The State of Information and Data Quality 2012 Industry Survey & Report

Understanding how Organizations Manage the Quality of their Information and Data Assets

By

Elizabeth Pierce, C. Lwanga Yonke, Piyush Malik, and Chitra Kagathur Nagaraj

International Association for Information and Data Quality (IAIDQ)
University of Arkansas at Little Rock, Information Quality Program (UALR-IQ)

November 2012

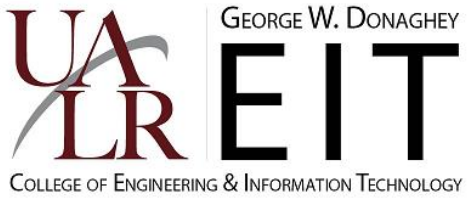


Table of Contents

Authors and Sponsors

Authors

Elizabeth Pierce is an Associate Professor of Information Science at the University of Arkansas at Little Rock (UALR). She joined the faculty in August 2006 as part of UALR's efforts to launch the first Master of Science in Information Quality. Prior to coming to UALR, Elizabeth Pierce taught for 11 years at Indiana University of Pennsylvania. She holds a PhD in Statistics & Management Science from the University of Michigan, a Master of Science in Computer Science from Iona College, and Bachelor of Science degrees in Quantitative Business Analysis and Mathematics from Penn State. She can be reached at expierce@ualr.edu.

C. Lwanga Yonke is a seasoned information quality and information management expert and leader. He has successfully designed and implemented projects in multiple areas including information quality, data governance, business intelligence, data warehousing and data architecture. His initial experiences were in petroleum engineering and operations. An ASQ Certified Quality Engineer, he holds an MBA from California State University and a Bachelor of Science in petroleum engineering from the University of California at Berkeley. Lwanga Yonke is a founding member of the International Association for Information and Data Quality (IAIDQ) and currently serves as an Advisor to the IAIDQ Board. He can be reached at Lwanga.yonke@iaidq.org.

Piyush Malik focuses on strategic initiatives and emerging technologies as part of IBM's Global Business Analytics & Optimization consulting practice. Currently he leads their Big Data Services Center of Excellence. Specializing in Information management strategy and architecture, information quality and governance, business intelligence, master data and advanced analytics, he has over twenty years of international consulting, practice building, sales and delivery experience with Fortune 500 clients across industries. He has served as the founding director of IBM's Global BAO Center of Competency and previously led the Information Integrity consulting practice at PricewaterhouseCoopers Management Consulting Services prior to its acquisition by IBM in 2002. Piyush has served on the Board of Directors of IAIDQ since 2008 and has been a charter member since its inception. He holds a Bachelor's degree in Electronics & Communications Engineering and a Master's degree in Management of Technology from IIT, Delhi. He can be reached at Piyush.Malik@iaidq.org. or via Twitter@pmalik1.

Chitra Kagathur Nagaraj recently completed her Master of Science in information Quality at the University of Arkansas at Little Rock. She has a bachelor's degree in Electronics and Communication from Visvesvaraya Technological University, Karnataka, India. Chitra has several years of IT experience, including work as a Quality Assurance analyst in India. Her interests include database, data warehouse and embedded systems.

About IAIDQ

Chartered in January 2004, the International Association for Information and Data Quality (IAIDQ) is a not-for profit, vendor-neutral professional society, and the only international professional organization for information and data quality practitioners. IAIDQ is advancing the quality of information and data around the world by building a community that supports learning and sharing knowledge for the benefit of all information consumers. IAIDQ offers a wide range of products and services to meet the needs of information/data quality professionals at all experience levels. To learn more about IAIDQ, please visit the website: iaidq.org.

About the Information Quality Program at UALR

The world's first graduate degree program in Information Quality was created by the University of Arkansas at Little Rock in 2006. The program was designed in collaboration with the MIT Information Quality program directed by Dr. Richard Wang, and a significant portion of the start-up costs were underwritten by Acxiom Corporation under then CEO Charles Morgan. The Information Quality Program at UALR (UALR-IQ) is housed in the Department of Information Science in the Donaghey College of Engineering and Information Technology. The purpose of the UALR-IQ Program is to meet the growing demands from academia, industry, and government for qualified professionals with graduate-level degrees (Certificate, Master's and PhD) that encompass the theory, principles, and practices essential to advancing the discipline of Information Quality and closely-related areas such as entity and identity resolution, identity management, information integration, governance, risk assessment, compliance, and enterprise architecture. To learn more about the IQ Program at UALR, please check out the website: http://ualr.edu/informationquality.

About the IAIDQ Industry Report Series

This report is part of a series of industry reports jointly produced by IAIDQ and the Information Quality Program at UALR (UALR-IQ). Earlier reports in this series include the following:

- The Job of the Information/Data Quality Professional (February 2011) authored by C. Lwanga Yonke, Christian Walenta, and John R. Talburt. Available at http://iaidq.org/publications/yonke-2011-02.shtml.
- Information/Data Quality Salary and Job Satisfaction Report (July 2009) authored by Elizabeth Pierce, C. Lwanga Yonke, and Apolonia Lintag. Available at http://iaidq.org/publications/pierce-2009-07.shtml.
- The State of Information and Data Governance: Understanding How Organizations Govern their Information and Data Assets (April 2008) authored by Elizabeth Pierce, Wendell Scott Dismute, and C. Lwanga Yonke. Available at http://iaidq.org/publications/pierce-2008-04.shtml.

Acknowledgements

The authors and the IAIDQ research partners would like to thank all the people who contributed to the 2012 State of Information/Data Quality Industry Report. We greatly appreciate and thank the many individuals who completed our survey. We also specifically express our sincere gratitude to the many people who volunteered valuable guidance and insight during the various phases of this project, including Abhijit Dev, Jeff Monica, Grant Robison, Ken Self and Rachele Williams.

Executive Summary / Key Findings

This report details the findings of a Spring 2012 survey conducted by the UALR-IQ Program and IAIDQ. The purpose of the survey was to understand how organizations are managing the quality of their information and data assets. The results are based on the responses of 270 individuals who completed the survey. Almost half of the respondents (49%) based their answers on their entire organization while the rest based their answers on a functional area (25%), a department (15%) or a subsidiary of the organization (11%). As a result, throughout this report, the term "organization" may refer to an entire enterprise, or it may refer to a portion of an enterprise.

Trends in Information and Data Quality

- Nearly half the survey participants rated the effectiveness of their information/data quality efforts as "OK-some goals are met" (46%), followed by "good - most goals are met" (22%) or "poor – few goals are met" (20%).
- Though organizations used multiple labels when referring to the activities associated with managing the quality of information and data assets, the term most often cited by respondents was Data Quality Management (36%).
- A strong majority of respondents (70%) agreed or strongly agreed that their organizations recognize and value information as a strategic asset and manage it accordingly.
- Nearly thirty-eight percent (38%) of survey participants said their organizations used both financial metrics and anecdotes/stories to assess the business impact of either poor or high data quality. The rest either used anecdotes/stories only (29%), financial terms only (14%) or other methods (10%).
- Organizations continue to face a long list of obstacles to their data quality efforts. The top four challenges cited by more than half of the respondents are: (1) lack of accountability and responsibility for data quality (55%), (2) too many information silos (52%), (3) lack of awareness or communication of the magnitude of data quality problems (51%), and (4) lack of common understanding of what data quality means (50%).
- Most respondents (65%) reported that their organizations did not have a documented information/data quality strategic or business plan that covers a time horizon of at least 2 years.

Organizational Structures for Information and Data Quality

Survey participants whose organizations have an information/data quality program currently in place reported the following:

- Information and data quality efforts were most often driven at the enterprise level (28%) or at a functional area level (28%).

- Information quality and data governance were most often led by the same person (31%). The individual leading the IDQ efforts often reported to an area other than IT/IS (68%)
- In the typical enterprise, Information/data quality efforts were often led by someone who sits three levels below the most senior leader in the organization (33%).
- Middle level non-IT managers (28%) were the typical members of the most senior information quality steering body. Twenty-six percent (26%) said the senior/executive management team (26%) was their top information quality steering body.
- Most senior information/data quality steering bodies met on either a monthly (20%) or quarterly (14%) basis.
- In general respondents said the most senior information/data quality steering body in their organizations had been in existence for 1 to 3 years (26%).
- While many survey participants (37%) reported their organizations did not have an established career path/staff ladder for their information/data quality staff, twenty percent (20%) reported their organization had plans to develop one, twenty-three percent (23%) had established an informal one, and nearly thirteen percent (13%) have a formal career path/staff ladder firmly in place.

Focus of Information and Data Quality Processes

- Survey participants named risk reduction and compliance as the primary direct or indirect business objective of organization's information and data quality efforts (39%), followed by cost reduction (19%) and increasing revenue (18%).
- The four most cited drivers, motivations or catalysts behind information and data quality efforts were (1) the general desire to improve the quality of data (68%), (2) data warehousing/business intelligence (47%), (3) compliance/risk/fraud/legal requirements (40%), and (4) master data management (MDM) projects (39%).
- While Customers (67%), Products and production (52%) and Financial (52%) led as the most common focus areas of information quality efforts, organizations are clearly seeking to improve the quality of a wide variety of data domains.
- To achieve their data quality aspirations, organizations are investing efforts in a broad range of information/data quality activities with the top five being: (1) data cleansing/remediation, (2) data quality monitoring , (3) propose, select or charter data quality improvement projects, (4) standardize data definitions across the organization, and (5) data quality assessment.

Information and Data Quality Tools

- Of the many categories of data tools, the following tools were the ones most often cited as being used in organizations and rated most important to information/data quality efforts: (1) data profiling and quality assessment (65%), (2) data quality monitoring (61%), (3) data

remediation/cleansing (57%), (4) data matching and reconciliation (data de-duplication) (53%), and (5) extract-transform-load (ETL) and other data integration tools (51%).

- For data profiling and/or data quality assessment, a large majority of respondents said their organizations used tools purchased from vendors (64%) in conjunction with ad-hoc queries (49%) and home grown tools (46%). Open source tools hold a lower but still significant share (20%).

Maturity of Information and Data Quality Efforts

- A comparison of three maturity attributes showed that Responsibility and Accountability was the most mature area for organizations (48% at a level of Defined or higher), followed by Goal-Setting and Measurement (42.5% at a level of Defined or higher). Policies, Plans and Procedures was the least mature area (37% at a level of Defined and higher).

The Characteristics of Organizations with Mature Information/Data Quality Processes

In addition to analyzing the survey population in general, we also focused on the 7% of respondents whose answers indicated that they came from organizations with very mature information and data quality practices in three areas: responsibility and accountability, policies, plans and procedures, and goal-setting and measurement. While survey participants from mature IDQ organizations and the general population shared similar views about IDQ challenges, key data domains, motivations for improving quality, and types of data quality activities/tools employed, we observed several differences.

Compared to the general population, organizations with mature IDQ processes are more likely to:

- View and value information as a strategic asset and to manage it accordingly.
- Have IDQ efforts that are enterprise-driven, meaning senior leadership is involved with managing the quality of key information and data assets across the organization with involvement by various functional areas and departments.
- Have built a close relationship between their information quality and data governance efforts.
- Have a documented IDQ strategic or business plan that covers a time horizon of at least 2 years.
- Have established some type of career path for their IDQ staff.
- Positioned their IDQ leaders at higher levels in the managerial hierarchy.
- Have greater representation of business (non-IT) managers and executives on their most senior IDQ steering body.
- Have a senior IDQ steering body that meets frequently (e.g. every month or every two weeks).
- Have a senior IDQ steering body that has been in existence for at least 5 years.
- Put greater efforts into their IDQ activities and placed greater importance on the use of IDQ tools.

Introduction

Overview and Objectives

This report presents the findings of a survey jointly conducted by the International Association for Information and Data Quality (IAIDQ) and the Information Quality Program at the University of Arkansas at Little Rock (UALR-IQ) between March 19 and April 20, 2012. The purpose of the survey was to better understand the current state of information and data quality programs and practices in organizations around the world. Our goal was to provide valuable insights for information/data quality practitioners, job seekers, employers, and the academic community in evaluating existing conditions, identifying best practices, and setting the agenda for future growth of the discipline.

Definitions

This report uses both "data quality" and "information quality" depending on the context most relevant to a particular situation. In addition, the report uses the label "business" as a synonym for "non-IT/IS".

Survey Methodology

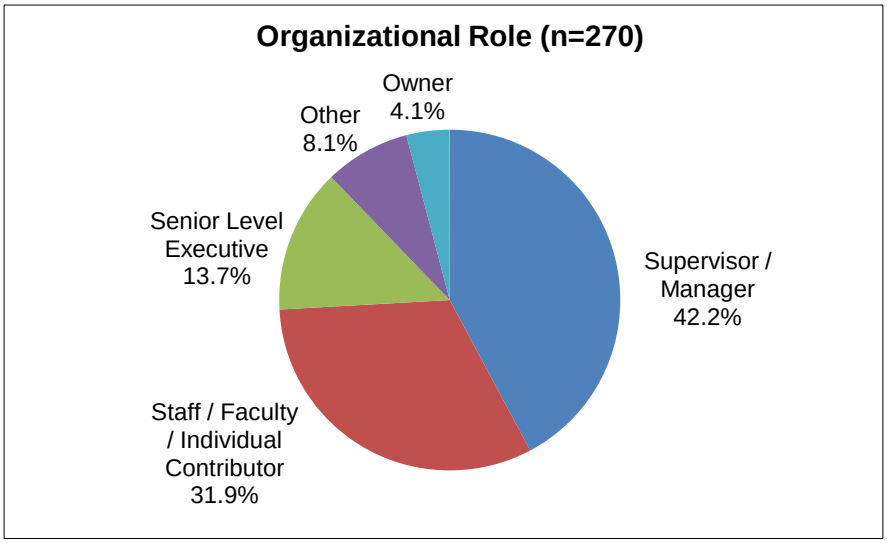
In early 2012, a team of UALR-IQ researchers and IAIDQ members developed a questionnaire to gather information regarding information and data quality (IDQ) programs and practices in today's organizations. The survey was officially launched via SurveyMonkey.com on March 19, 2012. IAIDQ sent several invitations via e-mail to individuals on its mailing list, asking them to complete the web-based survey. Invitations were also distributed via several data quality web sites and social networking groups. The survey closed on April 20, 2012.

Once the data collection period ended, the raw survey data were checked to eliminate any duplicates or abandoned survey responses (i.e., surveys where individuals exited the survey before completing any IDQ-related questions). A total of 296 participants started the survey. After duplicates and abandoned survey responses were eliminated, 270 participant responses remained. Because not every person answered every question, the number of respondents varies by question and is specified in the tables and charts featured in this report. In addition for some questions, multiple responses by individuals were allowed. This is designated where applicable in the tables and charts. Finally some survey questions included an "Other" option. The presence of an "Other" option allowed survey participants to write in their own responses in addition to selecting answers from a pre-determined set. Whenever possible, this report lists alternative answers provided by respondents when discussing the results for these open-ended questions.

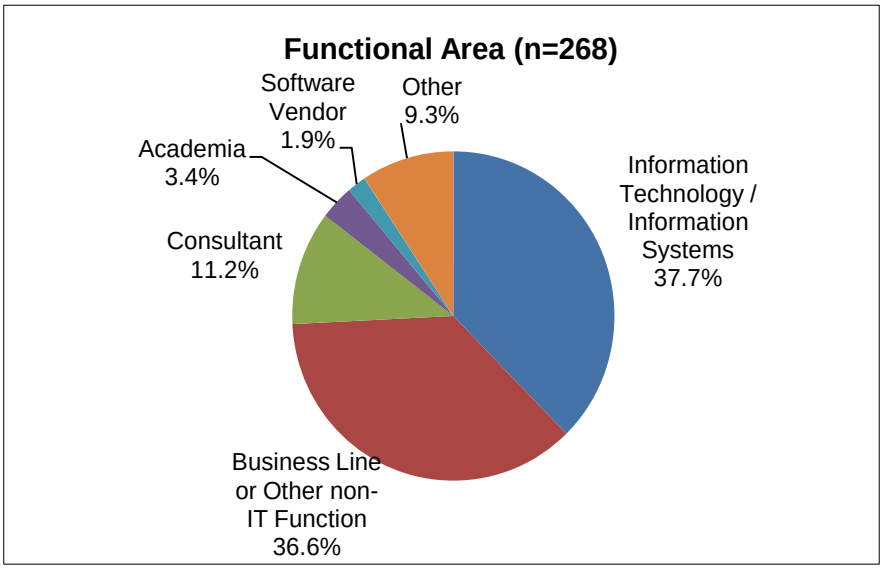
Survey Demographics

The 270 survey participants who completed our survey represent a diverse set of organizations from around the globe. A complete set of demographic charts are included in Appendix I of this report.

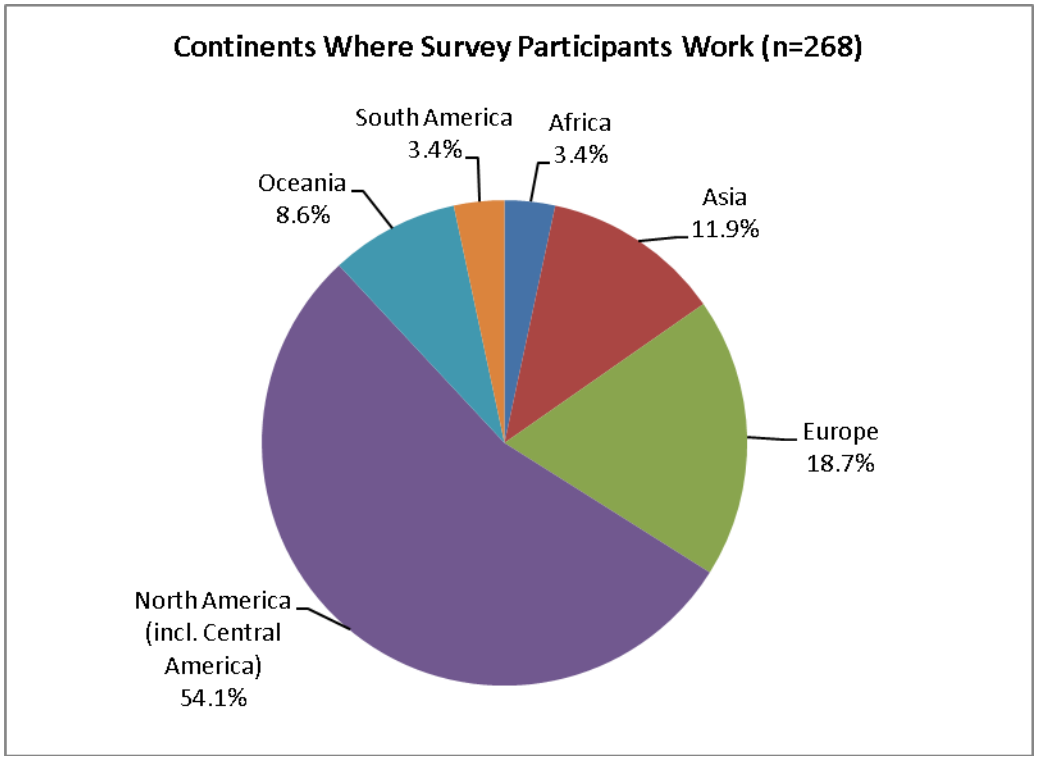
Roles that survey participants assume in their organizations: The majority of our survey participants identified themselves as either Supervisors/Managers (42.2%) or Staff members (31.9%).



Functional areas that best describe the work of survey participants: Most survey participants selected either IT (37.7%) or a Business line (36.6%) as the functional areas best describing the work that they do.



Where survey participants work: Survey participants from 33 countries volunteered their responses. Fifty-four percent (54.1%) of respondents work in North America (including Central America), followed by Europe (18.7%), Asia (11.9%), Oceania (8.6%), Africa and South America (3.4 % each). The countries with the largest survey participation were the United States (49.3%), Australia (7.5%), India (6.0%), United Kingdom (5.6%), and Canada (4.5%).



See Appendix I for an extended set of demographics charts. The essential statistics are summarized below.

Industries associated with survey participants' organizations: Survey participants represented over 25 different industries. The largest participation for our survey came from the Financial Services (13.3%), Energy/Oil & Gas (12.2%), Healthcare (8.5%), Consulting/Professional Services (7.8%), and Insurance (7.4%) industries.

Type of organizations for who survey participants work: Most survey participants work for either privately held (39.0%) or publicly traded companies (34.9%).

Organization size by employee count: Over half of the survey participants (54.8%) reported being employed by large organizations (more than 2,500 employees), roughly a quarter (29.3%) reported being employed by medium size organizations (100 to 2,500 employees) and the remainder (14.8%) reported being employed by smaller organizations.

Organization size by revenue or budget (for non-profits)s (USD): Over half of the survey participants (56.3%) stated they came from organizations with annual revenues in excess of 100 million USD. About a quarter (22.2%) stated their organizations as having smaller annual revenues, with the remainder (21.5%) unsure.

