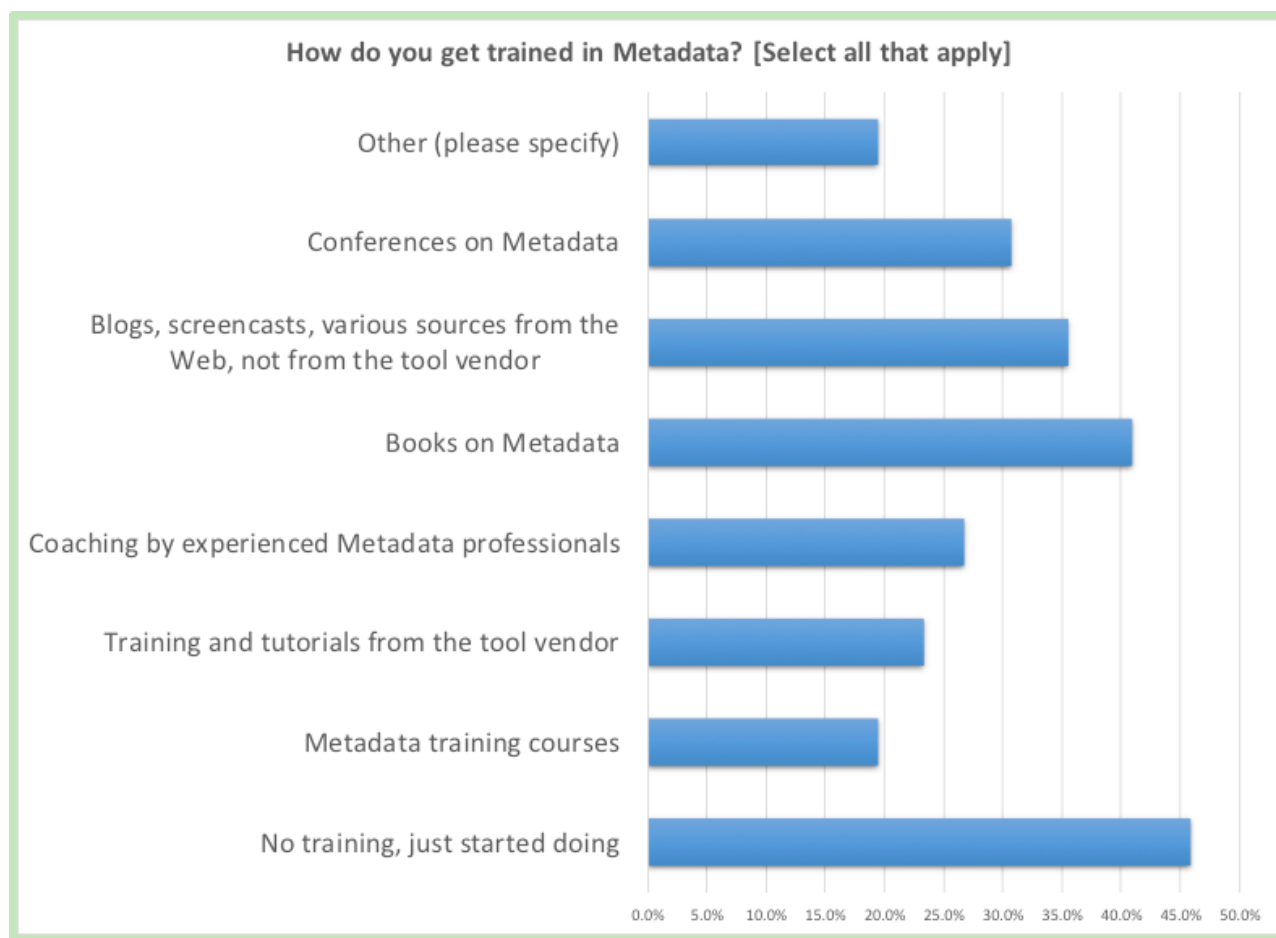


Figure 8: Metadata Training

The survey included a “comments” section for this question to give respondents the chance to give their own answers. Some of the main answers included working with peers and other team members, extensive experience as a data professional (10-20 years), and in-house training sessions.

d. Analysis of Results

There are a number of significant themes evident from the survey results that will continue to be discussed throughout this paper. In this section and from the opening discussion, it's clear that while many organizations have understood the importance of Metadata Management for a number of years, its prominence today is greater than ever before. Metadata is used by all members of an enterprise, whether business or IT, with business users making up the largest percentage. Training in Metadata is a necessary aspect of the entire equation, though it is not keeping pace with the growth of many emergent technologies such as Big Data, Semantics, and others.

5. Metadata Management and Usage within the Enterprise

The business case for Metadata Management is clear: it provides a necessary context around all of the data assets “under management” within an enterprise and thus allows the enterprise to improve efficiency in the respective operations connected with those assets. It has long been recognized that the monumental task of trying to manage all of an organization’s data assets (especially in larger organizations) is often outside of the realm of possibility. A focus on the most crucial assets with a well-formulated roadmap for further developments is a key element of any Metadata Management program. Therefore, the principal section of this report concerns itself with how Metadata Management is being executed within the enterprise, what the primary assets under management are, and what the plans are for future use cases and asset management.

We began by asking respondents if the management of Business or Technical Metadata was their main focus. The primary answer was both at 73.54%, with Business a significant second at 19.05%, and Technical in last at 7.41% [Figure 9]:

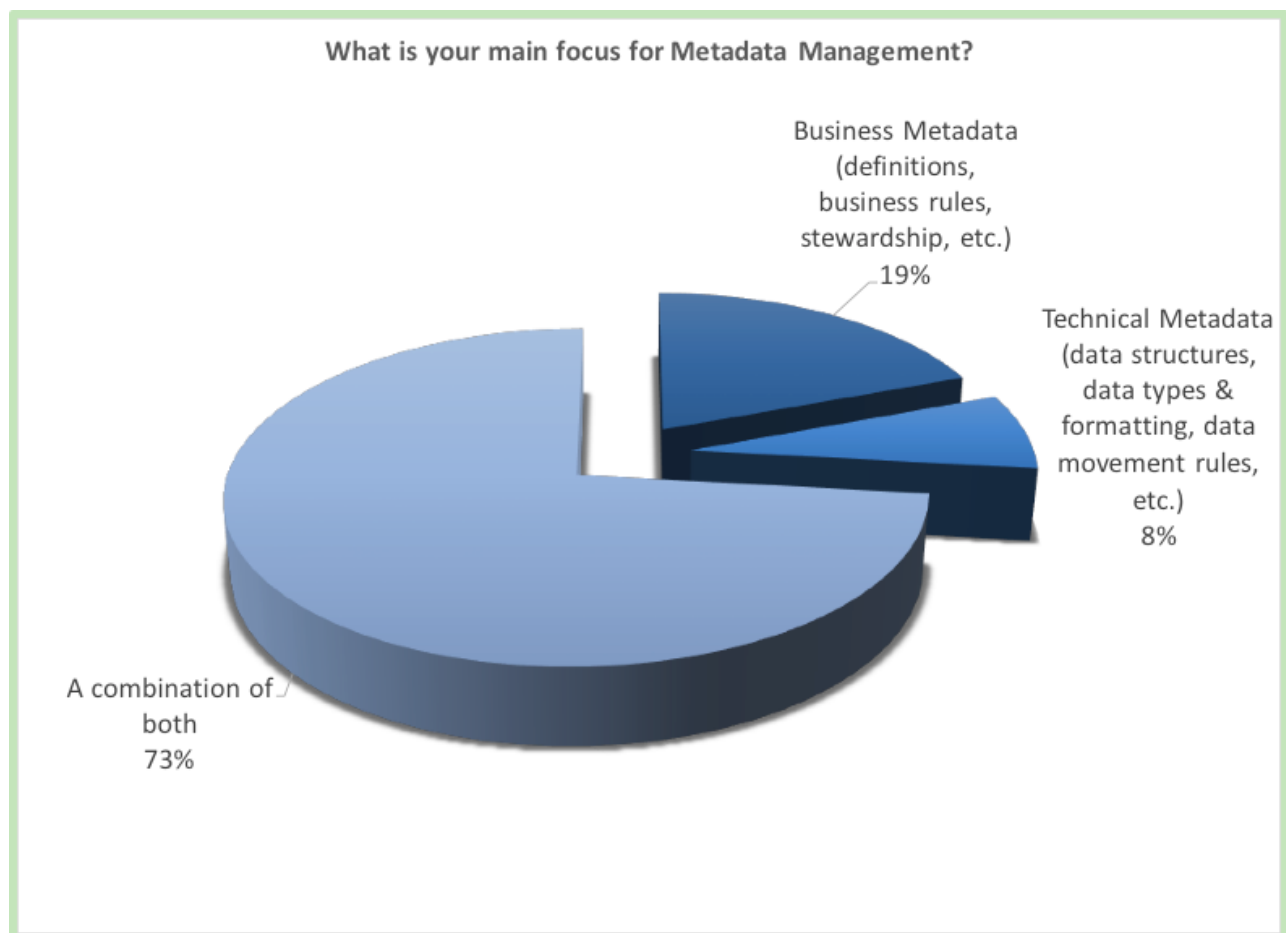


Figure 9: Focus of Metadata Management

a. “For what types of information do you currently manage Metadata? [Select all that apply]”

The top results for this question did not come as a surprise, but rather further clarified what Metadata assets most organizations are currently managing. Of those primary results, Data Warehouse and Relational Databases tied for first at 65.75%, with Data Models at 63.68%, Business Glossaries at 45.79%, BI Tools at 45.79%, and ETL Tools at 42.11%.

Other results of this question show the growth in new technologies and the need to manage Metadata for these systems, as well. Big Data Platforms came in at 20.54%, while Semantic Technologies were at 15.79%, and Open Data garnered 10% [Figure 10]:

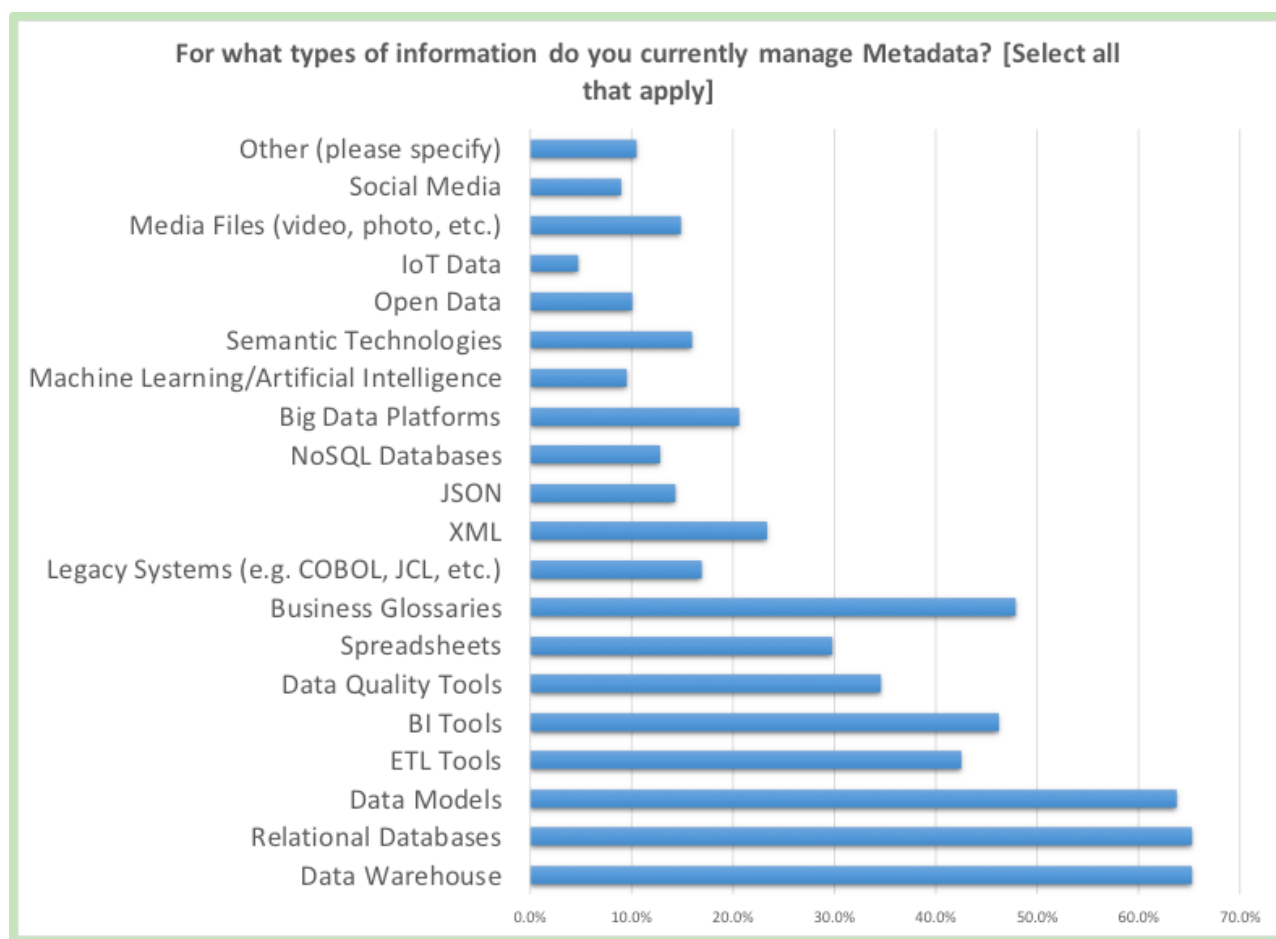


Figure 10: Metadata Assets Currently Managed

b. “For what types of information would you like to manage Metadata for in the future, but are currently unable to do so? [Select all that apply]”

Big Data Analytics and Data Science rank as two of the most anticipated Metadata use cases in the future at more than 80% when combined together.