Market scope of customers that survey participants' organizations serve: The majority of survey participants described the customer base that their organization serves as either international (49.6%) or national (33.3%) in scope.

Part of organization that survey participants had in mind when answering questions: Nearly half of survey participants (49.3%) indicated that they had the entire organization in mind when they answered the questions. About a quarter (24.8%) focused on a functional area, and the rest chose either a department (15.2%) or a subsidiary of the organization (10.7%) as the scope they had in mind when answering our survey questions. Throughout this report, the term "organization" may refer to an entire enterprise and/or to a portion of the enterprise.

Information and Data Quality (IDQ) Maturity

One aim in conducting this industry study was to gauge the maturity of IDQ practices in organizations. There are a number of available maturity models that can be used as a benchmark. For our survey, we chose the COBIT 4.1 Maturity Model[1]. COBIT was originally developed as a guide for IT governance and it incorporates a maturity schema as part of its framework. We modified the wording of the COBIT 4.1 maturity model descriptions to change the focus to the management of information and data quality. Our revised maturity model concentrates on three areas: (1) Responsibility and Accountability, (2) Policies, Plans and Procedures, and (3) Goal-Setting and Measurement. Five maturity levels are defined for each.

Responsibility and Accountability

Over half of the respondents (52.3%) reported their organizations had not yet reached the Defined level. At the Defined level, organizations have defined IDQ responsibility and accountability roles with individuals assigned to carry out those duties; however, issues regarding authority still remain. Nearly 37% of respondents indicated their organizations had reached the Defined Level. Fewer than 11% of survey participants believed their organizations had reached the higher stages of maturity in the area of Responsibility and Accountability for IDQ.

Which of the following statements best describes the responsibility and accountability for information and data quality among employees in your organization? Select ONE only. (220 respondents)			
Maturity Level	**Description**	**Response Percent**	**Cumulative Percent**
5 – Optimized	Information / data stewards and others with information/data quality roles are empowered to make information / data quality decisions and to take action. The acceptance of responsibility has been cascaded down throughout the organization in a consistent fashion. An effective governance structure has been established.	3.6%	3.6%
4 – Managed	Information and data quality responsibility and accountability are accepted and working in a way that enables information/data stewards and others with information/data quality roles to fully discharge their responsibilities. An appropriate reward structure is in place.	7.3%	10.9%
3 – Defined	Information and data quality responsibility and accountability are defined and information/data stewards have been identified. Occasionally, the information/data stewards and others with information and data quality roles may lack the full authority to exercise their responsibilities.	36.8%	47.7%
2 – Repeatable	One or more individuals have assumed responsibility for information quality and are usually held accountable, even if this is not formally agreed. There is often confusion and blame about responsibility when information and data quality problems occur.	25.5%	73.2%

[1] Questions based on content from COBIT 4.1, which is used by permission of the IT Governance Institute (ITGI). Copyright IT Governance Institute. All rights reserved.

1 - Ad-hoc	There is no clear definition of accountability or responsibility for information and data quality issues. People take ownership of information/data quality issues based on their own initiative as problems arise.	26.8%	100%

Policies, Plans, and Procedures

The majority of the respondents (63%) reported their organizations had not yet reached the Defined level in the area of Policies, Plans and Procedures. At the Defined level, organizations have defined and documented IDQ processes and policies along with more formal and structured practices for communicating these plans. About 29% of respondents indicated their organizations had reached the Defined Level. Fewer than 8% of survey participants thought their organizations had reached the higher stages of maturity in establishing Policies, Plans and Procedures.

Which of the following statements best describes the status of information and data quality <u>policies, plans, and procedures</u> in your organization? Select ONE only. (222 respondents)

Maturity Level	Description	Response Percent	Cumulative Percent
5 – Optimized	Benchmarking against external best practices and standards for information/data quality is applied. The effectiveness of information and data quality processes and policies is continually being improved. Management is engaged in proactive and ongoing communication of these practices.	1.8%	1.8%
4 – Managed	All aspects of information and data quality processes and policies are documented and repeatable. Policies have been approved and signed off by management. Standards for managing and improving the quality of information and data quality processes and policies are adopted and followed. Management is communicating these practices on a frequent and widespread basis.	5.9%	7.7%
3 – Defined	Information and data quality processes and policies are defined and documented for all the subject areas the organization is focusing on. Management is becoming more formal and structured in its communication of these practices.	29.3%	36.9%
2 – Repeatable	Some documentation and/or understanding of common information and data quality processes and policies are emerging, but are largely intuitive because of individual expertise. Management is communicating some of these practices.	35.1%	72.1%
1 - Ad-hoc	Information and data quality processes and policies are largely undefined. Several ad-hoc processes and policies exist, but management communication about these practices is sporadic.	27.9%	100%

Goal Setting and Measurement

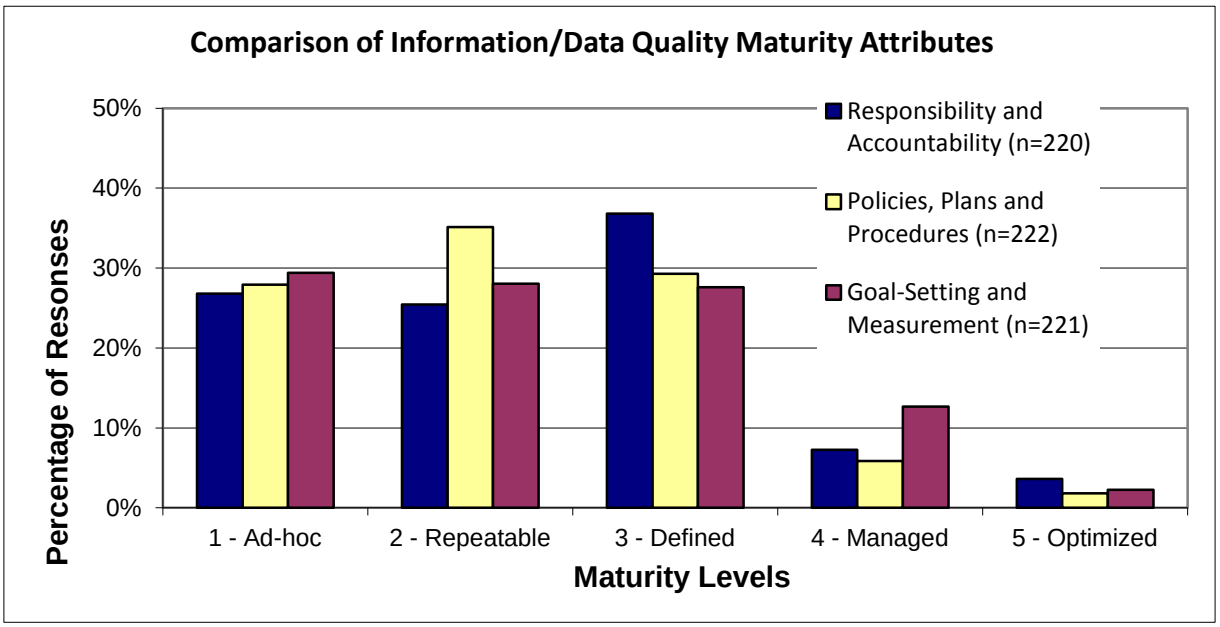
As for Goal Setting and Measurement of IDQ results, over half of the respondents (57.5%) reported their organizations had not yet reached the Defined level. At the Defined level, organizations have set some IDQ goals and metrics, but they do not clearly link to the strategic goals of the organization.

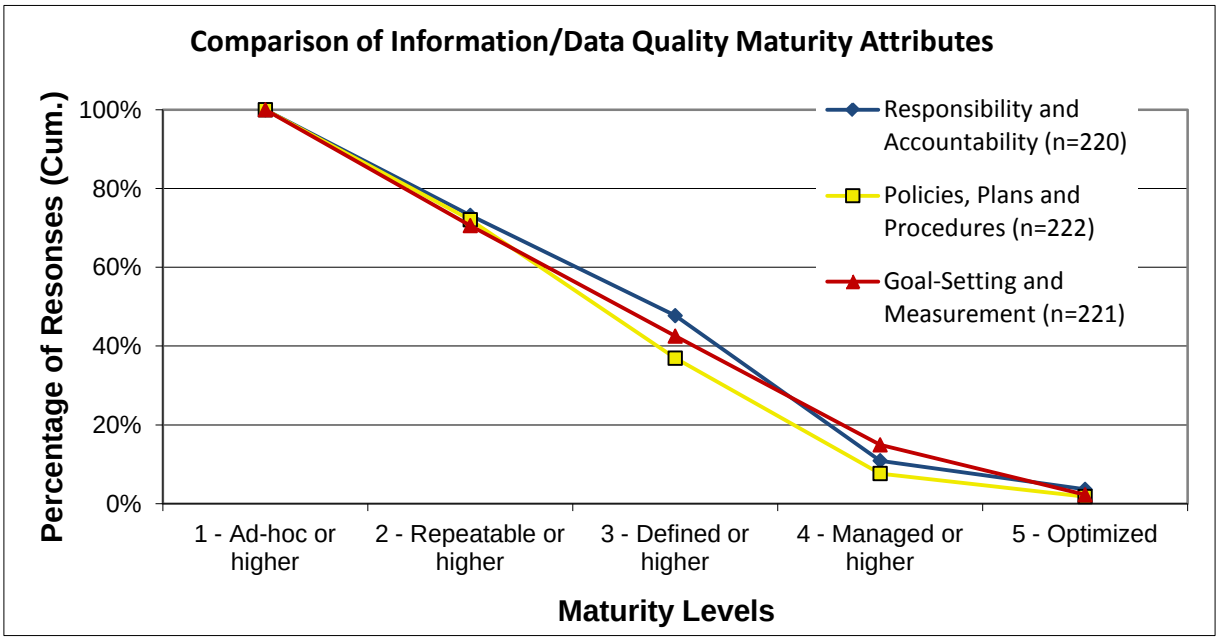
Measurement processes are also not consistently applied. In addition, communication about these IDQ goals and metrics is not widespread. Nearly 28% of respondents indicated their organizations had reached the Defined Level. About 15% of survey participants judged their organizations as being at the higher stages of maturity when it comes to ensuring that an effective system is in place for defining, measuring, monitoring and managing IDQ performance.

Which of the following statements best describes the status of information and data quality goal setting and measurement in your organization? Select ONE only. (221 respondents)			
Maturity Level	**Description**	**Response Percent**	**Cumulative Percent**
5 – Optimized	An organization-wide integrated information and data quality performance measurement system is in place. It links information/data goals to organizational strategic goals. Goals are routinely met. Deviations are consistently noted by management and root-cause analysis is applied. Continuous improvement of information and data quality processes is ongoing.	2.3%	2.3%
4 – Managed	Efficiency and effectiveness goals are set, communicated, measured, and linked to organization's strategic goals. Continuous improvement of information and data quality processes is emerging.	12.7%	14.9%
3 – Defined	Some information and data quality effectiveness goals and measures are set, but may not be widely communicated. There is no clear link to strategic organizational goals. Measurement processes for these goals are emerging but are not consistently applied.	27.6%	42.5%
2 – Repeatable	Some information and data quality goal setting occurs. Measurement of success against these goals is inconsistent and typically limited to a few areas.	28.1%	70.6%
1 - Ad-hoc	Information and data quality goals are not clear and no measurement exists.	29.4%	100%

Comparison of Results

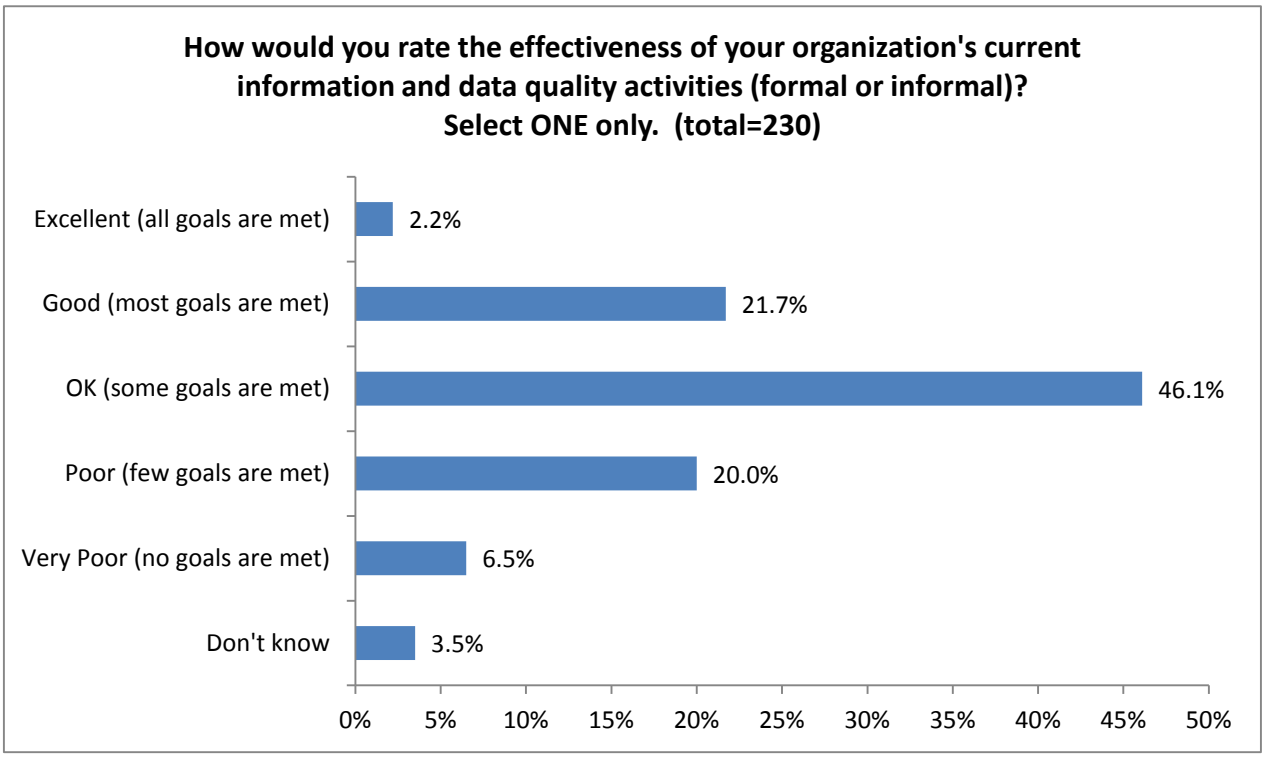
A side by side comparison of the results suggests that organizations are performing better when it comes to the maturity of their Responsibility and Accountability (47.7% Defined or higher), followed by Goal-Setting and Measurement (42.5% Defined or higher) efforts. Policies, Plans and Procedures appears to be the least mature area (36.9% Defined or higher). This may suggest a weakness in organizations' ability to sustain their IDQ efforts. By not embedding desired behaviors in policies, plans and procedures, organizations leave conformance up to individual preferences and management whim, and vulnerable to employee turnover. This may also explain the familiar stop-and-go nature of many IDQ programs.





How Effective are IDQ Activities?

Another useful indicator of an organization's maturity is the effectiveness of its IDQ activities. Forty-six percent (46.1%) of respondents reported that the effectiveness of their organization's current information and data quality activities (formal or informal) was Okay (some goals are met). On the positive side, 21.7% of respondents selected Good (most goals are met) and 2.2% selected Excellent (all goals are met). On the negative side, 20.0% of respondents chose Poor (few goals are met) and 6.5% chose Very Poor (no goals are met). The balance (3.5%) indicated that they did not know.



Do Mature Organizations Share Common and Distinctive Characteristics?

To identify the shared and distinctive characteristics of organizations with mature IDQ programs, we wanted to compare the responses of survey participants from mature IDQ organizations with the rest of the responses to see what lessons could be learned. To accomplish that task we first had to define "IDQ maturity".

Defining "IDQ Maturity": We sought to define a criterion that would be based on a rigorous and sound definition of maturity while including enough responses to be representative of mature IDQ organizations. We selected the following criteria:

IDQ mature organizations are those who are:

- At a maturity level of 3-Defined, or 4-Managed, or 5-Optimized, in the Responsibility and Accountability area
- AND at a maturity level of 3-Defined, or 4-Managed, or 5-Optimized in the Policies, Plans and Procedures area
- AND at a maturity level of 4-Managed or 5-Optimized for in the Goal-setting and Measurement area
- AND with a rating of Good or Excellent for the Effectiveness of their IDQ activities.

The rationale for these criteria was based on our assessment that more organizations had reached the higher levels of maturity when it came to Goal Setting and Measurement (~15%) compared to

Responsibility and Accountability (~11%) and Policies, Plans and Procedures (~8%). This may be due in part to the stronger wording of the requirements for the Defined levels for Responsibility and Accountability and Policies, Plans and Procedures. In contrast, we felt the Defined level for Goal-Setting and Measurement was not as demanding. Furthermore by adding the effectiveness condition, we sought to achieve some balance between the input/leading indicators of the Maturity Model, with the outcome/lagging indicator embodied in the Effectiveness rating.

These criteria yielded 19 responses or about 7% of the total survey population, a number sufficiently selective to separate these organizations from the rest. Six respondents from this group did not disclose their organizations. Of the 13 who did disclose their organizations, only 2 were overlapping (i.e., 2 individuals from the same company). The rest were either from different companies or individuals reporting on different parts of the same large multinational company. Because we potentially had overlapping results in the general population as well, we decided to compare the results as is. Despite these limitations, we believe this group is a worthy benchmark considering that among those with disclosed organizations, we noted that a number of them had strong reputations among their IDQ peers for having excellent data quality programs.

Identifying the differences between IDQ-mature and less mature organizations: In the subsequent findings in this report, we have distinguished the responses of the more mature organizations from the rest, to highlight the extent to which maturity corresponds to unique characteristics in information and data quality efforts trends, structures, processes, and tools. We have also included remarks where appropriate to describe these differences. These comments are highlighted in yellow.

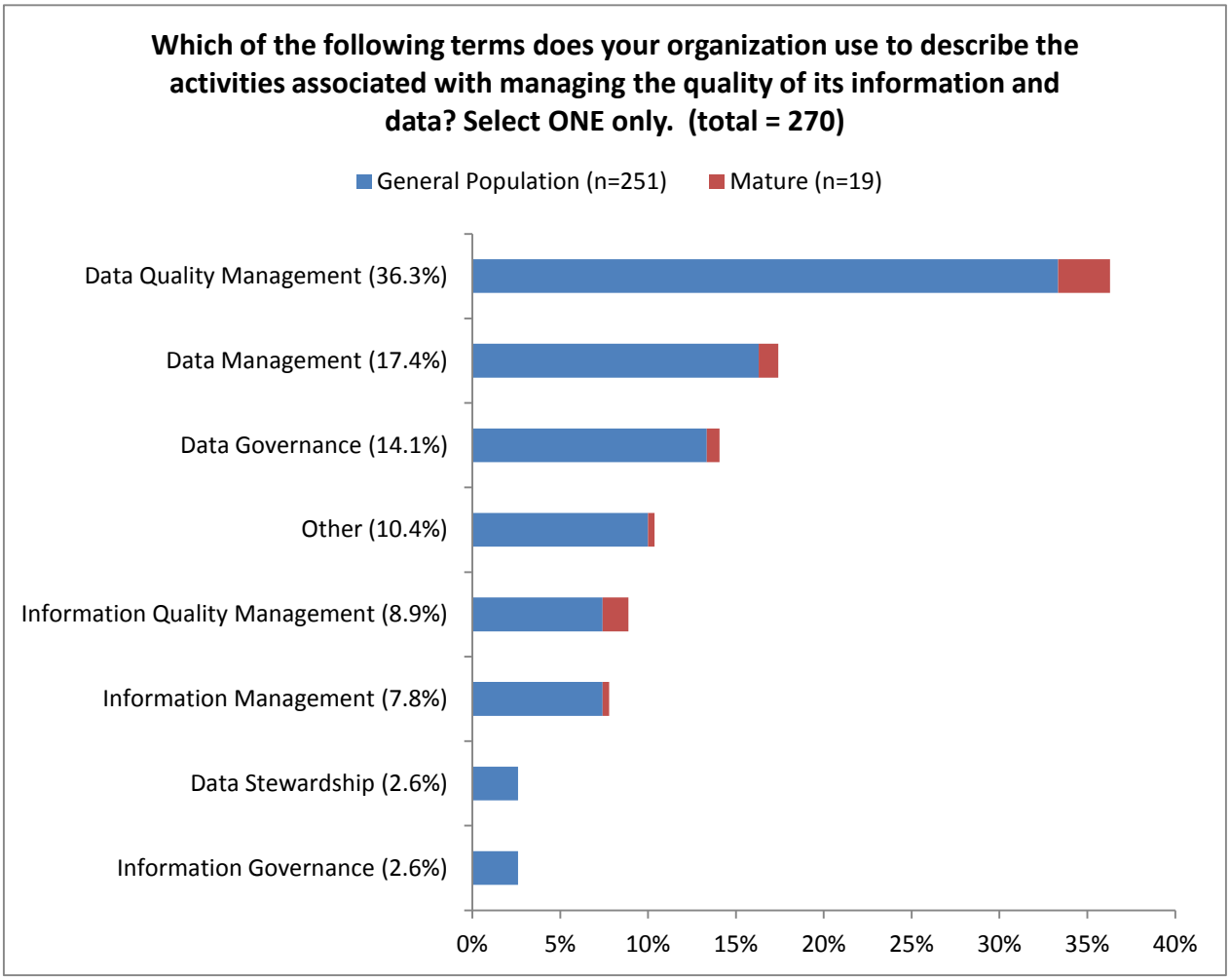
To further support the reasoning behind our comments, we have included some additional maturity comparison charts in Appendix II of this report.