Asking about future use cases for Metadata Management further shows the growth and importance of emerging technologies such as Big Data Platforms, as well as a continued focus and greater need for more traditional systems such as Data Quality Tools and Business Glossaries. When compared with the question above on current Metadata asset management, we gain a more well-defined picture of what many organizations judge to be most important, with the leading choices being [Figure 11]:

- Big Data Platforms: 38.73%
- Data Quality Tools: 32.95%
- Business Glossaries: 31.21%
- ETL Tools and Data Warehouse (tied): 30.64%

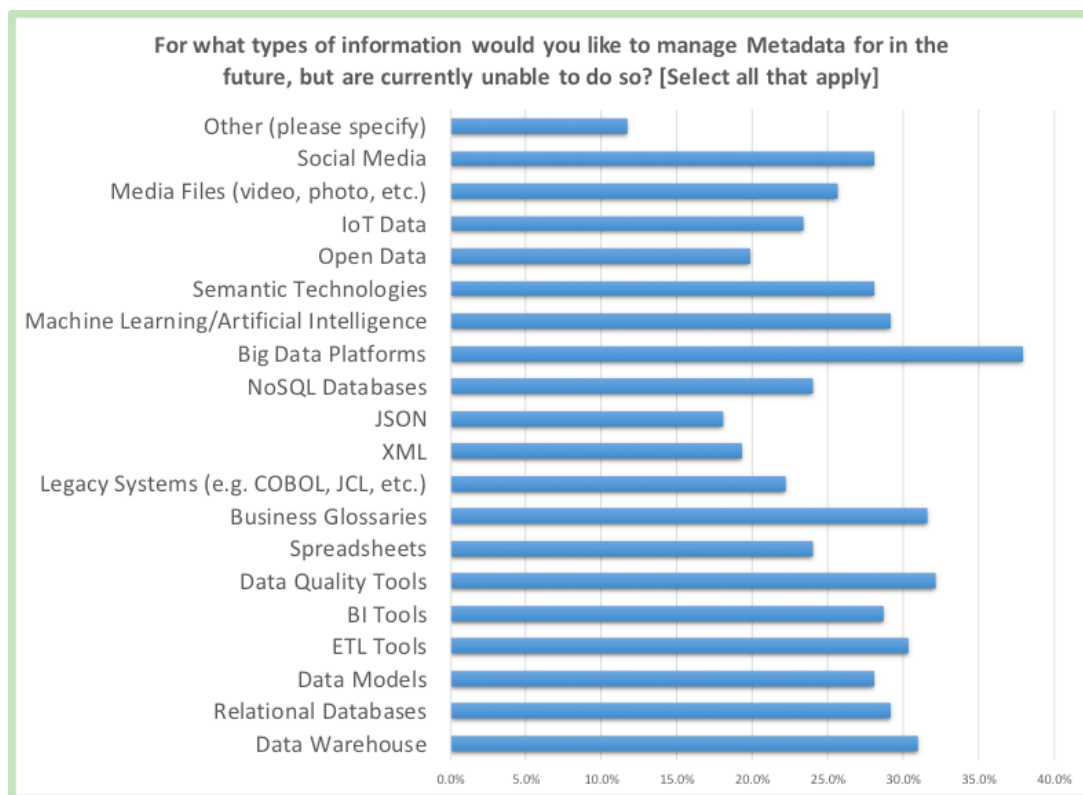


Figure 11: Future Metadata Asset Possibilities

Big Data is clearly in the minds of many organizations, with a little over 20% already working on managing their Big Data resources and close to 40% wanting to in the future. Some of the other newer trends and technologies, such as Machine Learning/Artificial Intelligence (28.90%), Semantic Technologies (27.75%), IoT Data (23.12%), Open

Data (19.65%), and NoSQL Databases (24.28%), are also certainly important concerns in terms of future Metadata Management for many enterprises. But nonetheless, the fact that ETL Tools (30.64%), BI Tools (29.48%), Data Quality Tools (32.95%), Data Warehouses (30.64%), Business Glossaries (31.21%), and Data Models (27.75%) still receive some of the highest responses demonstrates that while many organizations are looking to newer trends, they also must continue focusing on the “core” of Metadata Management.

c. “What are your current main Use Cases for Metadata? [Select all that apply]”

The results of this question reinforce our initial hypotheses about where Metadata Management currently exists within the majority of organizations practicing it. Conversations about Metadata coincide with conversations about Data Governance, Data Quality, Master Data Management, and Data Warehousing – such are the conventional use cases. Those also occupy the main responses to this question [Figure 12]:

- Data Governance: 63.83%
- Data Warehousing & BI Reporting: 53.19%
- Data Quality Improvement: 52.31%

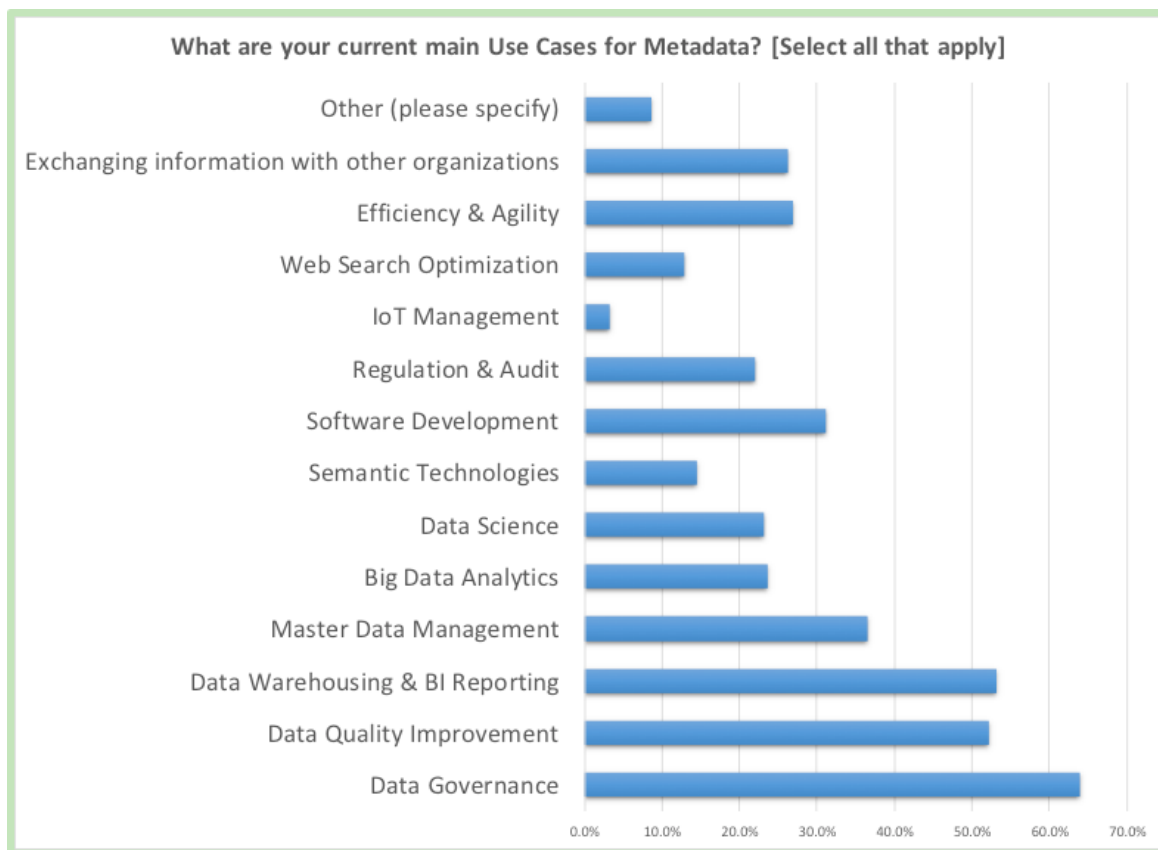


Figure 12: Current Metadata Use Cases

Some of the other significant use cases that deserve attention are Big Data Analytics (23.40%), Data Science (22.87%), Software Development (31.38%), Regulation & Audit (21.81%), and exchanging information with other organizations (26.60%).