Trends in Information and Data Quality (IDQ)

This section of the report explores some current trends regarding Information and Data Quality in organizations.

What Terms are Used to Refer to IDQ Activities?

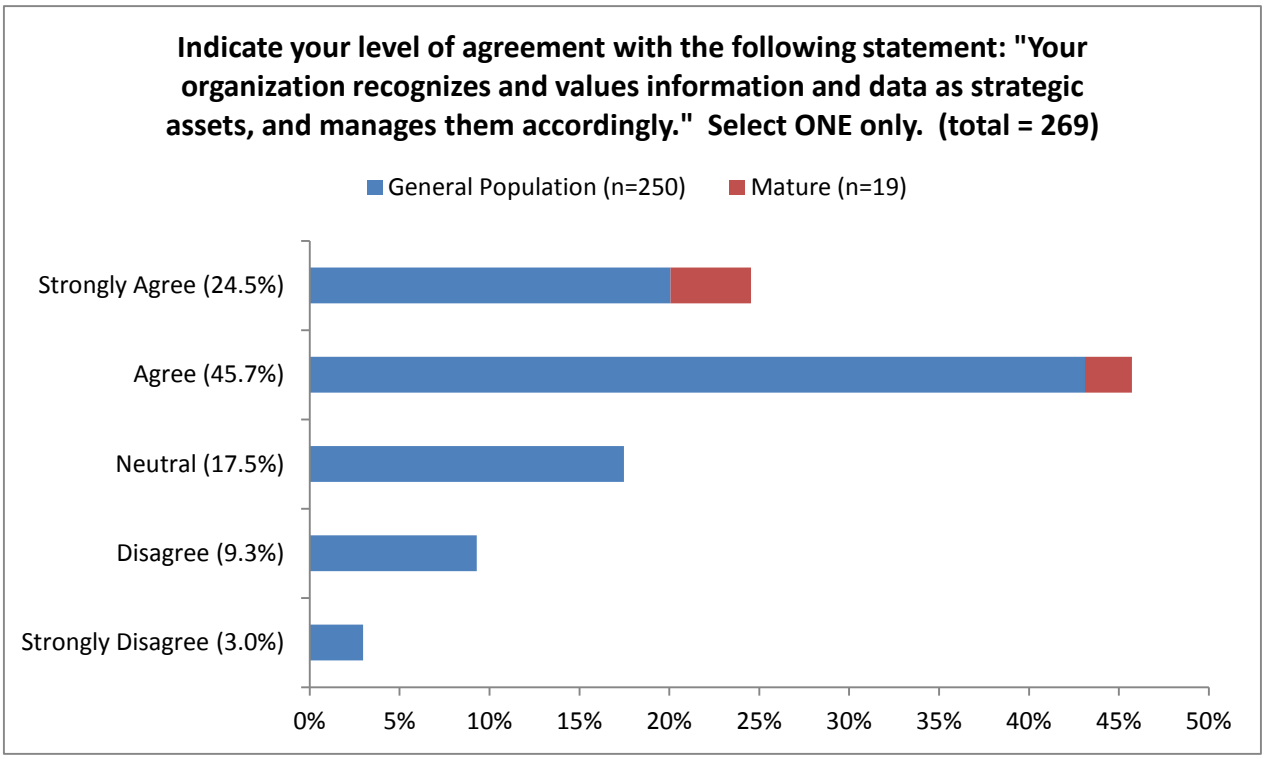
Survey participants chose Data Quality Management as the term most frequently used in their organizations to describe IDQ activities. As for "Data" vs."Information", a label containing the word "Data" was selected by nearly 70% of survey participants. In addition to the options presented, respondents contributed other terms such as "Enterprise Data Management", "Master Data Management", "Information Architecture", "Data Quality", "Data Integrity", and "Quality Management". Several people observed that their organization did not yet have an expression because no formal data quality efforts existed. Others remarked that their organization used multiple labels because of a lack of a consistent system for data and information management.



Do Organizations Recognize and Value Information and Data as Strategic Assets?

Recognition of information and data as strategic assets has improved since 2008 when we asked a similar question in our 2008 State of Information and Data Governance Industry Report. Back then, 58% of survey participants either strongly agreed or agreed with this statement with roughly 17% neutral and the rest negative. As of 2012, 70% of respondents either strongly agreed or agreed that their organization values information as a strategic asset and manages it accordingly, with 17.5% remaining neutral and approximately 12% disagreeing or strongly disagreeing.

Note: 100% of respondents from mature IDQ organizations either agreed or strongly agreed with this statement.

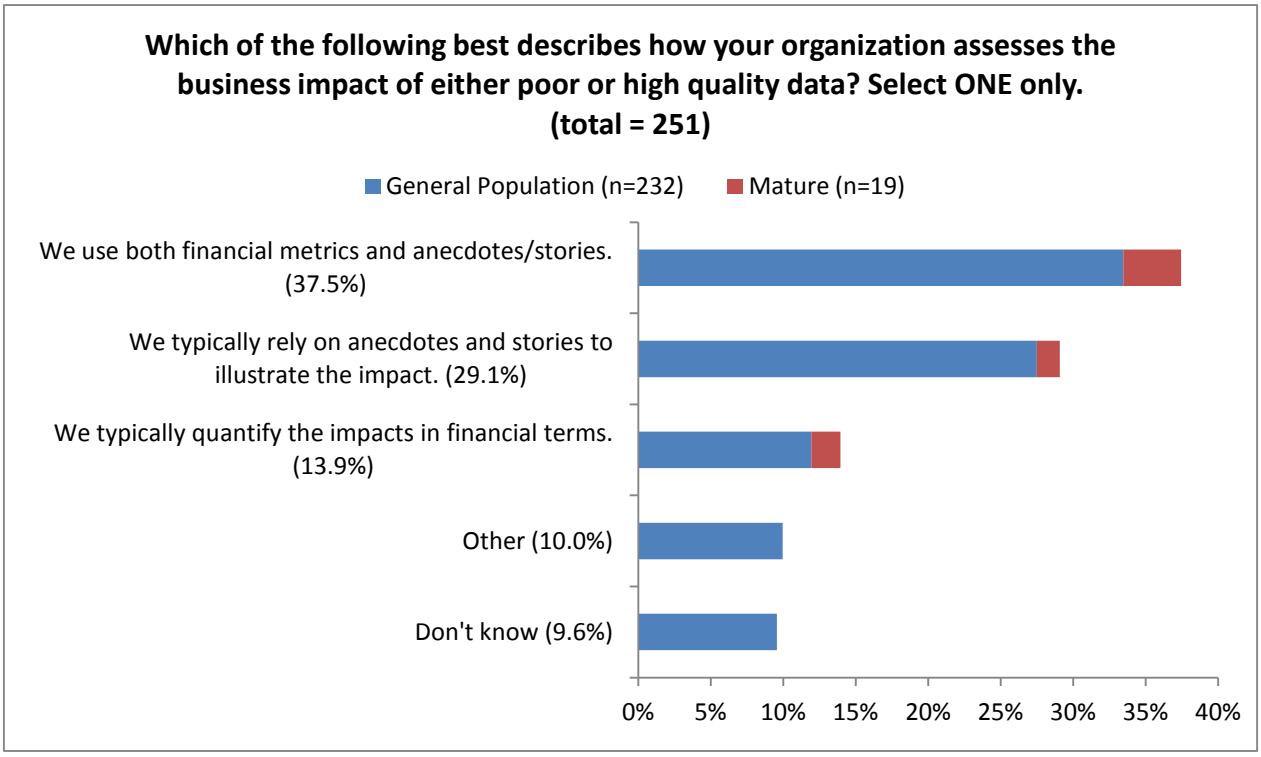


How do Organizations Assess the Business Impact of Either Poor or High Quality Data?

A combination of financial metrics and anecdotes to illustrate the business impact of either poor or high quality data was the most frequent choice (37.5%). Only 14% of respondents exclusively used financial metrics. Several people commented that their organizations did not yet have any procedures in place for assessing how data quality impacts the business. In addition, respondents contributed a variety of other methods they are employing, including:

- Full-time equivalent (manpower or man hours) needed to correct the problem.
- Turnaround time to resolve customer queries
- Data Audits
- Data Quality Thresholds

- Quantified Data Quality Scores such as Accuracy %
- Customer Satisfaction Ratings
- Customer Complaints
- Reputational Risk
- Clinical Safety
- KPIs used to assess customer targets in SLAs
- Defects



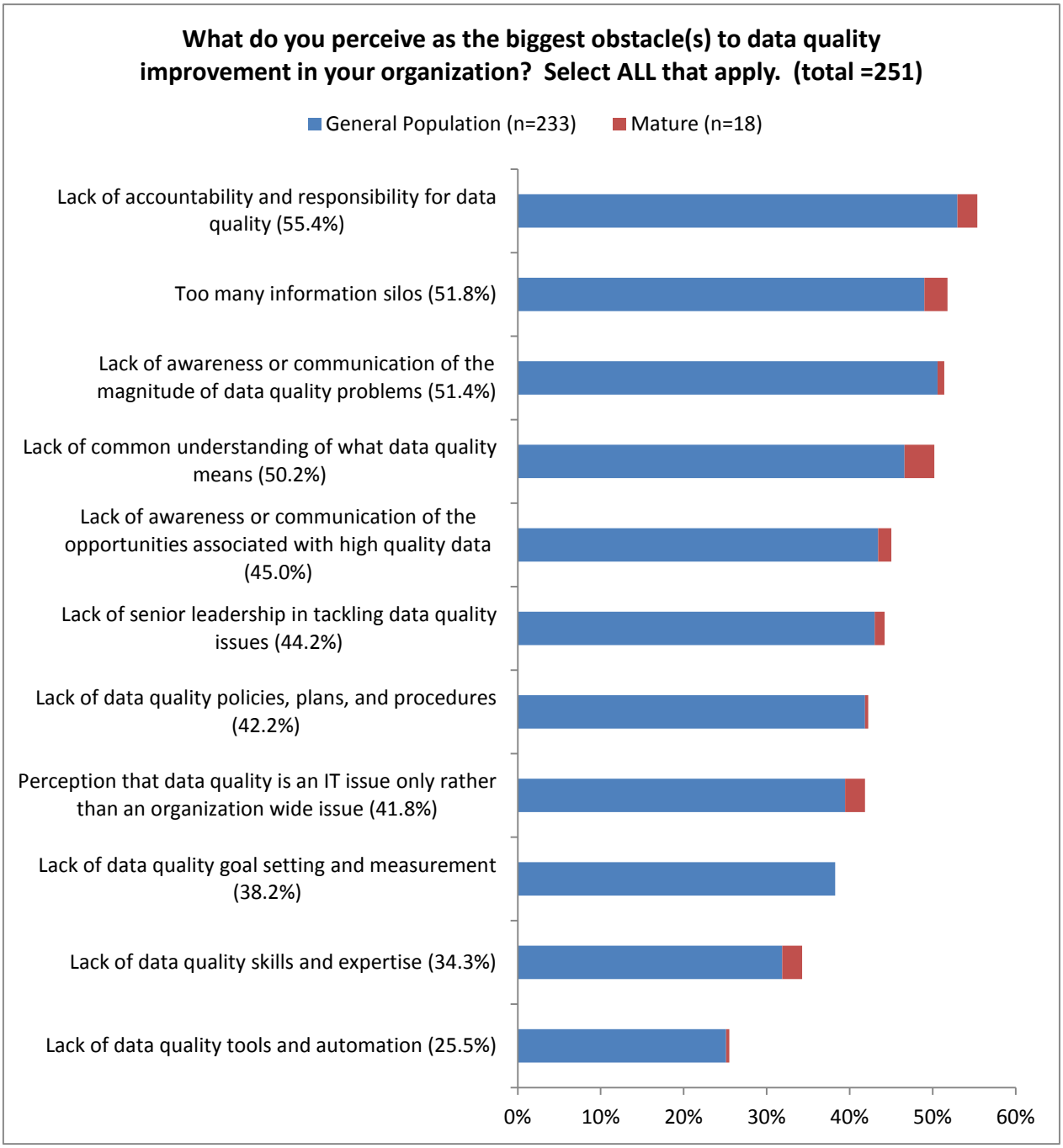
What are the Biggest Obstacle(s) to Data Quality Improvement in Organizations?

Survey participants reported that they are facing multiple obstacles to improving data quality in their organizations. While people most often selected "Lack of accountability and responsibility for data quality", "Too many information silos", "Lack of awareness of the magnitude of data quality problems", and "Lack of common understanding of what data quality means" as the biggest problems, all of our predefined choices received numerous votes. Furthermore, in addition to the challenges that we listed, people provided several others:

- Lack of resources including limited staff to manage data issues and promote data quality, cost to build a good data quality program, time to get proper tools and automation in place.
- Out of date policies, plans, and procedures.
- Lack of grass roots development of data quality as a strategic vision
- Lack of data quality rules that are customer focused

- Lack of understanding by data collectors of their impact on quality
- Lack of awareness of impact of frequent organizational changes on contextual meaning and usability of data assets
- Perception that Data Quality is a business issue only and cannot be helped with IT support

Note: While respondents from mature IDQ organization reported fewer obstacles related to governance issues, they also faced multiple issues inhibiting their IDQ efforts.

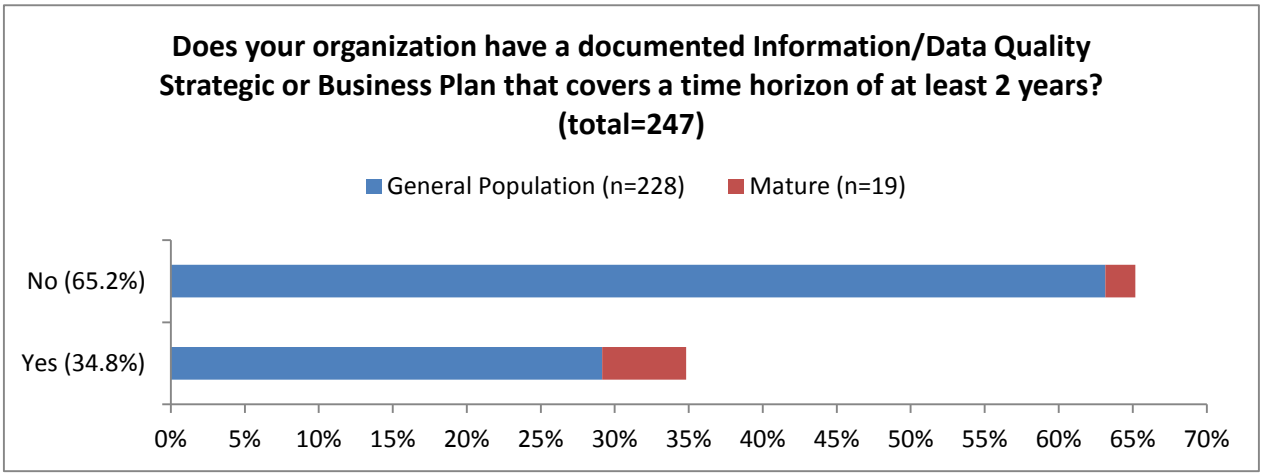


Note: For this question the terms information and data quality can be used interchangeably.

Do Organizations Have a Documented IDQ Strategic or Business Plan?

About a third of survey participants (34.8%) responded "Yes"; their organization does have a documented IDQ strategic or business plan that covers a time horizon of at least 2 years. The rest said "No".

Note: nearly 74% of respondents from mature IDQ organizations replied "Yes" – their organization did have a documented IDQ strategic or business plan.



Note: Only one selection was allowed for this question.

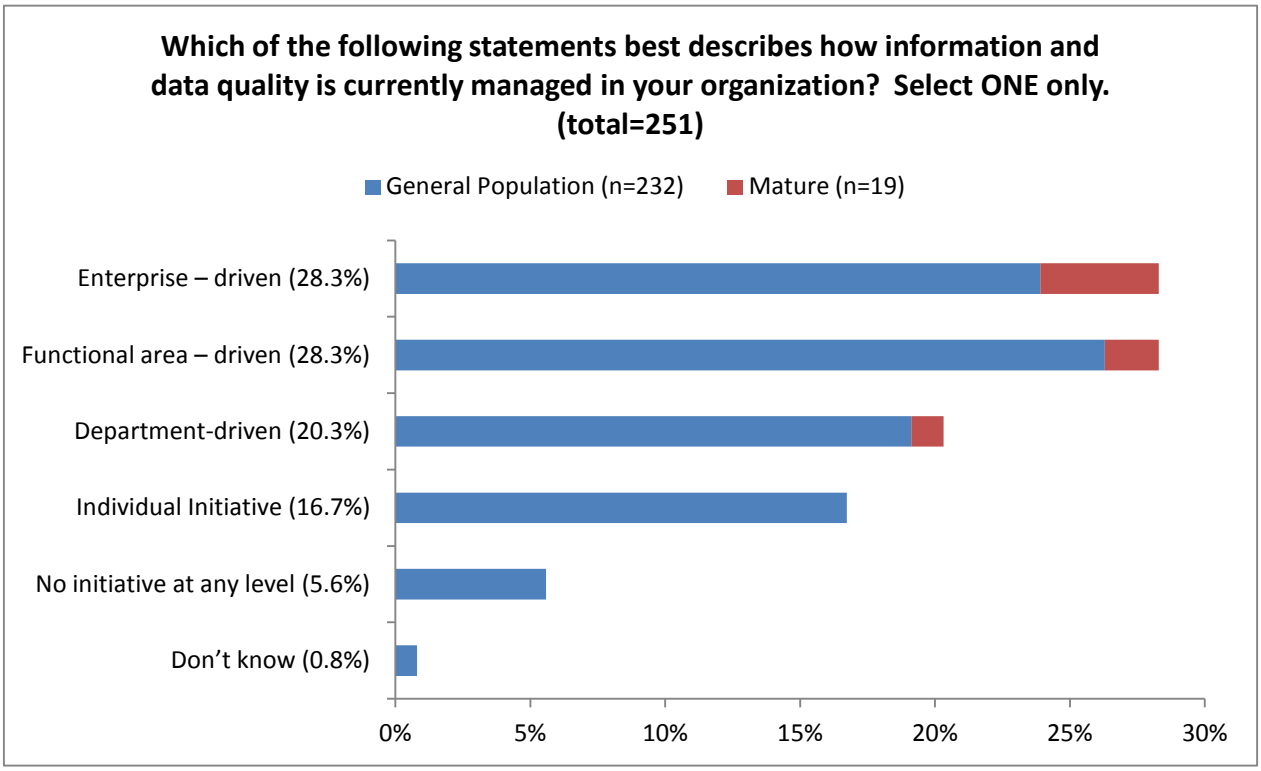
Organizational Structures for Information and Data Quality

This section of the report describes the organizational design and structural aspects of IDQ efforts.

How is IDQ Managed in Organizations?

According to our survey results, most IDQ management efforts are driven at either the enterprise or functional areas. 28.3% of respondents said that IDQ efforts are enterprise-driven, meaning senior leadership is leading the effort across the organization with involvement from functional areas and departments. Another 28.3% of respondents reported that IDQ efforts are driven by functional areas with participation from the departments that report to those areas. About 20% of survey participants responded that departments are driving the management of the quality of their information and data assets. 16.7% of respondents indicated that in their organizations, information and data quality management is left to individuals to pursue on their own initiative. Finally 5.6% of survey participants reported no information and data quality management at any level in their organizations.

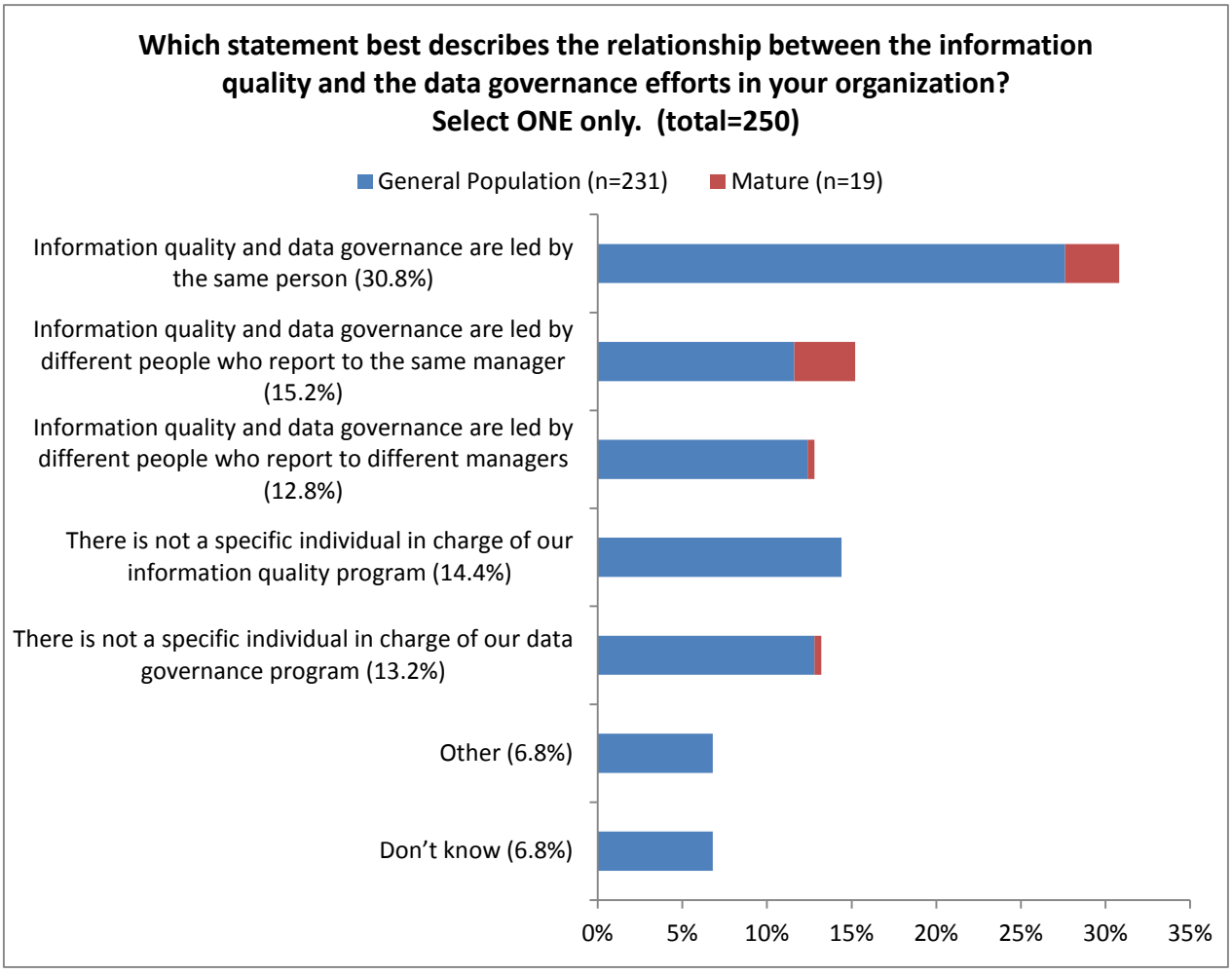
Note: Nearly 58% of respondents from mature IDQ organizations said that IDQ efforts at their organizations are enterprise-driven.



What is the Organizational Relationship between IDQ and Data Governance?

Because issues surrounding the quality, integrity, or usability of information sometimes fall under the scope of an organization's data governance initiatives, we asked survey participants to share with us the relationship between IDQ efforts and data governance efforts in their organizations. While nearly a third (30.8%) of respondents said that in their organizations information quality and data governance are led by the same person, the rest pointed out that a wide range of relationships exist. In addition to the options provided, a few individuals noted that in their organizations IDQ initiatives report directly into their Data Governance Group. Furthermore about 4% of respondents wrote that no relationship exists because their organizations either did not have a Data Governance program or were still in the very early stages of developing a Data Governance program.

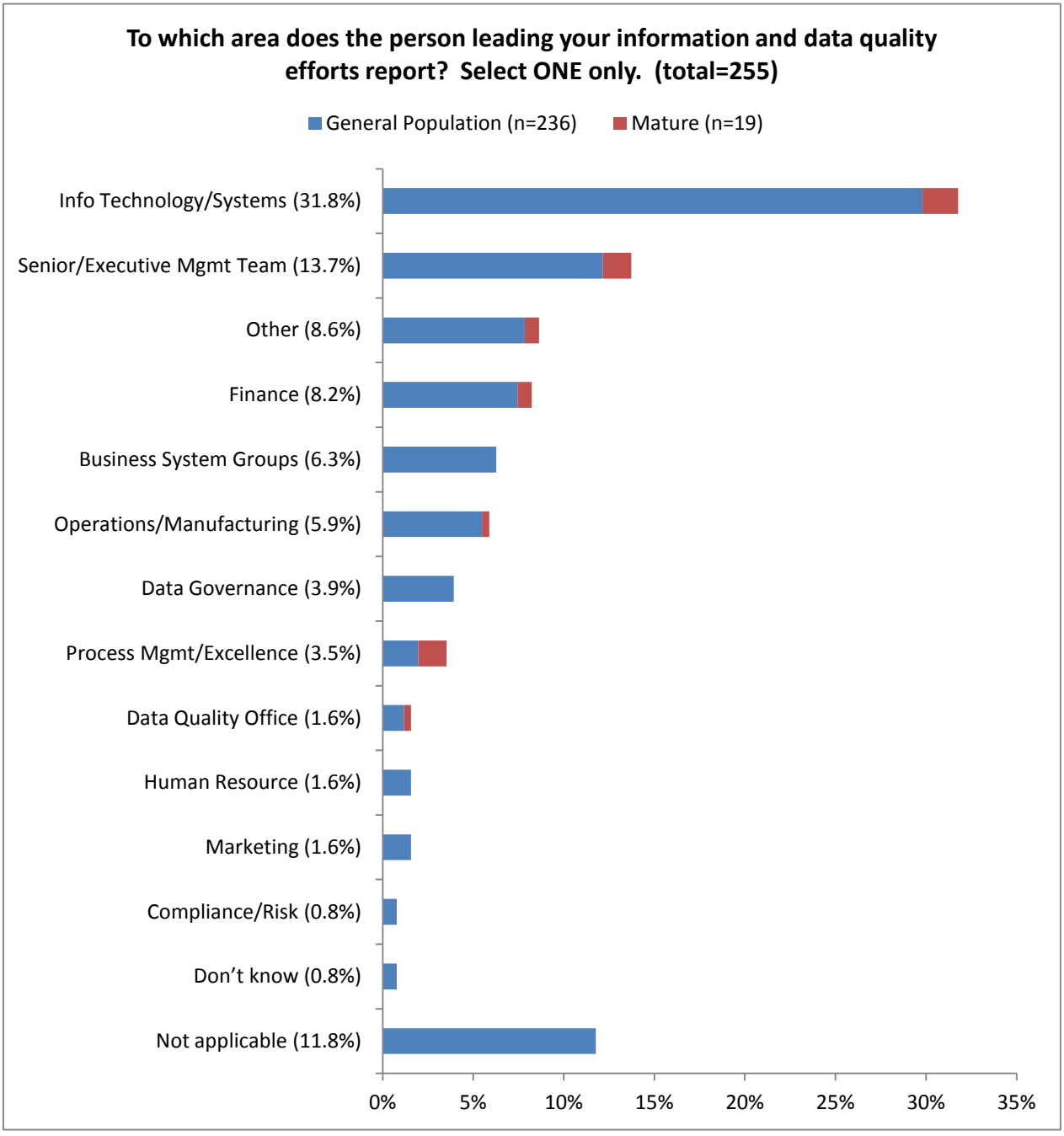
Note: Over 80% of respondents from mature IDQ organizations indicated a close relationship between their information quality and data governance efforts: 42% reported that these two efforts are led by the same person, 47% said they are led by different people who report to the same manager.



Note: The terms "information" and "data" can be used interchangeably for this question

Where Does the Person Leading the IDQ Efforts Report?

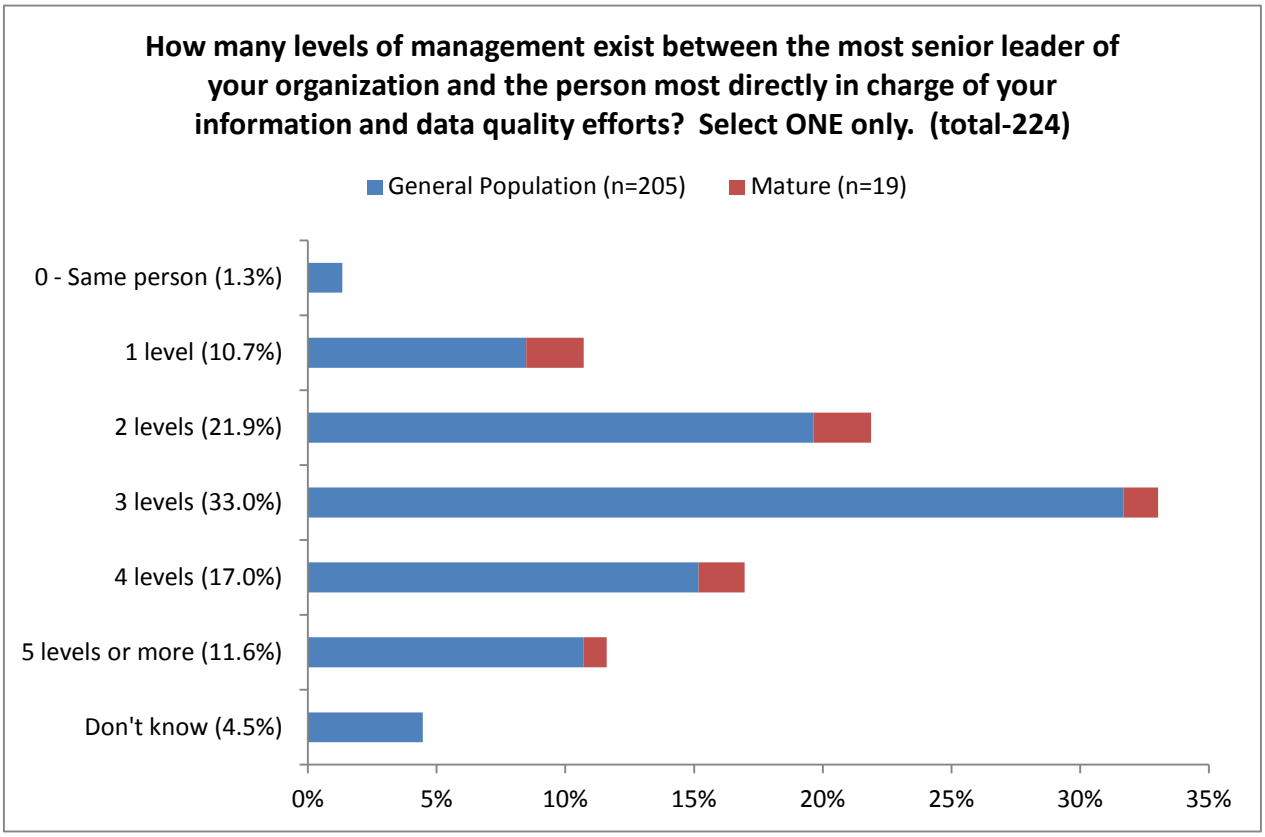
Survey participants chose Information Technology/Information Systems as the most common reporting area (31.8%) for individuals leading an organization's IDQ efforts. Conversely this means that most IDQ efforts are led by people who report outside of IT/IS. We feel this is very encouraging given the need to establish business ownership of IDQ efforts. Among the non-IT/IS areas, respondents selected a variety of areas with Senior/Executive Management Team (13.7%) leading the pack. In addition to the choices listed, respondents provided other reporting areas such as Business Intelligence, Supply Chain Management, Internal Audit, Research, Medical Affairs, Asset Management, and Data/Information Management Groups separate from IT. About 12% of respondents indicated "Not applicable", most probably because their organization does not have a specific individual leading their organization's IDQ efforts.



What is the Seniority of IDQ Leaders?

As a follow up to the previous question, survey participants who indicated that their organizations had an IDQ leader were asked how high in the organization that leader reports. Roughly a third (33.9%) said that fewer than three levels separate the most senior leader of their organization and the person most directly in charge of their IDQ effort. One-third (33.0%) reported exactly three levels. The remainder indicated more than three levels or said they were unsure.

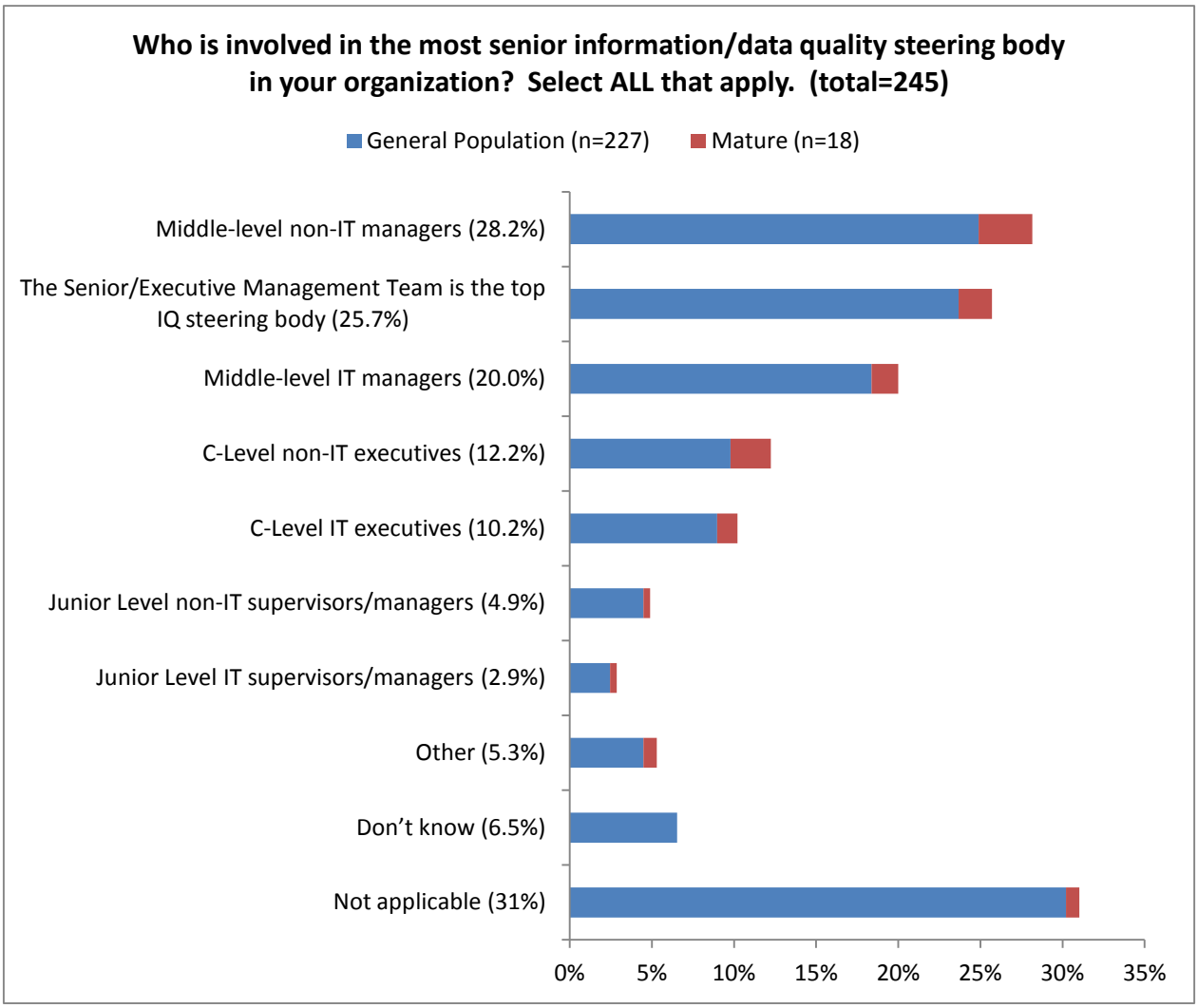
Note: Mature IDQ organizations seem to place their IDQ leaders at higher reporting levels. Nearly 53% of respondents from mature IDQ organizations indicated that fewer than three levels separate the most senior leader of their organization and the person most directly in charge of their IDQ efforts. Of the remainder, slightly fewer than 16% reported 3 levels of separation and about a third (~32%) said that there were more than 3 levels of separation.



Who in the Organization is Involved in the Most Senior IDQ Steering Body?

In many organizations information and data quality initiatives and processes are guided by one or more bodies such as a Data Council, Steering Committee or the equivalent. Our survey participants most frequently cited middle-level business managers (i.e., non-IT) (28.2%) and middle-level IT managers (20.0%) as the ones most involved in their organizations' IDQ steering body. About a quarter of respondents (25.7%) chose the Senior/Executive Management Team as the top IDQ steering body in their organizations. Other positions suggested by respondents include Data Owners, Architects, and Directors. Nearly a third (31.0%) selected "Not applicable" meaning those survey participants came from organizations that do not have a senior IDQ steering body.

Note: compared to the general population, mature IDQ organizations are more likely to include middle-level non-IT managers (~44%) and C-level non-IT executives (~33%) on their most senior IDQ steering body.

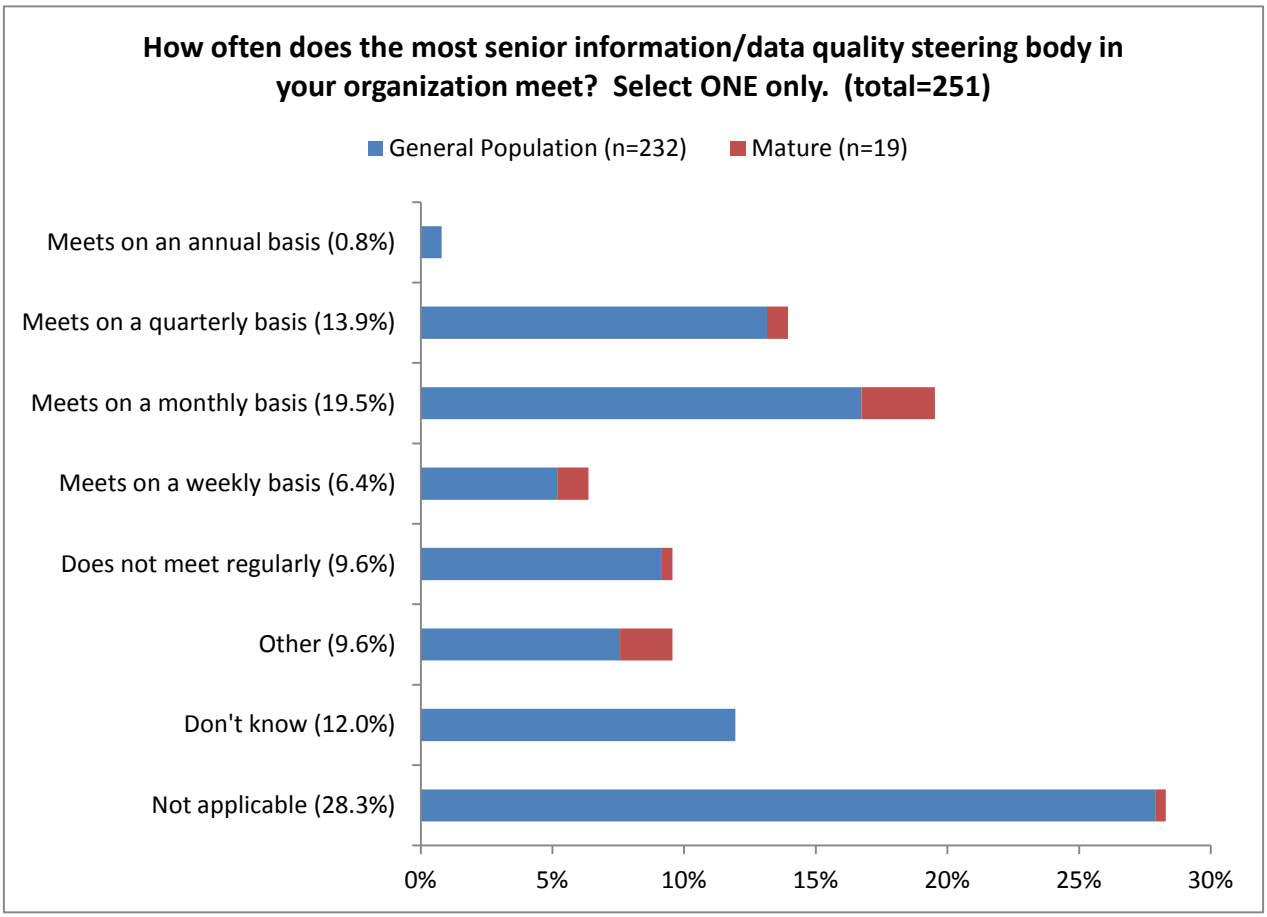


Note: By "non-IT" we mean lines of business or departments other than Information Technology / Information Systems.

How Often Does the Most Senior IDQ Steering Body in Organizations Meet?