There is a solid mix of use cases represented here. Data Governance and Data Warehousing are typically the most common, but organizations are branching out into other areas as well, showing again the diversity of Metadata Management across the entire enterprise.

A point of note: Efficiency & Agility received 27%, demonstrating that Metadata helps to improve efficiency and can be used in Agile development practices. It is also interesting to note that legacy systems (e.g. COBOL) are still a consideration with 17% of respondents managing legacy Metadata.

Data Governance is the most prevalent current Metadata use case, along with Data Quality Improvement, Data Warehousing, and BI Reporting.

The diverse responses from traditional systems to emerging technologies to legacy systems show the challenge and complexity of Metadata Management in today's organizations.

d. “What are the future planned Use Cases of Metadata within your organization (that you are NOT doing today)? [Select all that apply]”

Figure 13 below underscores what some of the earlier figures in this section of the report have already made clear – Big Data is firmly entrenched in the thoughts and plans of many organizations, coming in as the number one future planned Metadata use case at 43.79%. Other significant percentages include Master Data Management (41.42%), Data Governance (32.54%), Data Science (36.69%), and Data Quality Improvement (31.36%):

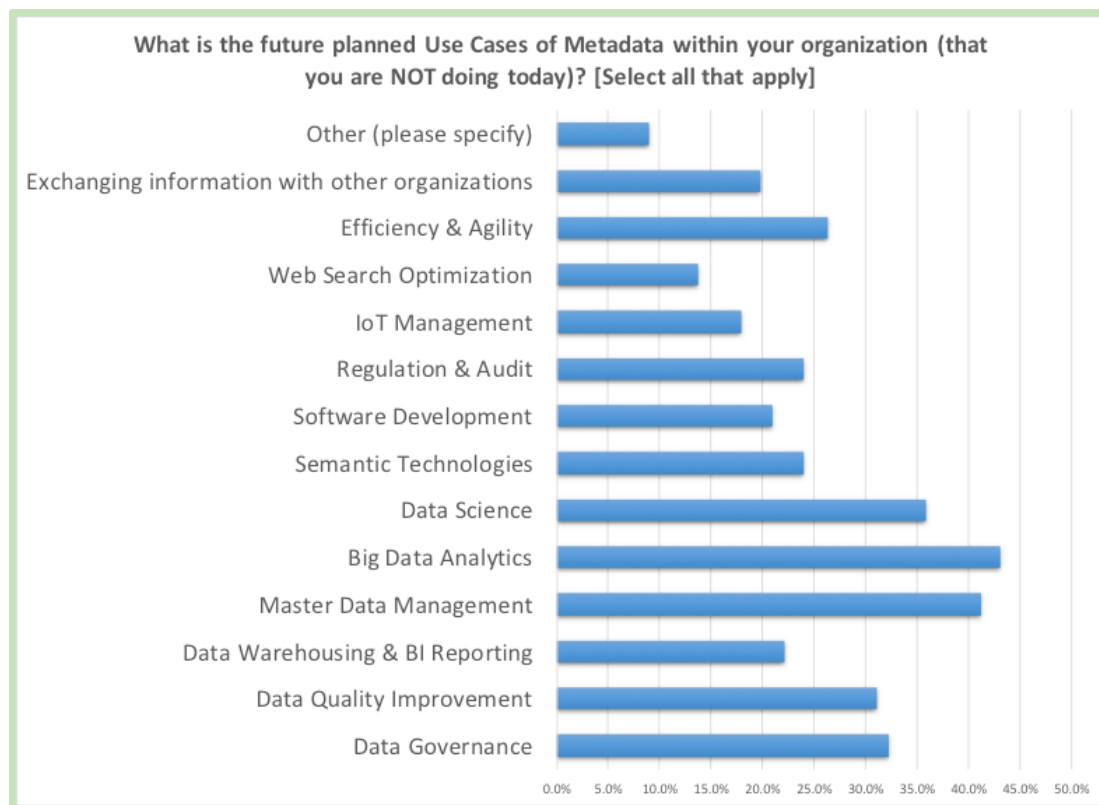


Figure 13: Future Planned Metadata Use Cases

Big Data Analytics is a huge driver and also a huge challenge for many organizations. As discussed above, many of the newest tools/technologies do not currently have good Metadata Management options available, though organizations are still moving forward with their adoption at ever larger scales. The ramifications of such a clear lack of options are currently unclear but will certainly have long-ranging effects across the entire Data Management spectrum for years to come.

This question broke Big Data and Data Science into separate categories, though in practice both are necessary components of a larger whole. Therefore, if we put them together the percentage is considerably greater -- over 75%.

It's clear that enterprises are continuing to increase their adoption of traditional use cases as well, since Data Governance, Master Data Management, Data Warehousing, Data Quality, and others still have

substantial numbers. And legacy systems remain a clear priority with 20% of respondents looking to add management for these systems in the future.

e. “What are your Use Cases for showing relationships between Metadata objects? [Select all that apply]”

A big value of Metadata is in showing the relationships between key data. These use cases delve into which types of relationships organizations are interested in. The final question in this section took a broader approach to dealing with use cases. The top three answers were [Figure 14]:

- Data Lineage/Audit Trails: 75.40%
- Impact Analysis/Where Used: 61.50%
- Data Model Design Layers: 54.55%