For those survey participants whose organizations have a senior IDQ steering body, meeting monthly was chosen most frequently (19.5%) followed by meeting quarterly (13.9%). In addition to the selections provided, a number of respondents wrote-in alternative meeting schedules for their most senior IDQ steering body, such as every two months, every two weeks, and twice a year.

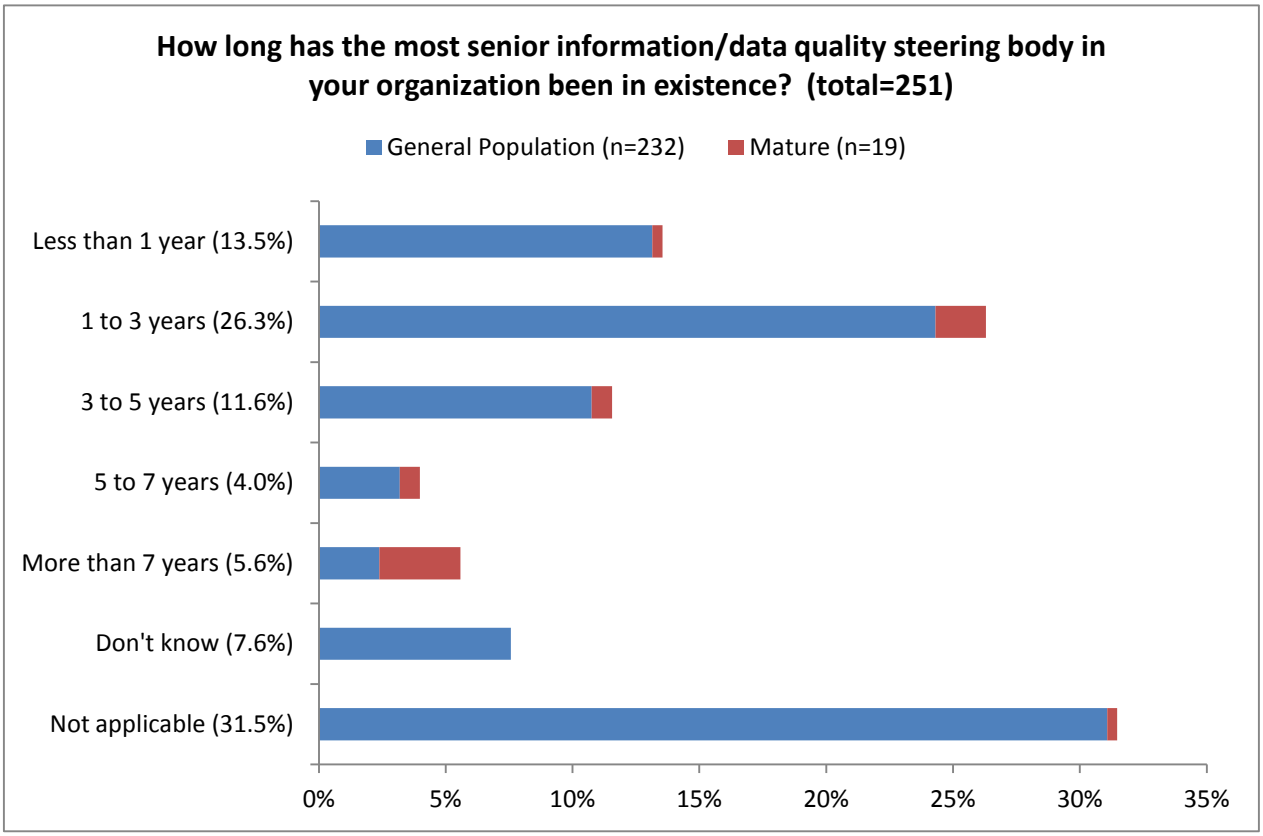
Note: Senior IDQ steering bodies in more mature organizations appear to meet more frequently. Nearly 37% of respondents from these organizations said their senior IDQ steering bodies met on a monthly basis and nearly 16% met on a weekly basis.



How Long has the Most Senior IDQ Steering Body Existed in the Organization?

Survey participants whose organizations have a senior IDQ steering body indicated that this body is relatively new. The most common responses were 1 to 3 years (26.3%), less than 1 year (13.5%) and 3 to 5 years (11.6%).

Note: Perhaps not too surprisingly, about 42% of respondents from mature IDQ organizations said their most senior IDQ steering body had been in existence for at least 7 years with another 10.5% reporting their most senior IDQ steering body had been in existence for 5 to 7 years.

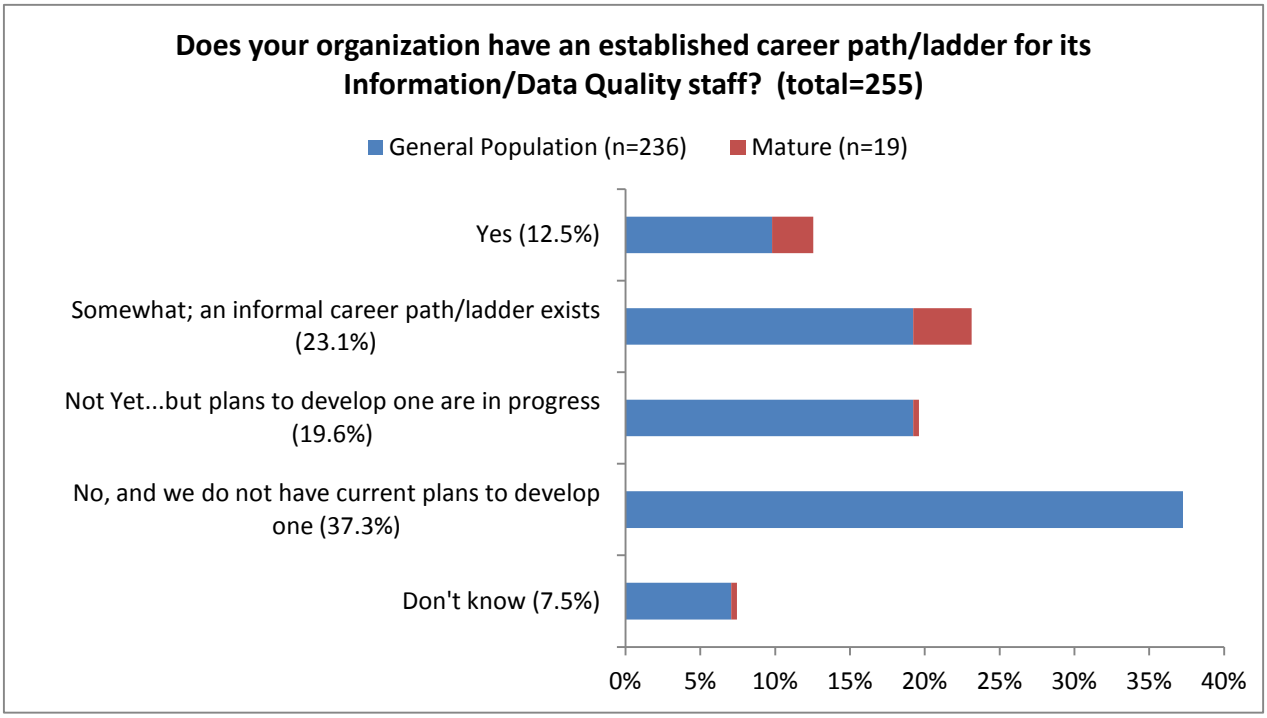


Note: Only one selection was allowed for this question.

Have Organizations Established Career Paths for their IDQ Staff?

Though the most frequent response (37.3%) was that organizations neither have a career path for their IDQ staff nor plans to develop one, there are some promising signs of progress in this area. Over half of the respondents indicated that their organizations are in some phase of establishing a career path for IDQ professionals.

Note: Mature IDQ organizations tend to have some type of career path for their IDQ staff. Nearly 37% of respondents from mature IDQ organizations said Yes - their organization does have an established career path while close to 53% replied Somewhat - an informal career path/ladder exists for their IDQ staff.



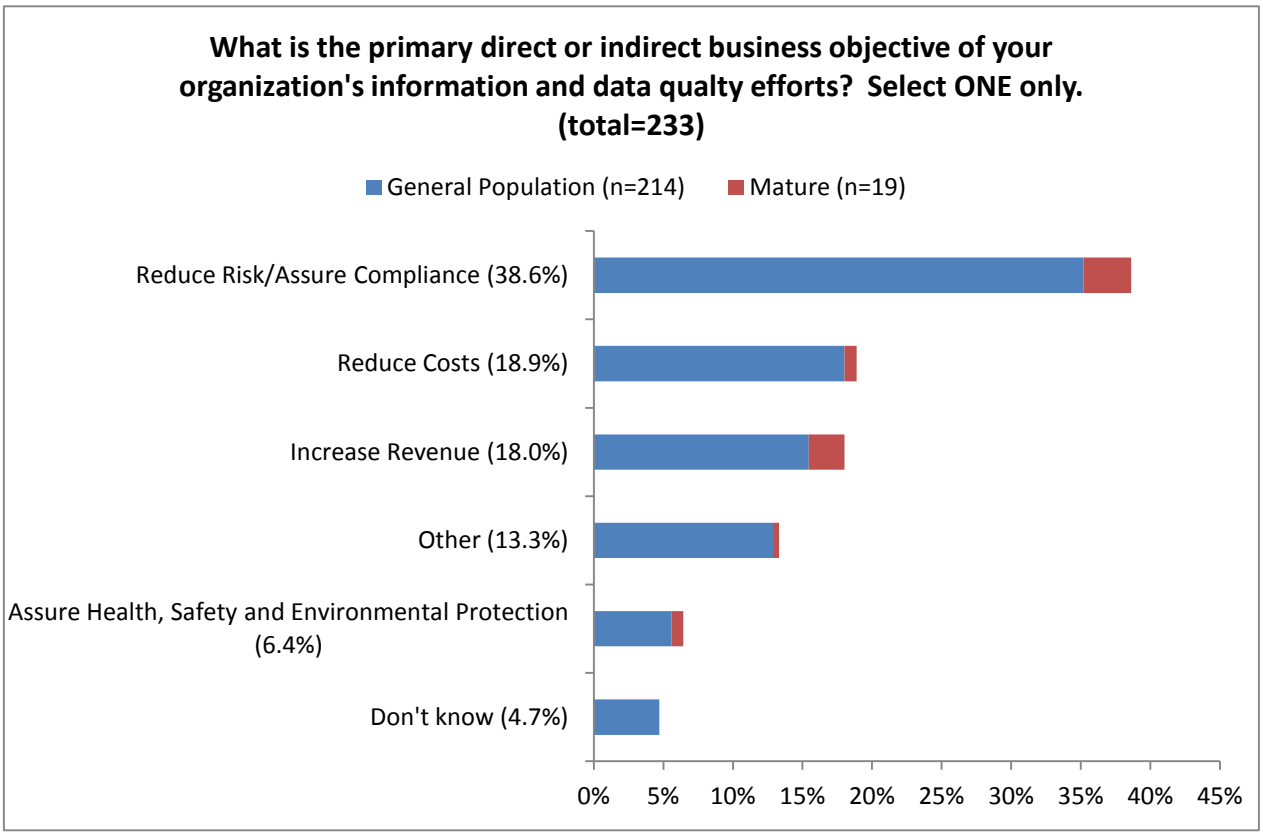
Note: Only one selection was allowed for this question.

Focus of Information and Data Quality (IDQ) Processes

This section of the report discusses the focus of organizations' information and data quality processes.

What are the Primary Business Objectives for IDQ Efforts?

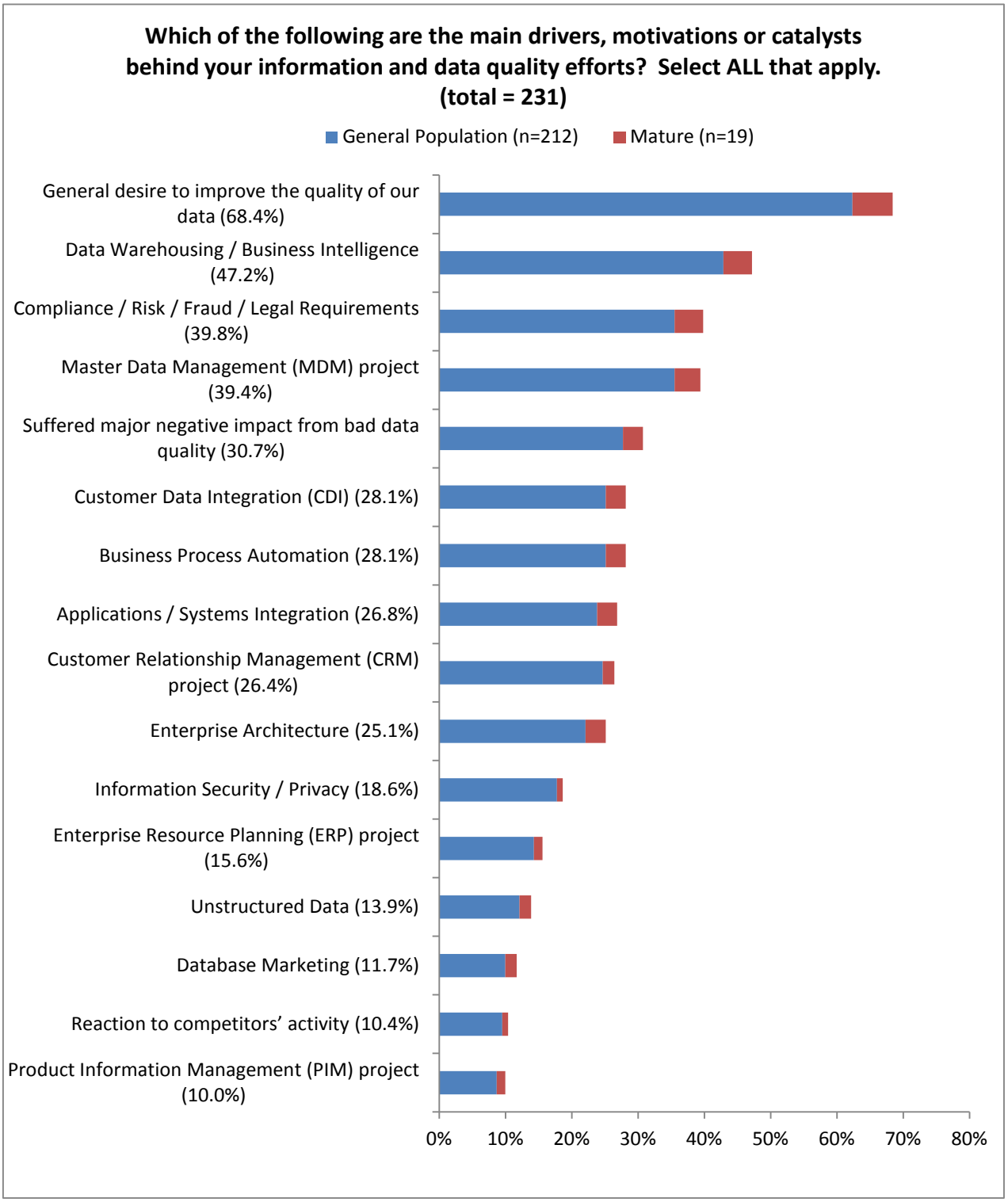
Linking information and data quality activities to business needs is essential. Our survey participants told us that the most common primary direct or indirect business objective of their organization's information and data quality efforts was to "Reduce Risk and Assurance Compliance" (38.6%). This objective was selected twice as much as "Reduce Costs" (18.9%) or "Increase Revenue" (18.0%). These results make us wonder if information quality efforts would be more supported and better sustained if they were more directly connected to the revenue/profit side of the financial statements. In addition to the objectives listed in the survey, people mentioned other objectives such as better customer satisfaction/service, improved business decision making, reliable reporting, and accurate information/data-based products.



What are the Main Drivers of IDQ Efforts?

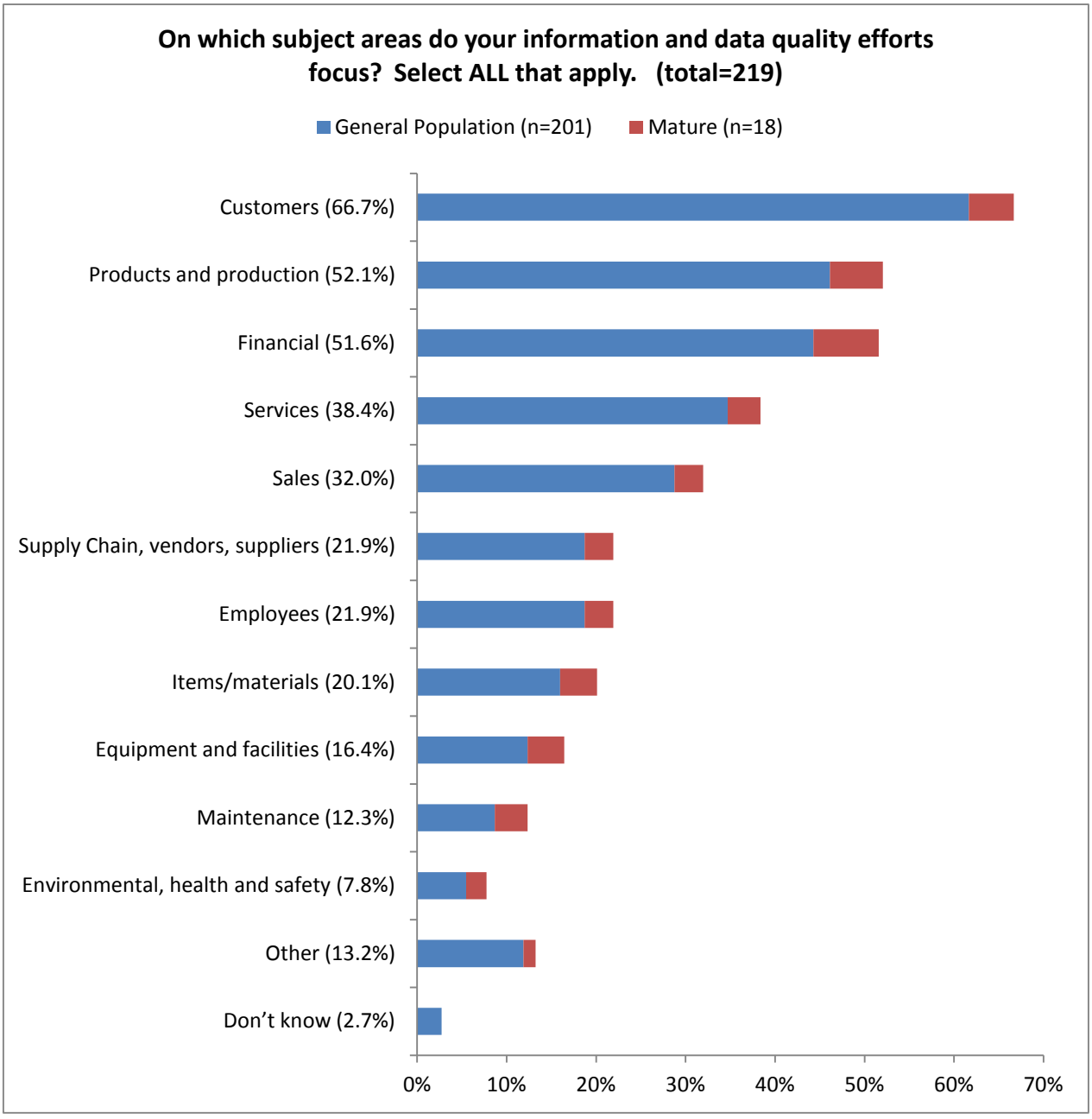
Survey participants named the general desire to improve the quality of their data as the main driver for their information and data quality efforts (68.4%). In addition to the items displayed in the following chart, people selected motivators such as sales force automation (9.5%), big data (9.1%), service-

oriented architecture projects (8.7%), mergers/acquisitions (8.2%), and cloud computing (2.2%). About 5% of people also wrote in other drivers such as the increasing complexity of the business, cost/asset management, process improvement, archival requirements, patient safety, profitability measurement and reporting requirements as other catalysts.



What is the Subject Area Focus of IDQ Efforts?

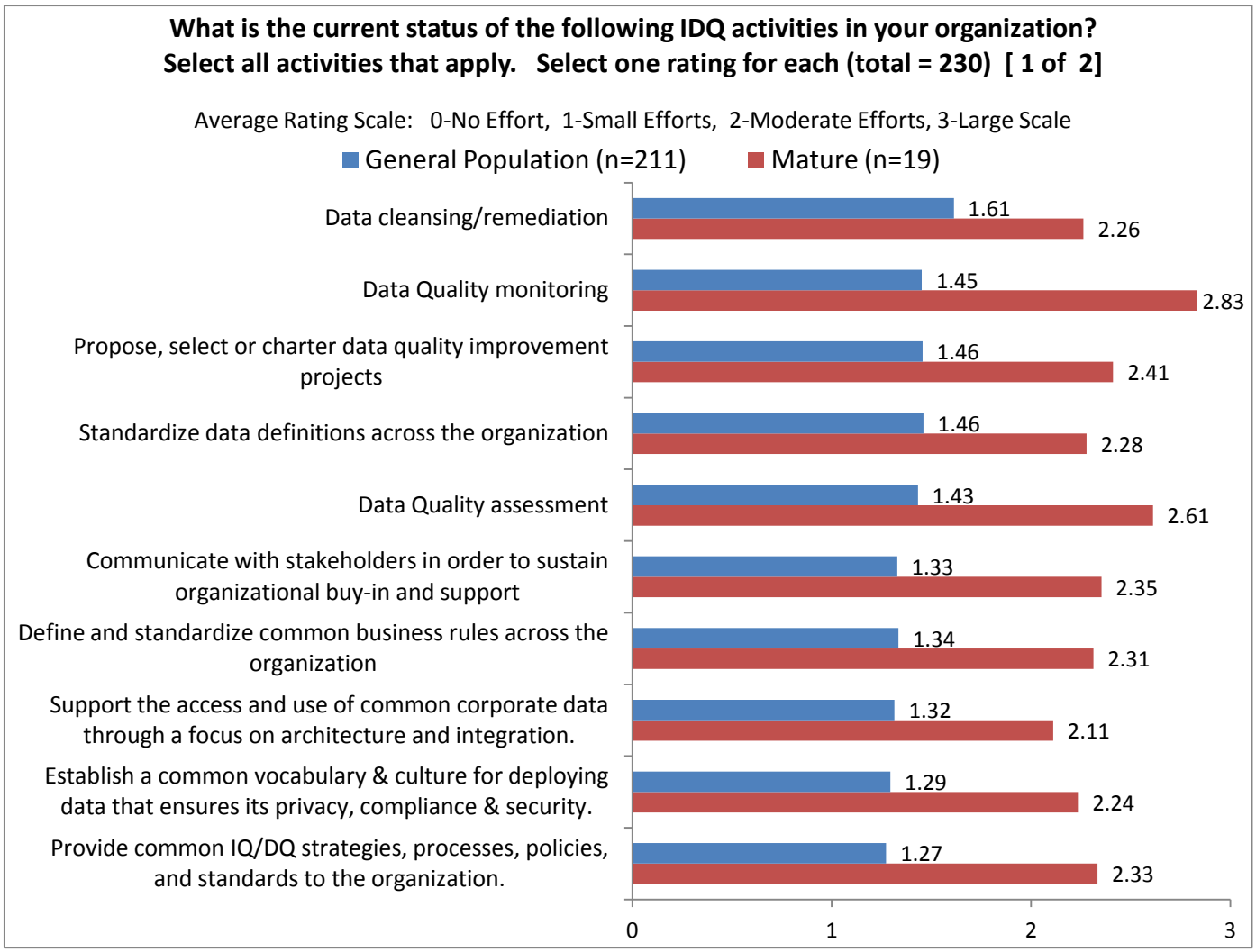
While Customers (66.7%), Products and Production (52.1%), and Financial (51.6%) were the subject areas most often cited by respondents, it is clear that organizations are seeking to improve the quality of a wide variety of data domains. These findings mirror those of our 2008 Information and Data Governance Survey. In addition to the subject areas that we listed, respondents provided several more subject areas such as Property/Real Estate, Education, Research/Scientific, Health Care/Patients, Regulatory Compliance, and Road/Transportation.



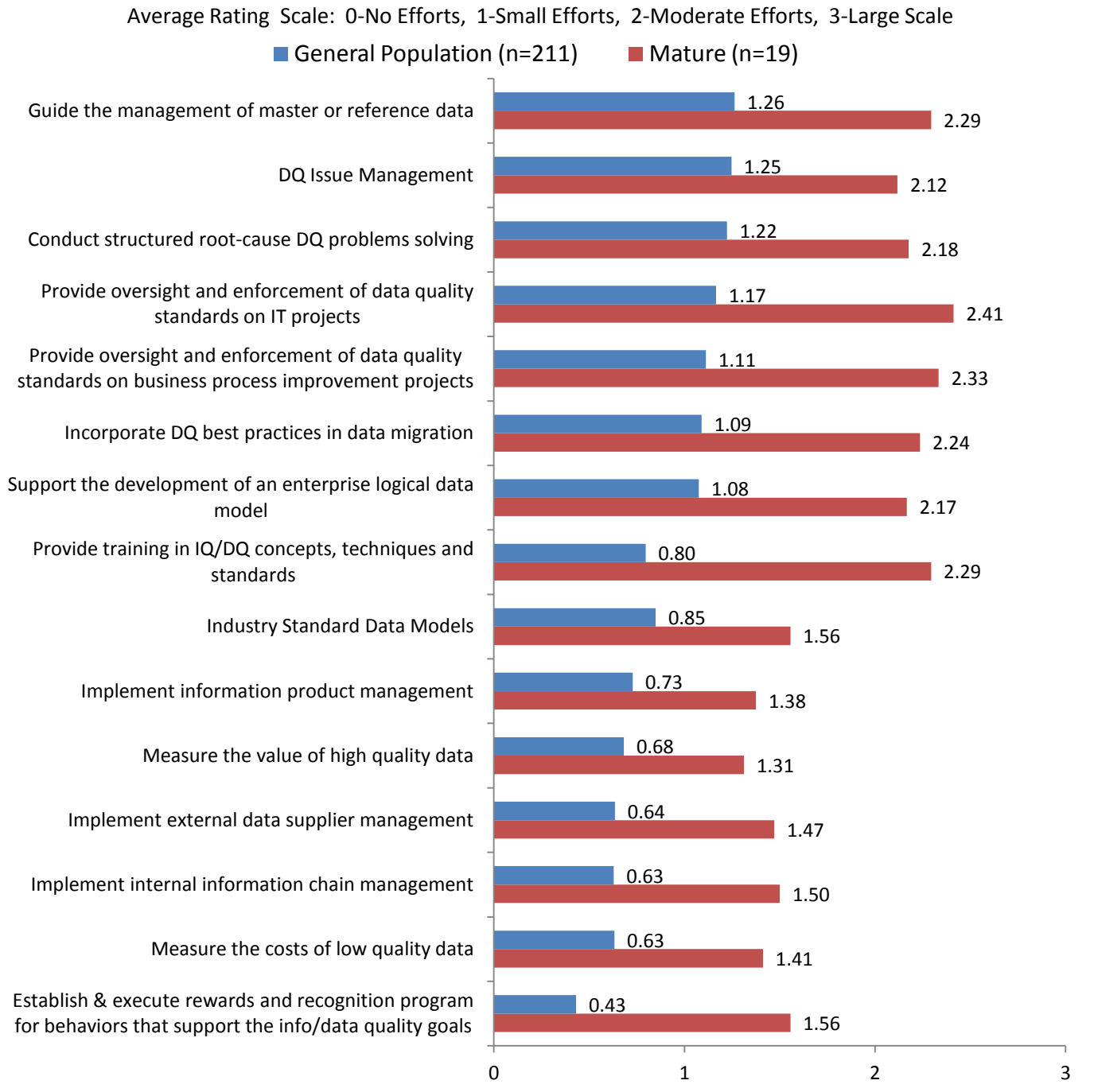
How Much Effort is Invested in IDQ activities?

We asked respondents to rate the amount of effort their organizations invest in various IDQ activities. Based on the average effort invested, the top five information and data quality activities of organizations overall were (in decreasing order): (1) Data cleansing/remediation, (2) Data Quality monitoring, (3) Propose, select or charter data quality improvement projects, (4) Standardize data definitions across the organization, and (5) Data Quality assessment.

Note: Mature IDQ organizations had a different top five list: (1) Data Quality monitoring, (2) Data Quality assessment, (3) Propose, select or charter data quality improvement projects, tied with Provide oversight and enforcement of data quality standards on IT projects and (5) Communicate with stakeholders in order to sustain organizational buy-in and support. We feel this last one is significant. In addition compared to the general population, mature IDQ organizations typically engage in larger efforts in each of the activities listed. This suggests that they embrace a broader, more full-scale approach to information/data quality management.



What is the current status of the following IDQ activities in your organization? Select all activities that apply. Select one rating for each (total = 230) [2 of 2]



Information and Data Quality (IDQ) Tools

This section of the report examines the tools that organizations are currently using for their IDQ efforts.

What Types of Tools are Being Used in IDQ Efforts?

The data quality tools market has grown rapidly over the past several years, increasing organizations' ability to assure data quality. According to our respondents, the top five categories of IDQ tools being used by organizations are:

(1) Data profiling and quality assessment
(2) Data quality monitoring
(3) Data remediation/cleansing
(4) Data matching and reconciliation (data de-duplication)
(5) Extract-Transform-Load (ETL) and other data integration tools.

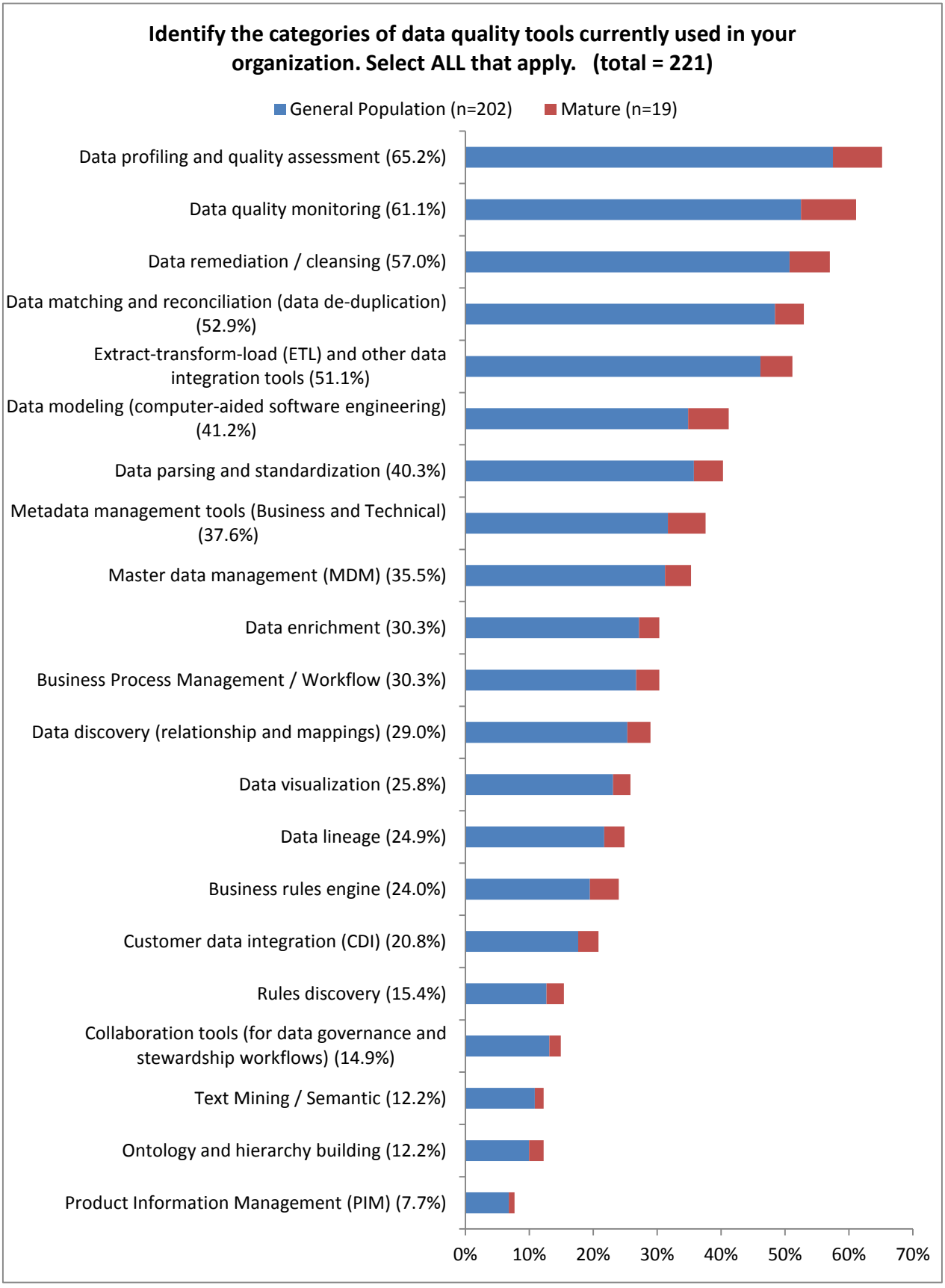
In addition to the ones listed, nearly 6% of survey participants wrote-in other categories such as "Statistical Analysis", "Microsoft Excel", "SQL scripts", "Quality Assurance/Quality Control" and "Program Management."

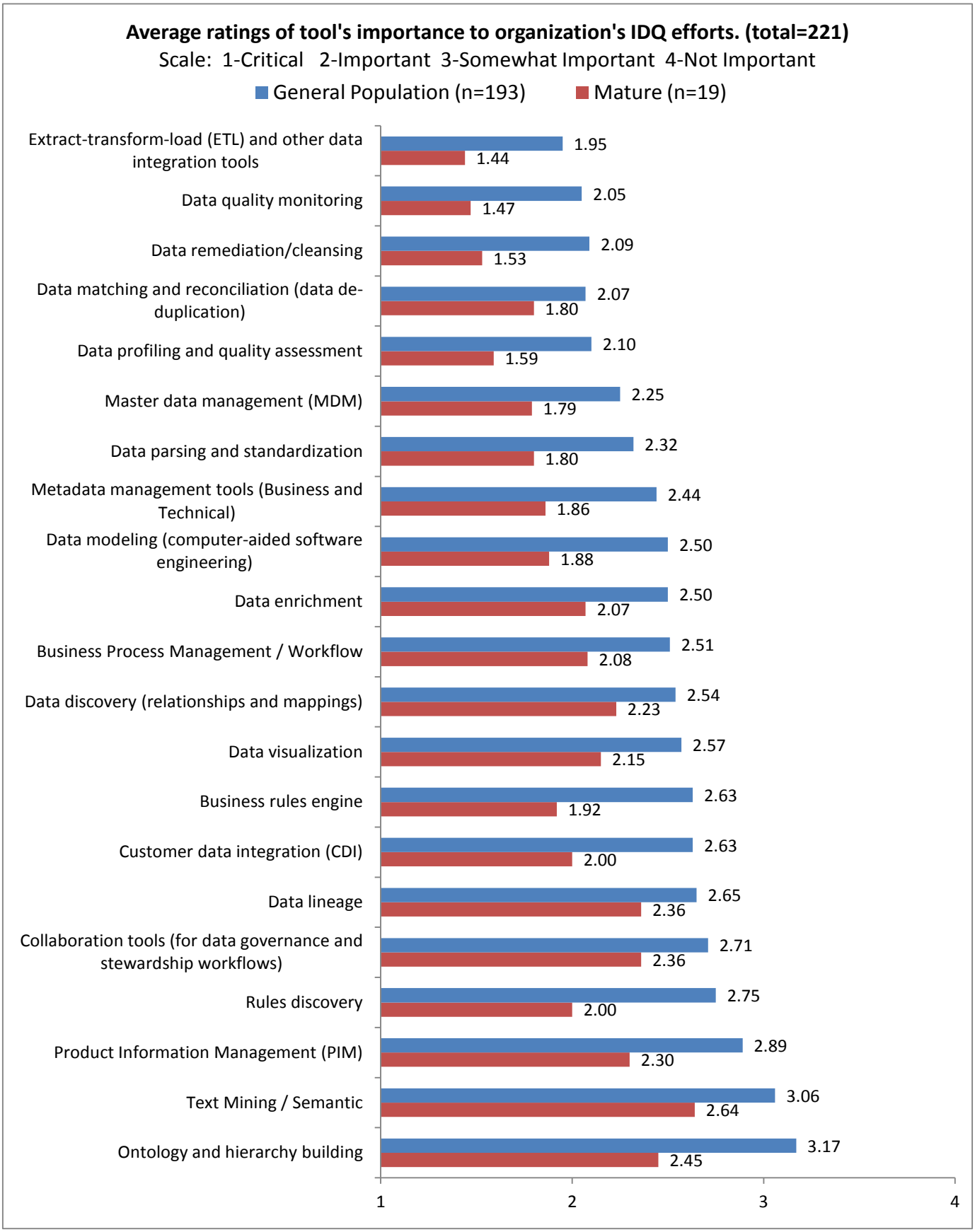
How Important are Tools to IDQ Efforts?

Survey respondents also rated the importance of the tools that they used to their organization's information and data quality efforts. Not surprisingly, the five tools most frequently used (listed in the previous question) were also rated the most important tools:

(1) Extract-Transform-Load and other data integration tools.
(2) Data quality monitoring
(3) Data remediation/cleansing
(4) Data matching and reconciliation (data de-duplication)
(5) Data profiling and quality assessment.

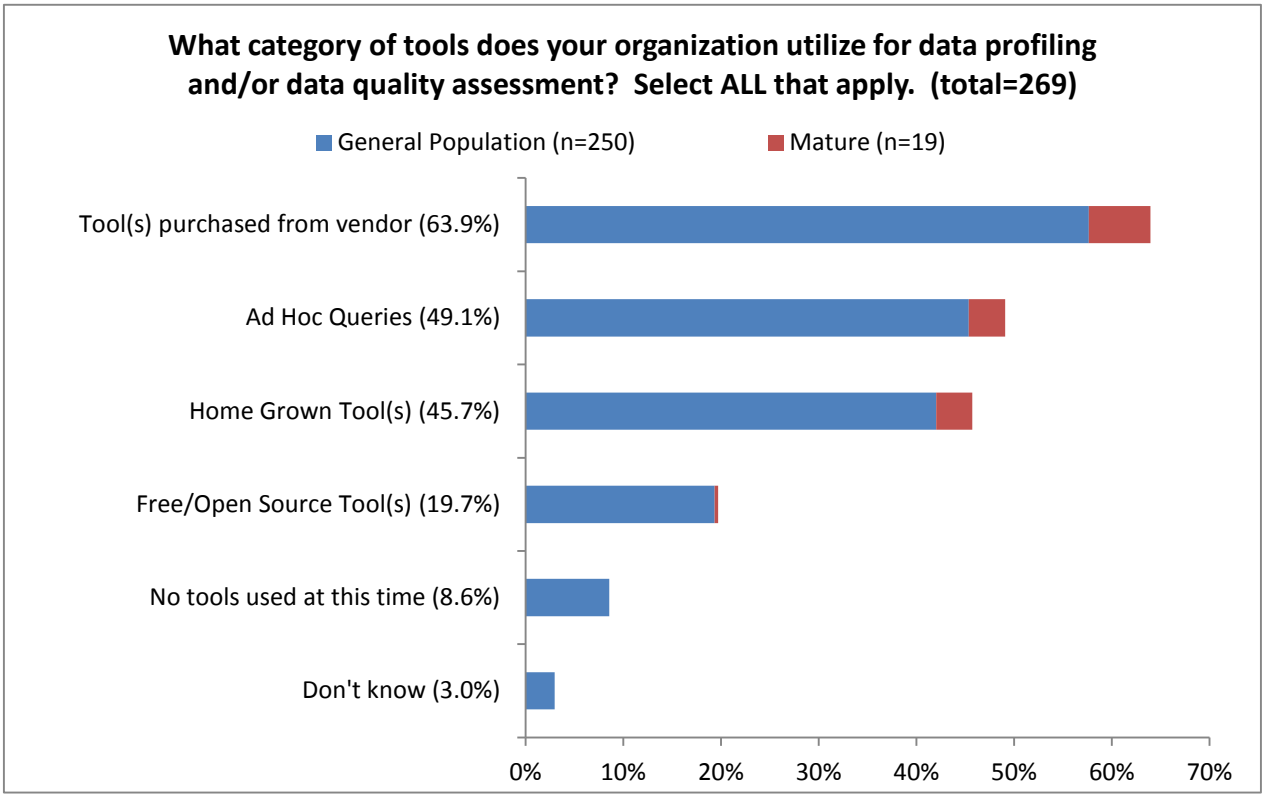
One interesting item to note is that the order of the five most important tools and that of the five most used tools are very similar with the following exception. While Data profiling and quality assessment tools were cited as the most used, respondents rated Extract-Transform-Load (ETL) and other data integration tools as the most important to their organization's information and data quality efforts. *Note: In comparing the detailed survey responses between mature IDQ organizations and the general population, we observed that mature IDQ organizations generally place greater importance on the utilization of tools as part of their IDQ efforts.*





Where do Organizations Get Their Tools for Data Profiling and Assessment?

Our survey responses revealed that most organizations are using tools purchased from vendors to conduct data profiling and data quality assessments (63.9%). In addition, close to half of our respondents reported using ad hoc queries and home grown tools as well.