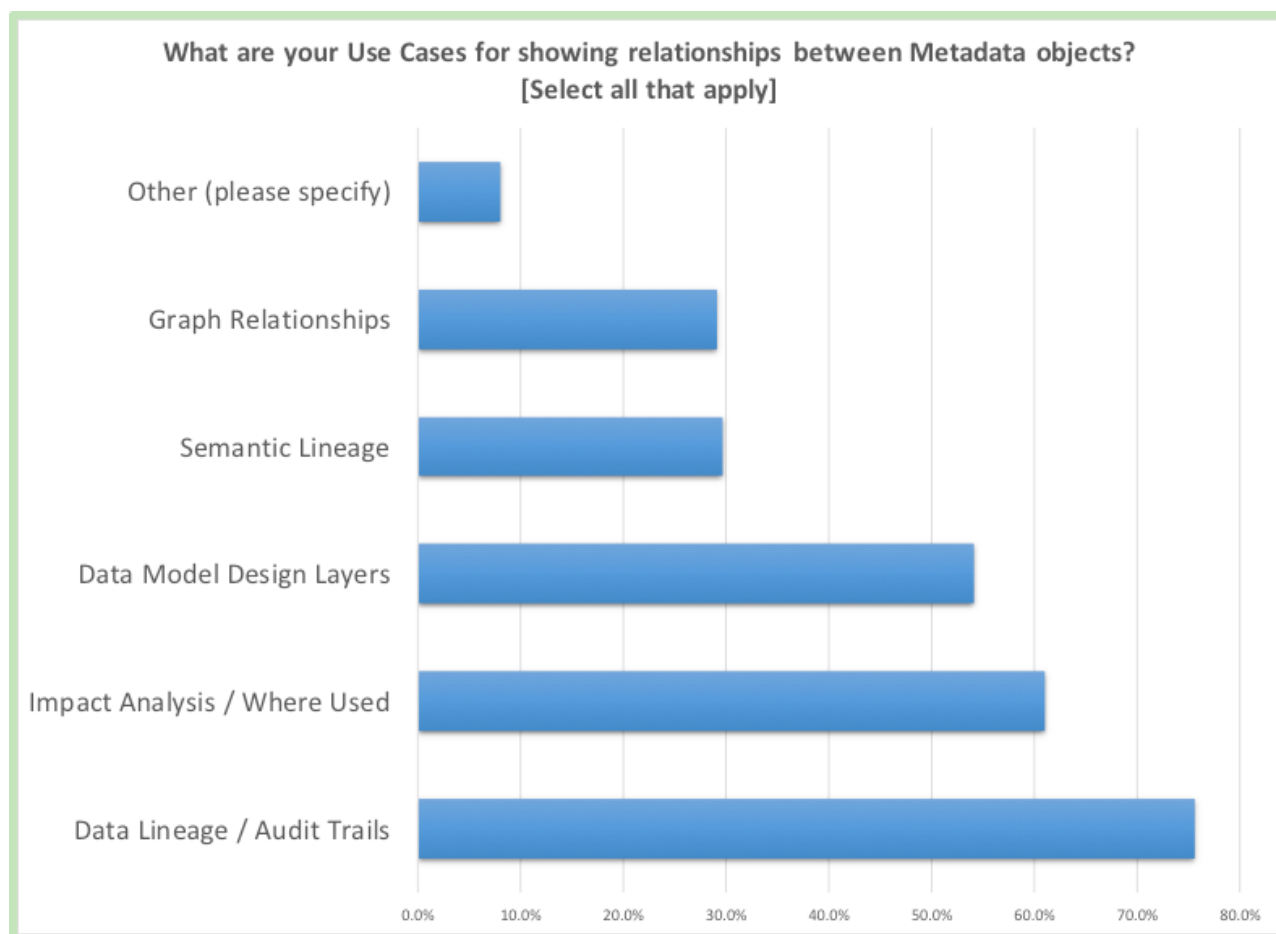


Figure 14: Metadata Object Relationship Use Cases

These results fit well into what was said about the use cases and business drivers. Data Lineage & Audit fit with Data Governance, as does Impact Analysis. Data Model Design Layers fit with aligning business and IT. Impact Analysis, Data Lineage, and Data Model Layers fit with Data Warehousing. Graph Relationships fit with some of the newer NoSQL technologies and Social Media, and support earlier findings in that they are still only beginning to see greater adoption.

f. “Do you have a defined Metadata Strategy in your organization?”

The ability to ensure actionable, consistent, and relevant control of an enterprise’s data ecosystem requires a well-formulated Data Strategy. Such a strategy provides a foundation for all enterprise efforts, helps to unite IT and business goals, creates a roadmap for future success, defines expectations, and gives a cohesive vision that everyone can understand. Comprehensive Metadata Management necessitates such a strategy and is a crucial element for its success, thus a defined Metadata Strategy should be a primary feature of an enterprise’s overall Data Strategy.

This question gave respondent’s four choices to choose from and helped bring some clarity to the overall picture of where Metadata exists within many organizations. Do organizations have a well-defined Metadata Strategy? Or are their practices defined only for some use cases? The answers are telling [Figure 15]:

- We have an ad-hoc, loosely defined Metadata Strategy for certain use cases: 35.87%
- We have a Metadata Strategy that is a subset of a larger Enterprise Data Strategy: 26.09%
- No, we do not have a Metadata Strategy: 24.46%
- Yes, we have a clearly defined Metadata Strategy: 13.59%

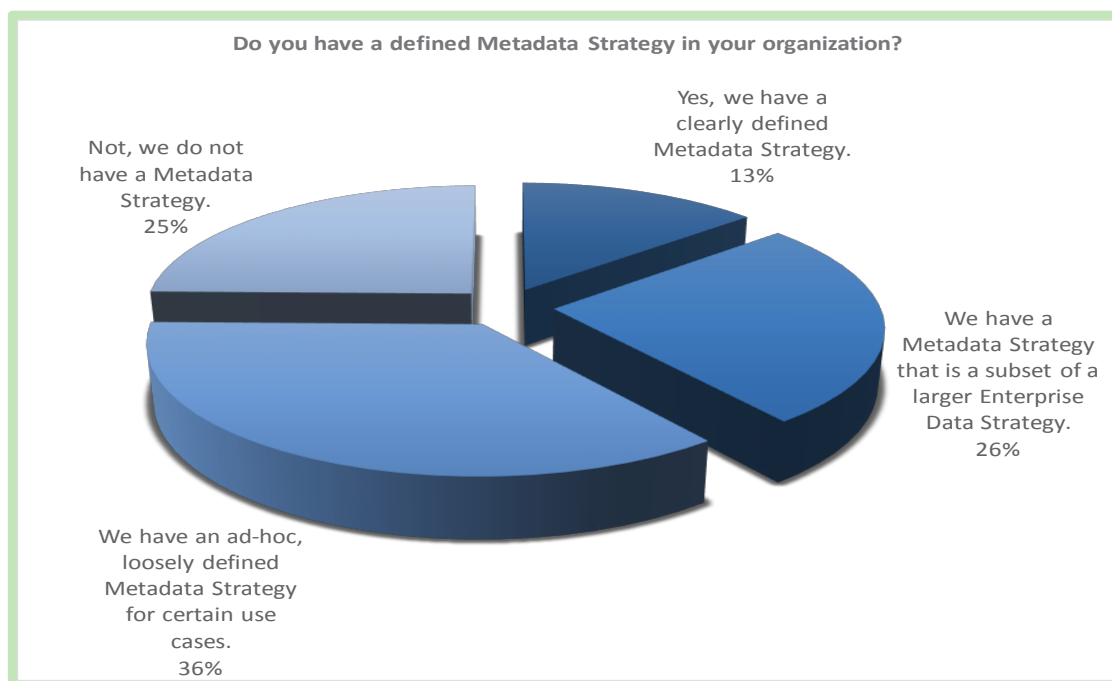


Figure 15: Defined Metadata Strategy

A main takeaway from this question really ties into some of the primary themes of this report. On one side, Metadata Management is critical to the overall organization, and over 25% have embedded their Metadata planning as part of a larger enterprise strategy. At the same time, given the wide array of use cases for Metadata, many organizations have taken a more ad hoc approach (over 25%). To support these various use cases, approximately 14% have a dedicated Metadata Strategy to support the existing and emerging Metadata trends.

Concerning, however, is the fact that 25% of organizations do not have a defined Metadata Strategy in any form, which is not consistent with the majority opinion that Metadata is critical to the organization. With the ever-changing technical and business landscape, many organizations perhaps do not know where to begin.

g. Analysis of Results

The wide range of responses shows the complexity of the technical landscape for the majority of organizations. While a high percentage of Metadata Management focuses on the traditional heavy-hitters such as Data Warehousing and relational databases, new technologies such as Big Data and graph databases are also a high priority. It is interesting to note that legacy systems (e.g. COBOL) are still a considerable priority and challenge to organizations with 17% of respondents managing legacy Metadata currently and 20% expressing a need to add support for legacy systems in the future, perhaps as “human metadata” leaves the organization in the form of retirees. The breadth of technologies highlights the complexity of enterprise-scale Metadata Management and the challenges faced by Metadata tools, standards, and technologies supporting them, which we’ll discuss in the next section.