We also asked survey participants to share with us the product brand and names of tools they get from vendors or open sources. The responses we received covered some 50 different products ranging from spreadsheets to sophisticated tool suites. Here are the names most frequently mentioned by our respondents:

- IBM and its InfoSphere Suite of Products (e.g., DataStage, QualityState, AuditStage, Information Analyzer, and Business Glossary)
- SAS and its DataFlux Products
- Informatica and its Master Data Management and Data Quality Tools Suite (Data Quality, Data Explorer, and AddressDoctor)
- SAP and its Data Quality Management Software Suite
- Talend Data Integration Software and Data Management Tools Suite
- Trillium and its suite of Software System Products
- Oracle and its Enterprise Data Quality Products.

Conclusion and Recommendations: The Agenda Forward

The main goal of this report was to understand and document how organizations are managing the quality of their information and data assets. We achieved this by examining current information and data quality trends, organizational structures, processes and tools, for organizations worldwide. Three findings stand out.

- Organizations across the globe and across all industries have adopted information and data quality as a means to stay in business (by reducing risk and improving compliance), or to improve profits (by reducing costs or increasing revenues). Information and data quality management is no longer a business secret and it is not a fad.
- A small but hopefully growing group of organizations have been able to sustain their focus on information and data quality over a long time, and as a result, they have achieved high levels in several or all the important information and data quality maturity areas.
- Unfortunately for various reasons, in all organizations, including the ones with mature information quality processes, significant obstacles continue to stymie progress and hurt the effectiveness of information quality efforts. These obstacles must be addressed and removed if information and data quality is to continue to make progress as a discipline, and most importantly, for organizations and society to be able to reap the considerable benefits of high information quality.

Action must be taken on several fronts.

For Practitioners in Organizations

1. **Learn from the organizations with mature information quality processes**. Adopt the best practices that have been implemented. Learn from the growing library of case studies becoming available through conferences. Use industry reports such as this one to benchmark internal practices and to set improvement goals for the future.
2. **Shift the focus to the revenue side.** One reason why IDQ efforts at many organizations are still at a low maturity level and struggling for executive support may be the perception that improving data quality is mainly viewed as a compliance objective rather than a revenue or cost reduction objective. While reducing risk will get information and data quality activities onto the business' agenda, it is important to demonstrate what data quality means in terms of financial gain or competitive advantage in order to get it to the top of the agenda and keep it there.
3. **Grow your skills**. The wheel does not have to be re-invented each time. Information and data quality conferences are now being held across the globe. Numerous information quality books have been published over the past few years, in many languages other than English, and more are on the way. There are opportunities to earn an IQ certification and to attend publically available information quality training courses.

4. **Leverage information and data quality consultants.** Expertise needs to be built in-house, however that does not happen overnight. External consultants brought in at the right time can be a strategic advantage to IDQ programs as they can help reduce the lead time to maturity.
5. **Leverage information and data quality-related professional associations**. They have beneficial products and services that can help you grow your knowledge and expertise

For Thought-Leaders in the Information/Data Quality Field, such as IAIDQ, Academics, Consultants, Leading Practitioners

Many of the obstacles that remain have at their root an awareness issue. Many organizations still lack a comprehensive understanding of the impacts that information and data quality have on their operations, employees and customers. Even when they are aware of the impacts, many organizations are not yet convinced that there exists a knowledge base of research and best practices that can be applied to improve the situation. Just as the American dairy farmers and milk processor industry used their "Got Milk?" campaign to encourage the consumption of cow's milk, IAIDQ and other IDQ leaders need to work together on a grass roots campaign to educate people on IDQ issues, to promote the development and exchange of the IDQ knowledge base, and to provide support and strategies for those trying to establish and grow an IDQ culture in their organizations. A coordinated grass roots effort could be conducted by IAIDQ and other IDQ leaders using a variety of cost efficient techniques. These efforts should not solely be directed towards IDQ professionals. It is critical that IDQ leaders reach out to across business, government, and academic communities to raise interest in addressing of IDQ issues using several communication strategies.

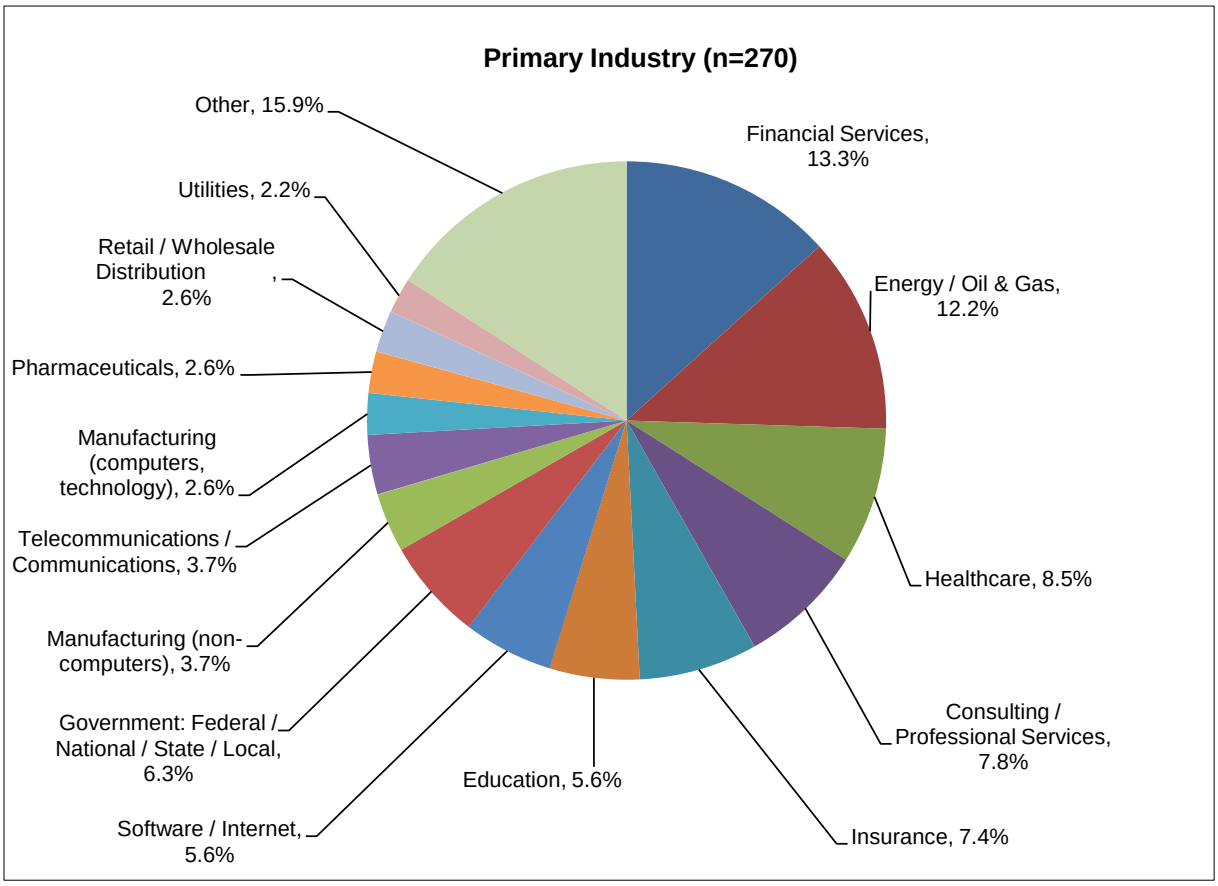
1. **Utilize social media.** Social media provides opportunities to serve, support, and sell IDQ through numerous channels. Wikis and directories can help people locate information on IDQ topics and events. Websites, webinars, YouTube videos, and white papers can help make the IDQ knowledge base accessible worldwide. Social networks such as LinkedIn can be leveraged more to help people network with other IDQ professionals.
2. **Foster face-to-face interactions.** While social media can help disseminate the IDQ message, face to face interactions can be more effective for fostering innovation and the adoption of IDQ principles and techniques. Conferences, professional training, communities of practice, and university course offerings are opportunities for people to congregate and to share their experiences and lessons learned regarding IDQ. Encouraging the formation of IDQ communities of practice in companies or across whole industries is another way to create a more peer-based structure for sharing IDQ best practices.
3. **Expand professional certification efforts.** The Information Quality Certified Professional (IQCP) credential sponsored by IAIDQ contributes greatly to the IDQ discipline. The certification process helps to define the scope of the discipline, to articulate skills and knowledge needed for the various IDQ task domains, and to provide a benchmark to test one's mastery of the IDQ

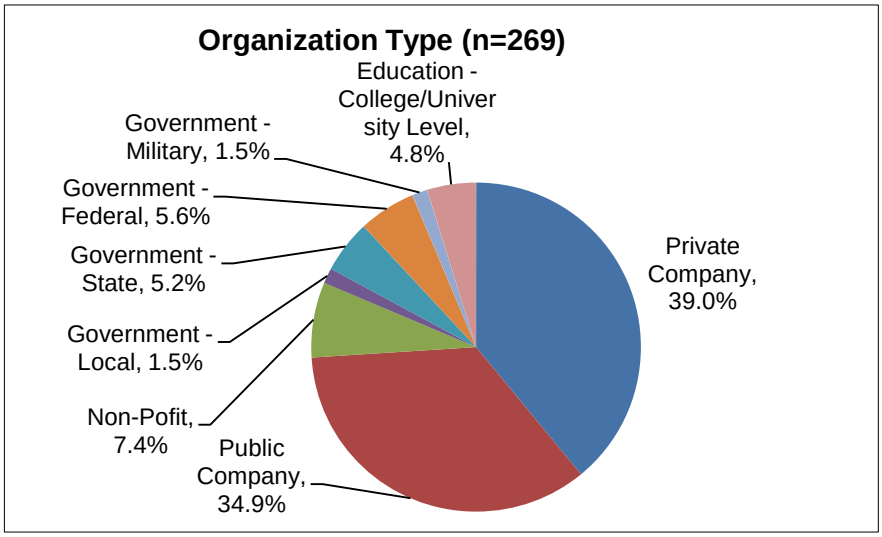
subject matter. The domains, tasks, knowledge and skills defined for the IQCP are valuable resources in the development of courses, training programs, and educational materials. These are essential components of continuous professional growth.

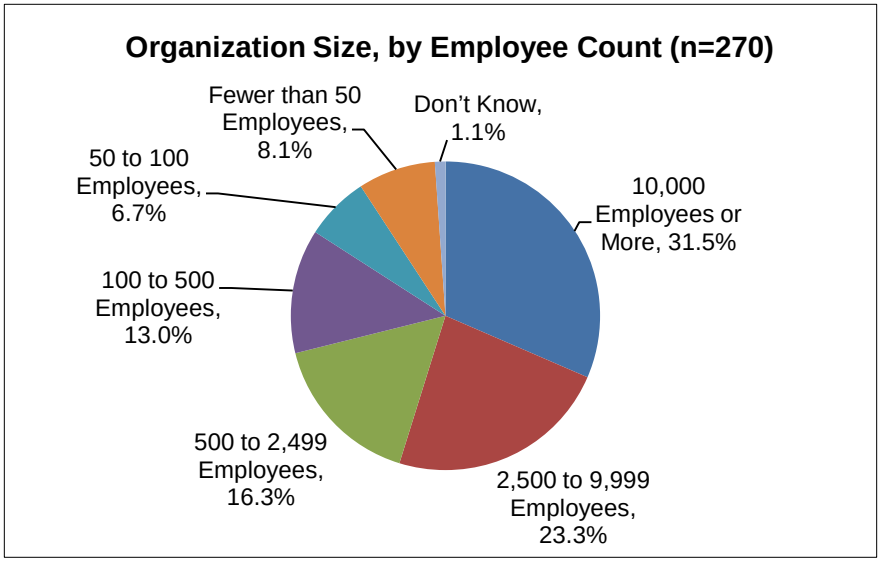
4. **Increase the availability of public IDQ training courses**. There is increasing demand for comprehensive and reliable IDQ training.
5. **Establish academic programs focusing on information and data quality**. Increase the number of IDQ courses taught at colleges and universities, and the number of program offering degrees in the discipline.
6. **Persuade others to get involved.** IDQ professionals can help to shape the future of IDQ. As alumni, IDQ professionals can work through their alma maters; encouraging academic programs in business and IT/IS to include topics on data governance and information quality management. IDQ professionals can offer talks through their local chamber of commerce, civic group, or Toastmasters Club on the impact that better data can have. By reaching out beyond the IDQ discipline, IDQ professionals can help motivate government, business, and community leaders to take steps to improve the quality of their data.

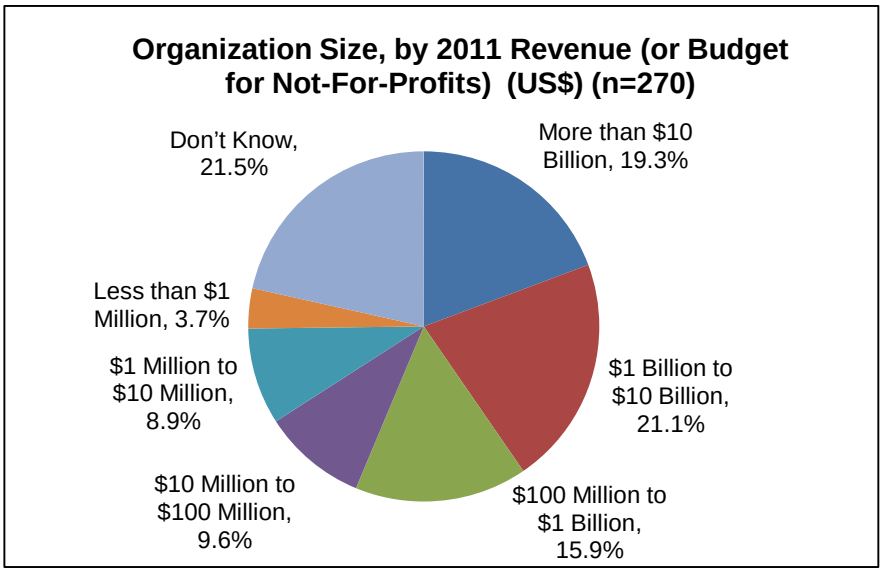
In closing, it is our hope that this report will be a catalyst for effective action along the lines we recommend above, and will promote an active discussion among IDQ professionals on where we stand when it comes to improving the quality of the information and data used in our organizations and how we as a profession can make greater advances in the future in order to better benefit society.