6. Metadata Tools, Functionalities, and Standards

The sheer number of data sources available to modern enterprises today makes the management of Metadata a difficult task: from legacy, transactional, and relational databases systems to Big Data platforms to Semantic resources and sensor data to BI & Reporting data, and the list could go on for pages. This is the challenge in selecting a toolset to manage Metadata within an organization. The creation of in-house Wikis helps some, while commercial Metadata Repositories, Business Glossaries, and Data Dictionaries have become advantageous solutions for others. In many cases, combinations of these tools work together.

The survey evaluated the current use of Metadata tools and their most important functionalities with three questions.

a. “Which of the following Metadata Solutions does your organization currently use? [Select all that apply] “

The results show the varied applications in use for enterprise Metadata Management today. We allowed the respondents to “Select all that apply” for this question since many organizations employ more than one option and we wanted to get the most comprehensive viewpoint as possible. There was a significant difference between the top four choices, which all had comparable numbers in the 45-55% range, and the less chosen options with ranges from 4-11%. The top choices by far were [Figure 16]:

- Tool specific repository (Data Modeling, ETL, etc.): 53.26%
- Metadata Repository: 50.54%
- Business Glossary: 46.74%
- Homegrown Solution (Wiki, SharePoint, etc.): 45.65%

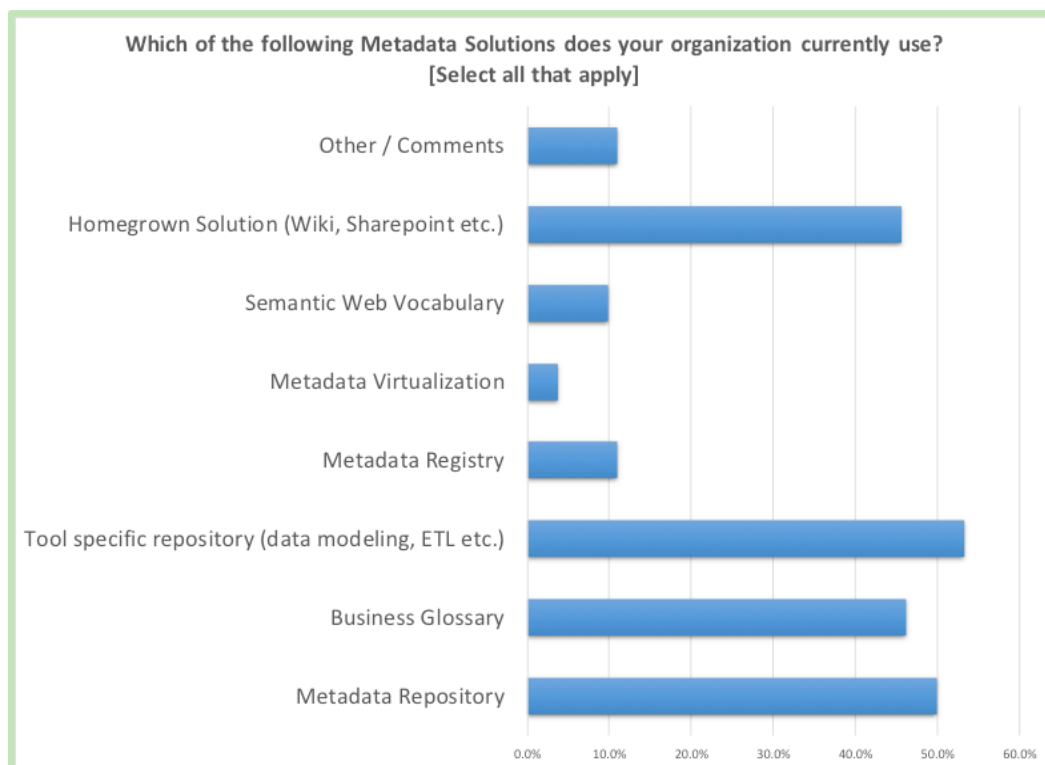


Figure 16: Metadata Solution in Current Use

The mix of solutions in use shows the variety of needs/ technologies/audiences that must be addressed. Often Metadata is associated with a central Metadata Repository, but clearly that is not the only solution being represented. Organizations are employing a range of tool choices and, as stated by many respondents, want more of them.

“Metadata is critical as it provides the common lexicon by which the entire organization can converse. Without it, there can be only rudimentary knowledge sharing.”

b. “How do you feel about the quality of current tools for Metadata Management available in the market today?”

Are the available tools sufficient for the needs of today’s typical organization? As was discussed above, use cases such as Big Data Analytics, Semantic Web Technologies, Data Science, Regulation & Audit, and more traditional use cases such as Master Data Management, Data Quality, and Data Governance are all within the scope of current needs and future planning.

Metadata Search, Storage Repository, and Interchange with other tools ranked as the three most important tool functionalities.

Although most respondents showed a net positive reaction to the currently available Metadata Management toolsets, the responses showed that there are gaps in supporting the rapidly-changing technical landscape. The leading answers were:

- They work well for some functionality, but don’t address the new technology options available: 26.78%
- They work well, and meet the majority of our needs:19.13%
- They are too cumbersome and expensive for what we need to do: 15.3%
- They don’t meet our current Metadata needs: 10.93%

It should be noted that, perhaps because of the recent growth in Metadata Management and new users entering the field, many respondents were not aware of available Metadata Management tools at all (13.8%) [Figure 17]:

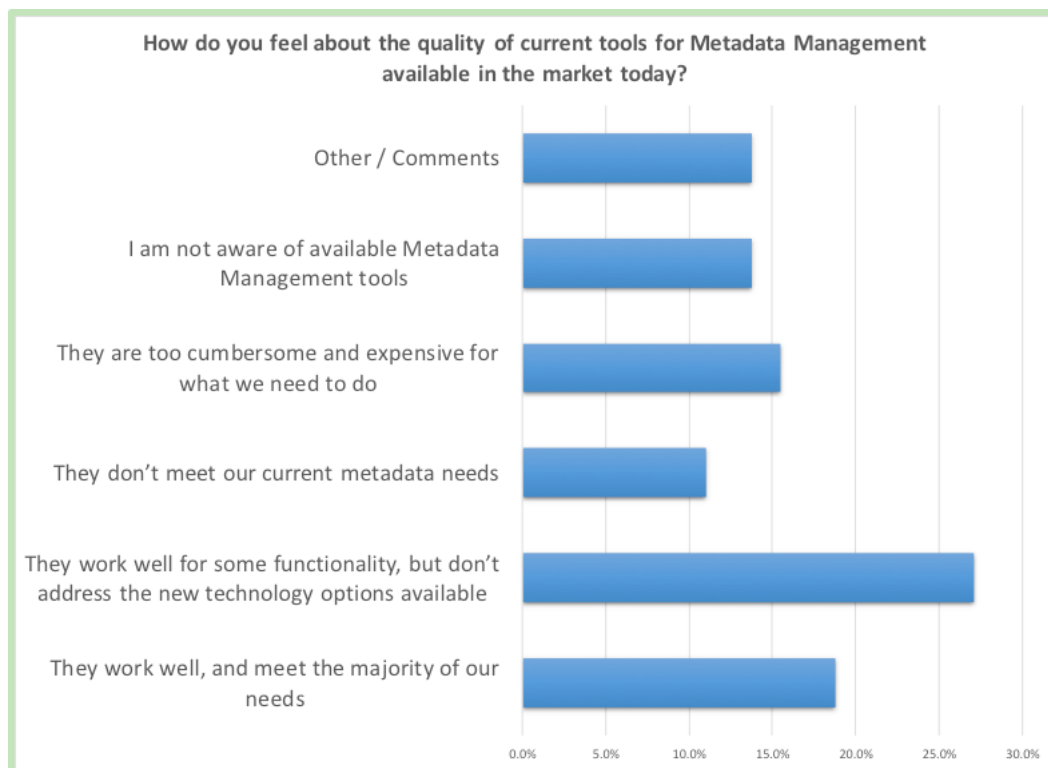


Figure 17: Metadata Tool Quality

The survey allowed respondents to add their own comments to this section. The most prevalent comments mentioned both cost issues and a lack of comprehensive functionality. Some of them include:

- “The comprehensive tools are too expensive and the affordable ones are too limited to a particular technology, e.g. tied to a specific ETL platform.”
- “Seems like you would need a variety of different Metadata tools – depending on your use case – for example some tools are stronger in Data Governance, some in Data Catalog functionality, some (especially those from vendors who also sell ETL tools) tend to be stronger in end to end lineage and incorporating Data Quality Metadata.”
- “No tool can do everything well. All tools should support open, flexible metamodels; integration capability; and exit ease.”
- “This is not a mature market. Very few tools support a broad set of use cases. There is a lot of manual entry required. The best tools support primarily IT functions or primarily business functions.”
- “I would love to see a ‘1 stop shop’ tool that also allows us to create data models within the Metadata Repository.”

The need to address multiple sources and audiences is evident here again, as is the necessity for flexibility and agility. A number of respondents commented about the ability to support flexible metamodels/models and the constant requirement of user customization. Integration was another key theme—the ability to share common Metadata across tools, technologies, and users was a constant concern for many respondents.

c. “Which Metadata tool functionality is important to you? [Select all that apply]”

This question of the survey was used to bring the entire issue of tools and functionalities together. Respondents were given the choice of selecting many answers since there are numerous different functionalities required from a given Metadata tool. Of those, the top four choices were [Figure 18]:

- Metadata Search: 79.01%
- Storage/Repository/Database: 69.06%
- Interchange with other tools: 55.80%
- Web-based Portal: 54.70%

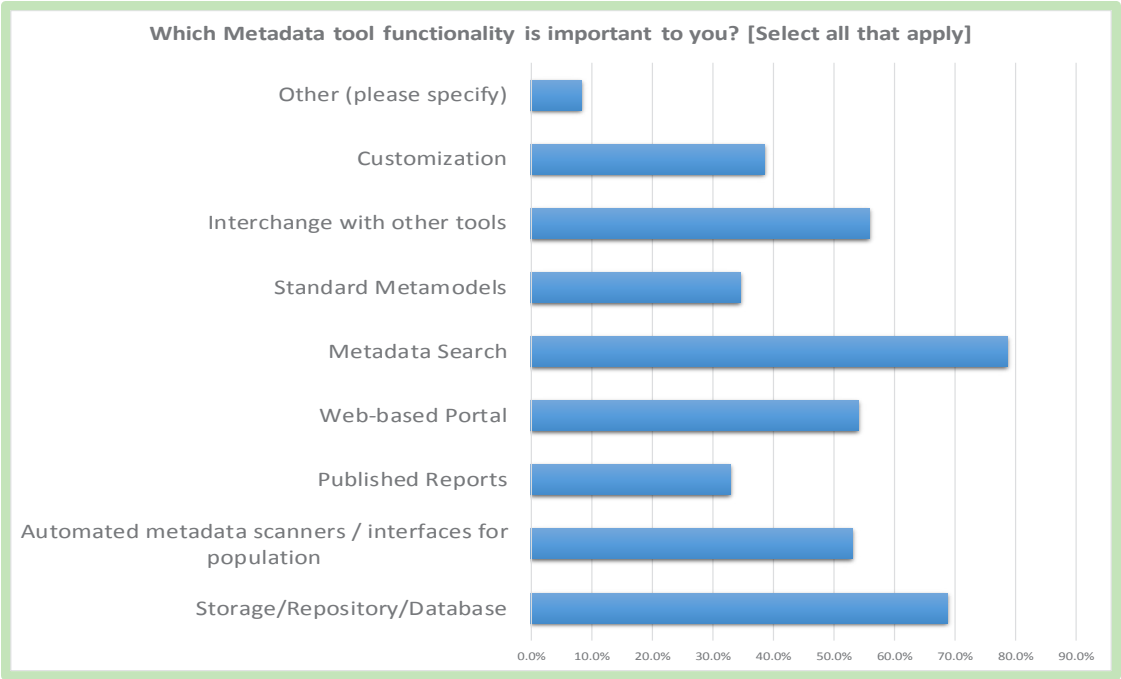


Figure 18: Metadata Tool Functionality

d. “Are you using any Metadata Standards in your organization? [Select all that apply]”

The use of Metadata Standards is also an element that can improve consistency through the adoption of common practices for the sharing and integration of Metadata from a variety of sources. Metadata Standards are certainly going to be more relevant for certain industries and use cases. Some of the many published and commonly used Metadata Standards are:

- **Common Warehouse Metamodel (CWM):** Used for Data Warehousing, Relational Databases, ETL, etc.
- **Open Data:** Employed especially within government and scientific institutions. Open Data can be freely used and redistributed by anyone.
- **Dublin Core:** Provides a common Metadata Standard for various resources such as libraries and media.
- **Schema.org:** A Semantic Web vocabulary that helps webmasters and websites mark-up their pages in a shared format.
- **International Organization for Standardization (ISO):** Publishes numerous standards for Metadata such as ISO/IEC 11179 for data exchange, ISO 19115 for geo-spatial data, and many others.

Considering the demographics of the survey respondents, we would presume that a number of organizations would be using some sort of Metadata Standard. 11% of respondents were in Government, 10% were in Insurance, 9% in Healthcare, 7% in Banking and Finance, among other fields. However, this was not the case. A total of 53.41% said “We are not using any Metadata Standards.” The other top answers were Dublin Core (13.07%) and ISO 11179/Data Exchange (9.09%) [Figure 19]:

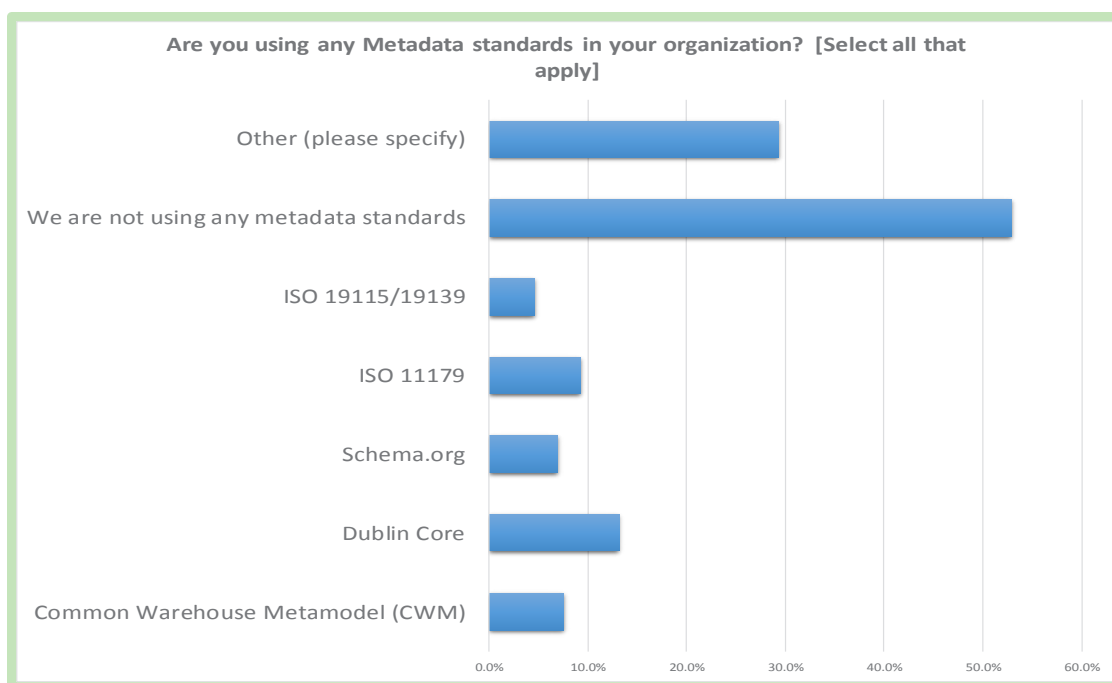


Figure 19: Metadata Standards

The survey concluded by asking respondents the open-ended question: “Do you feel that there industries, subject areas, or technologies that are currently lacking a Metadata Standard that you’d like to see?”

The main answer was simply “no” with well over 50% of respondents giving that answer, along with “I’m not sure.” The rest of the answers had a wide-range, with some of the primary choices being:

- Banking and Finance
- Energy and Utilities
- Big Data
- Internet of Things
- Healthcare

Many respondents mentioned that they were using their own standards or customizing existing models. One interesting comment was, “We prefer a flexible Metadata framework that could represent any standard.”

e. Analysis of Results

The responses demonstrate the complexity of the data environments within many enterprises today and the variety of functions required from search to storage to interchange. Respondents repeatedly commented that they seek better integration and interchange between tools, especially in terms of metamodels, and exchange between users, tools, and standards. The fact that the adoption of common standards is not keeping pace with the growth of Metadata sources is an issue that needs to be addressed, as well. With the rapidly-changing data landscape, these challenges will continue for both Metadata Management providers and the users of these systems.

7. Conclusion

The importance of Metadata Management to organizational success is obvious. It aids in the critical alignment of business and IT resources and is a foundational element in leveraging an enterprise's information assets.

Current enterprise Metadata Management drivers include many of the traditional or core Data Management practices, including Data Governance, Data Quality, Data Modeling, Master Data Management, Business Intelligence, and Data Warehousing. Yet, while many of these practices have been the focus of organizations for many years, it is clear from this report that they remain at the forefront of development and discussion – large percentages of organizations are still implementing them with varying success.

The emergence of additional use cases built around new technologies is only making integration more complex. Big Data Analytics, requirements for new regulation and audit mechanisms, Semantic Technologies, Data Science, IoT, Advanced Analytics, and a seemingly endless assortment of other new technologies are magnifying the challenges of integrating, implementing, and managing data assets into existing architectures. The drivers for these newer technologies are just as important as they are for the traditional technologies. Governing both is a requirement that will only increase moving forward.

With the rise of the data-driven organization and the myriad technologies that support it, Metadata Management will continue to grow in importance. Business users are helping to lead this growth as they see the value of information in context and the governance and audit that Metadata can provide. With the rise of Self-Service BI & Analytics, more and more business users are getting closer to the information that drives their business and are aware of the need for quality, compliance, and context. Metadata Management is a critical practice that helps align business drivers with the numerous technical data platforms that support them. This survey highlights the growing demand for Metadata as well as the challenges in supporting the ever-changing data landscape.

About the Authors



Donna Burbank, is a recognized industry expert in information management with over 20 years of experience in data management, metadata management, and enterprise architecture. She currently is Managing Director of [Global Data Strategy Ltd](#), an international data management consulting company. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, and business leadership.

Donna has extensive experience in metadata management, beginning her metadata career as an international consultant for the PLATINUM metadata repository. After working as a metadata consultant for many years, she moved to a product management role with the CA Metadata Repository solutions, defining product direction based on customer & market needs. As the director of product management at Embarcadero Technologies, she was responsible for the creation of a repository-driven, metadata-driven data modeling solution that encompassed both data and process modeling. Donna has also worked for many years with the CA ERwin solution as the VP of product marketing for data modeling. Active in industry organizations, Donna was a part of the review committee for the Object Management Group's Information Management Metamodel (IMM) and a member of the OMG's Finalization Taskforce for the Business Process Modeling Notation (BPMN). In all of these roles, metadata was a key component.

She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored several books including: [Data Modeling for the Business](#), [Data Modeling Made Simple with ERwin Data Modeler](#), and was a contributor to Metadata Solutions. She can be reached at donna.burbank@globaldatastrategy.com and you can follow her on Twitter @donnaburbank.



Charles Roe has been a professional freelance writer and copy editor for more 15 years, and has been writing for the Data Management industry since 2009. He is the founder of CRSScribes.com, his own writing/editing business. Charles has written on a range of industry topics in numerous articles, white papers, and research reports including Data Governance, Big Data, NoSQL technologies, Data Science, Cognitive Computing, Business Intelligence & Analytics, Information Architecture, Data Modeling, Executive Management, Metadata Management, and a host of others. Charles is backed with advanced degrees in English, History, and a Cambridge degree in Language

Instruction. He has worked for almost 20 years as an instructor of English, History, Culture, and Writing at the college level in the USA, Europe, and Turkey. He writes creatively in his spare time.

Sponsors



technologies

ASG Technologies brings peace of mind to every enterprise with information access, management and control for our customers. These solutions empower businesses to enhance workforce productivity, gain an accurate and timely understanding of the information that underpins business decisions and address compliance needs with improved visibility of cross-platform data from legacy to leading edge environments. More than 70 percent of global Fortune 500 companies trust ASG Technologies to optimize their IT investments. ASG is a global provider of technology solutions with more than 1,000 people supporting more than 4,000 midmarket and enterprise customers around the world. For more information, visit www.asg.com.



Data Advantage Group™ is a leading provider of enterprise metadata management and data governance solutions. Data Advantage Group's MetaCenter® platform enables organizations to govern their information assets while lowering costs, improving agility and reducing operational risks. The MetaCenter Platform's unique loosely coupled architecture delivers 3X the ROI of competing solutions by dramatically lowering the cost of ownership while enabling superior long-term success rates.

Founded in 1999, Data Advantage Group has been consistently delivering value to its customers through the MetaCenter® Platform for over 15 years. The MetaCenter® Platform's design has been driven by the specific needs of Data Advantage Group's customers for a cost effective and easy to use enterprise class metadata management and data governance solution.