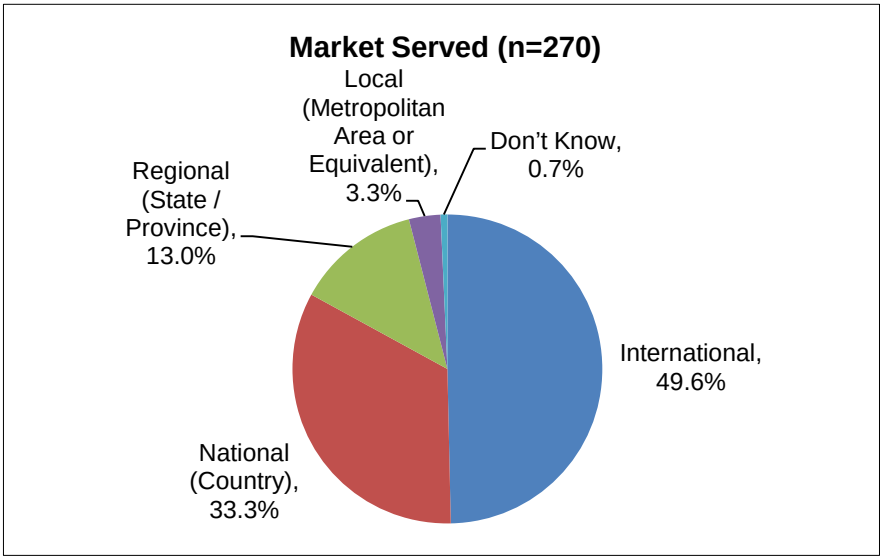
Appendix I. Extended Demographics

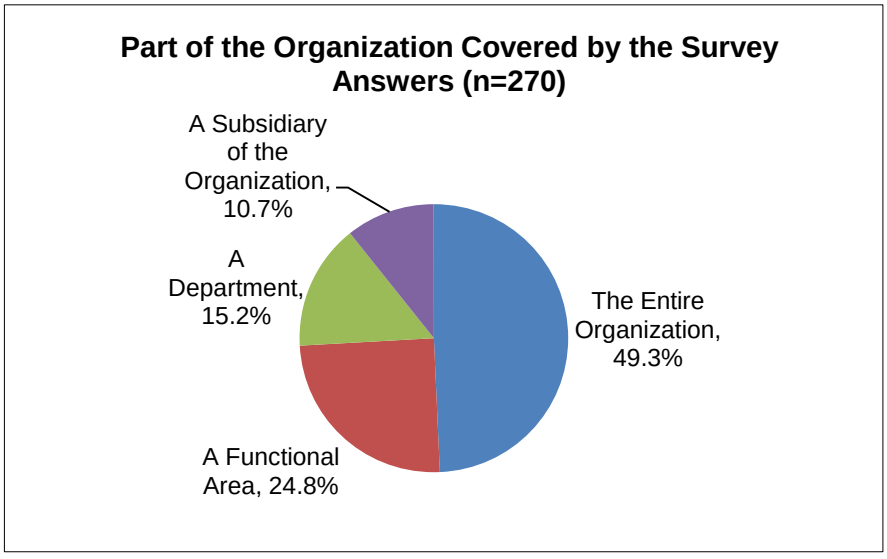




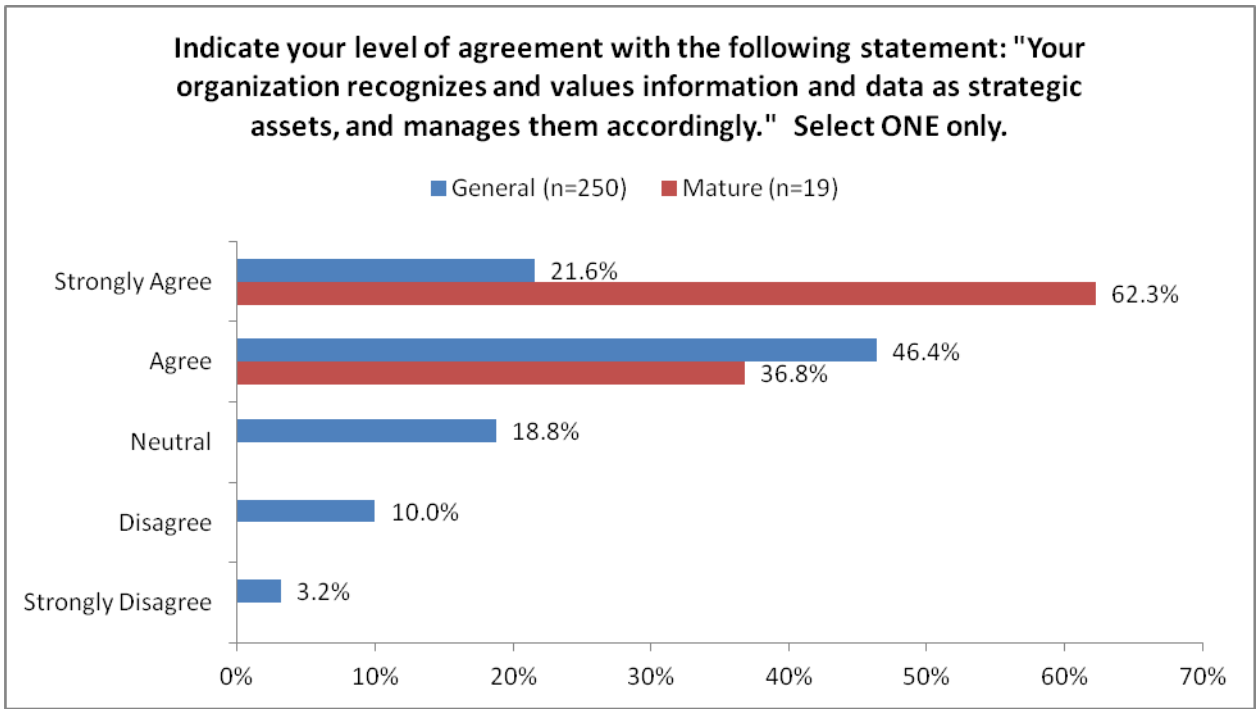


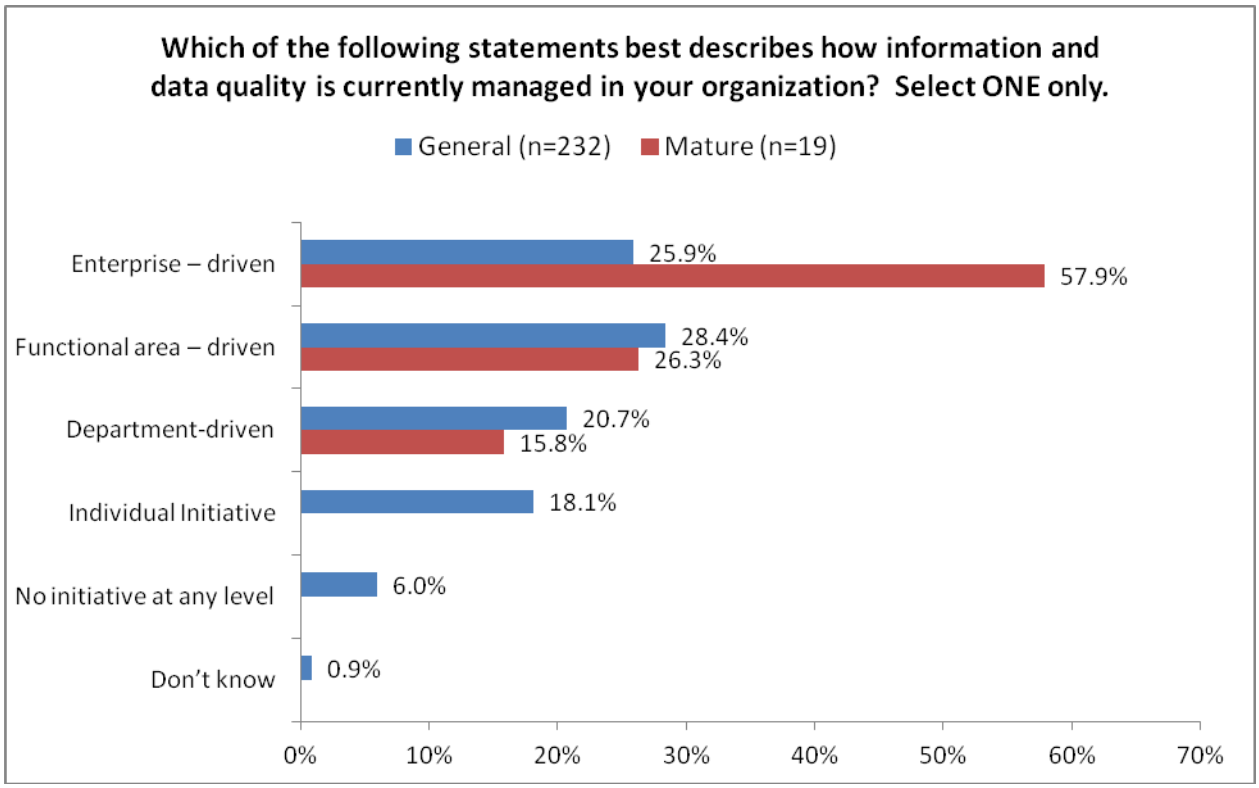


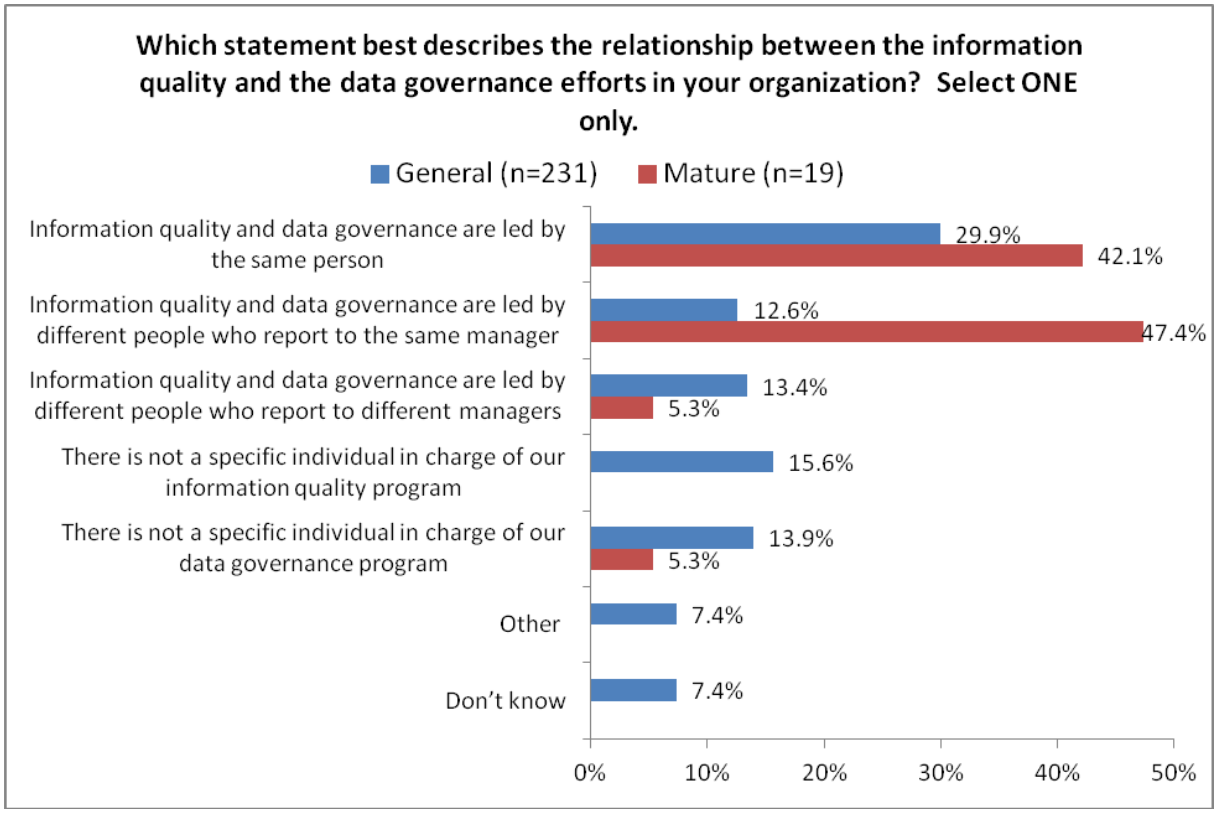


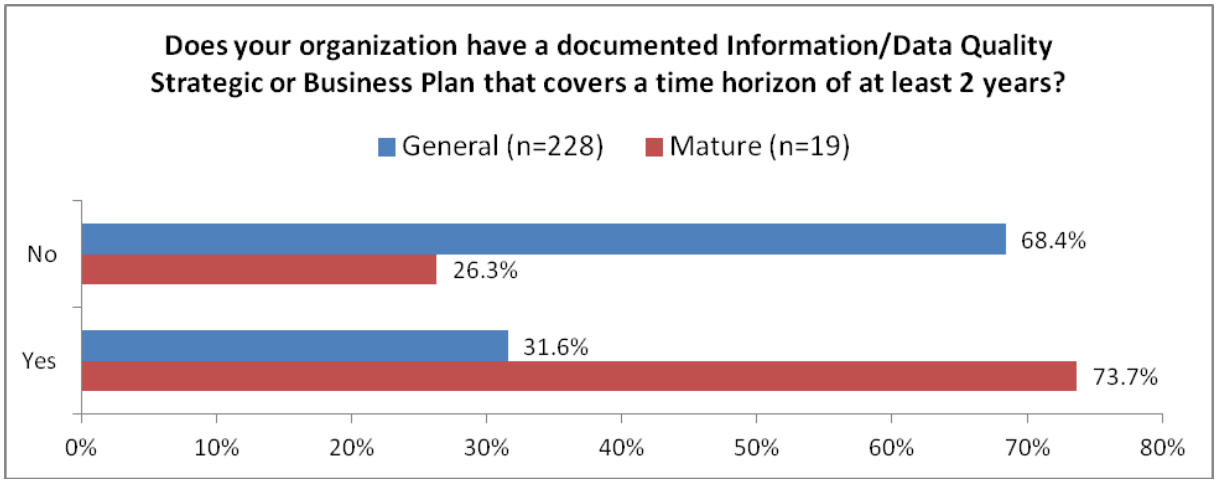


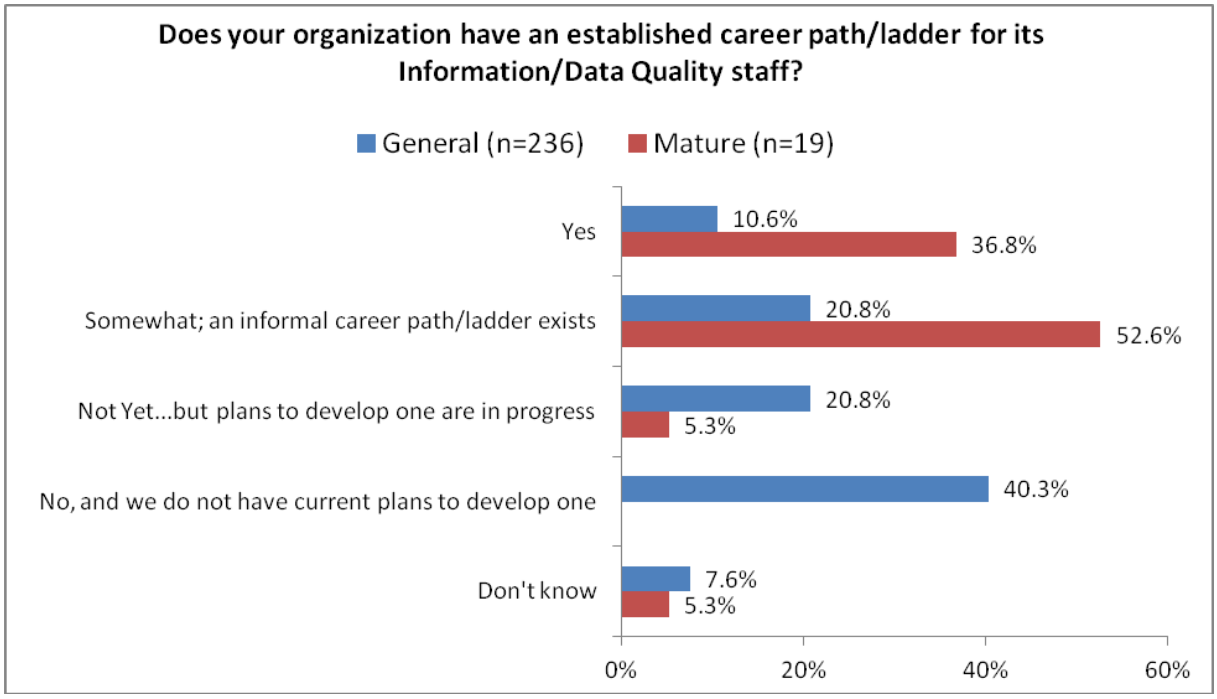
Appendix II. Extended Maturity Charts

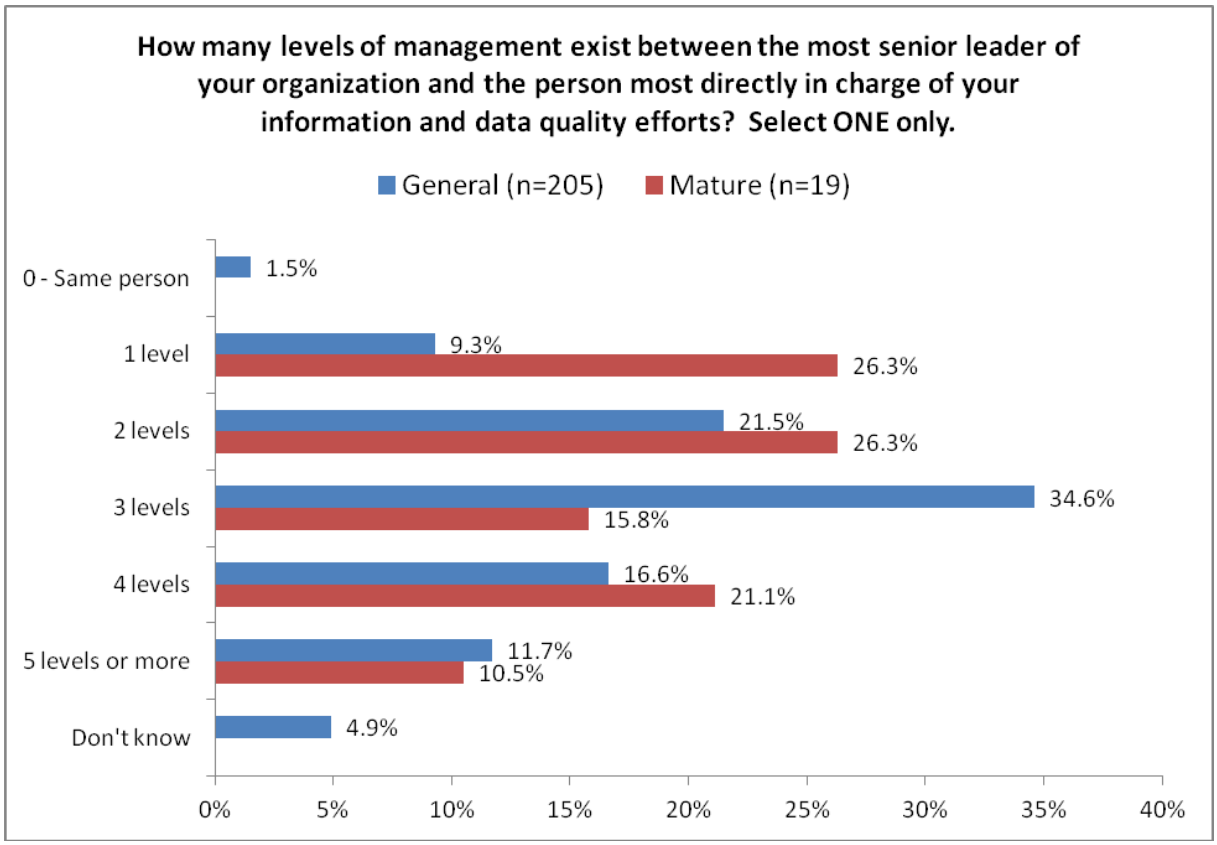


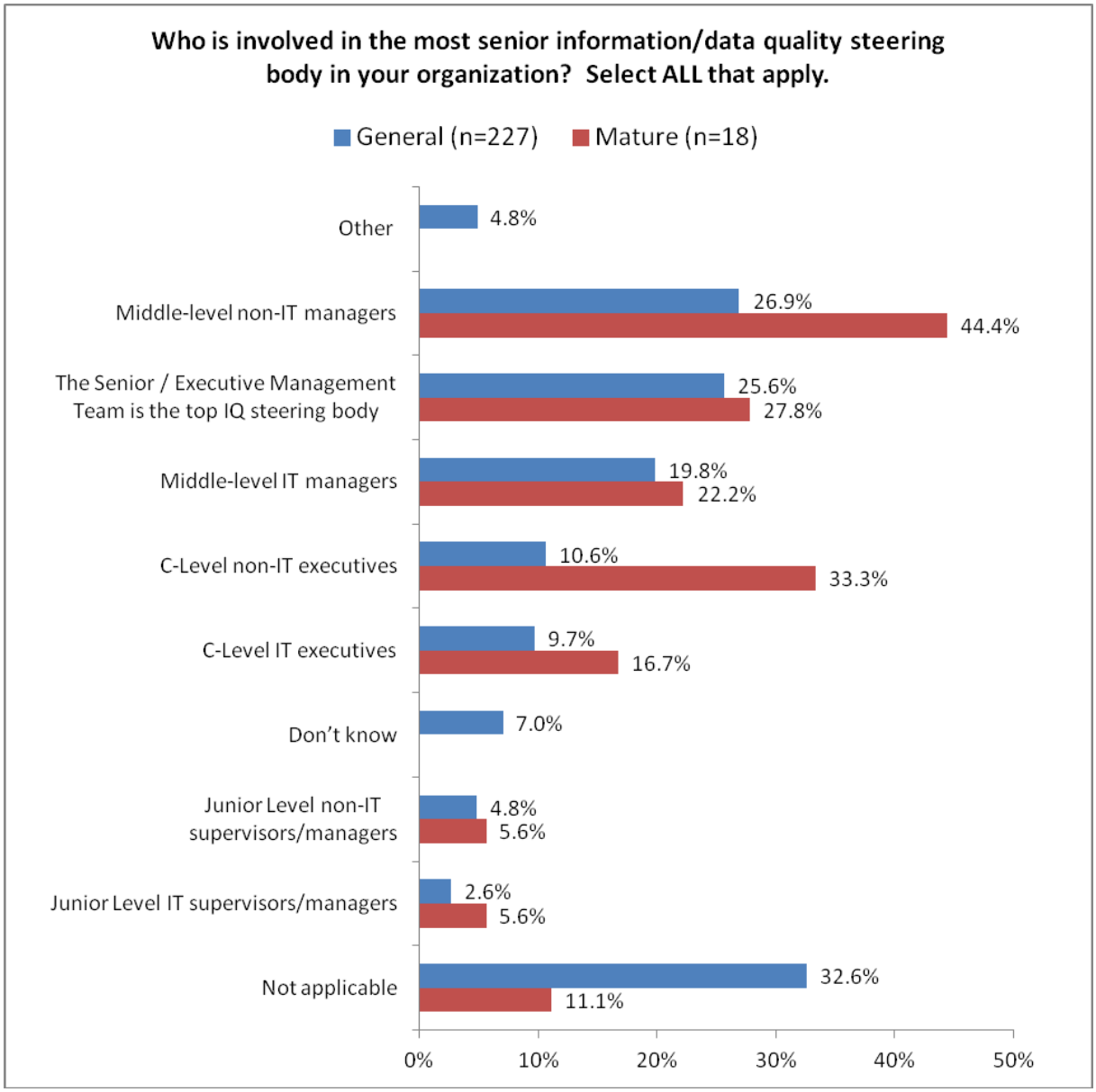


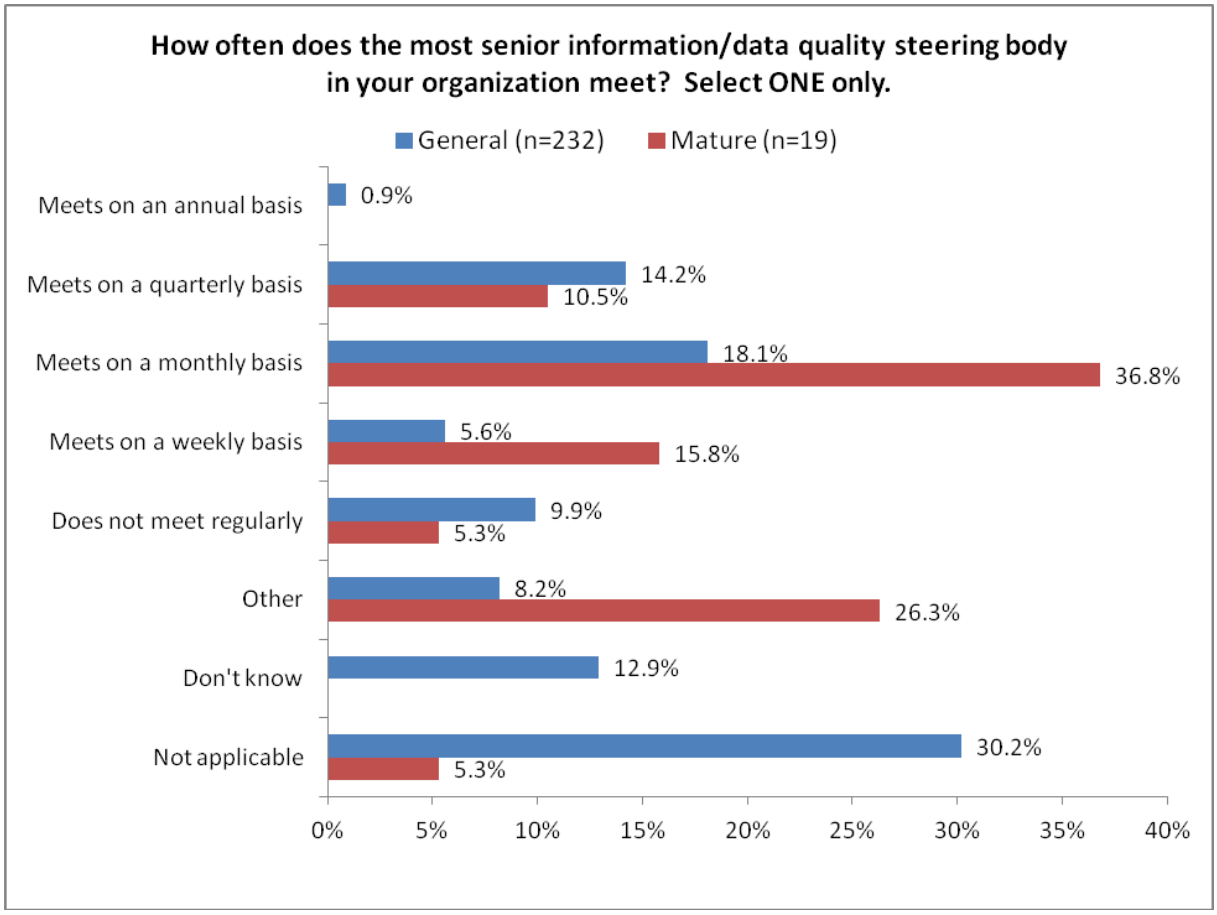


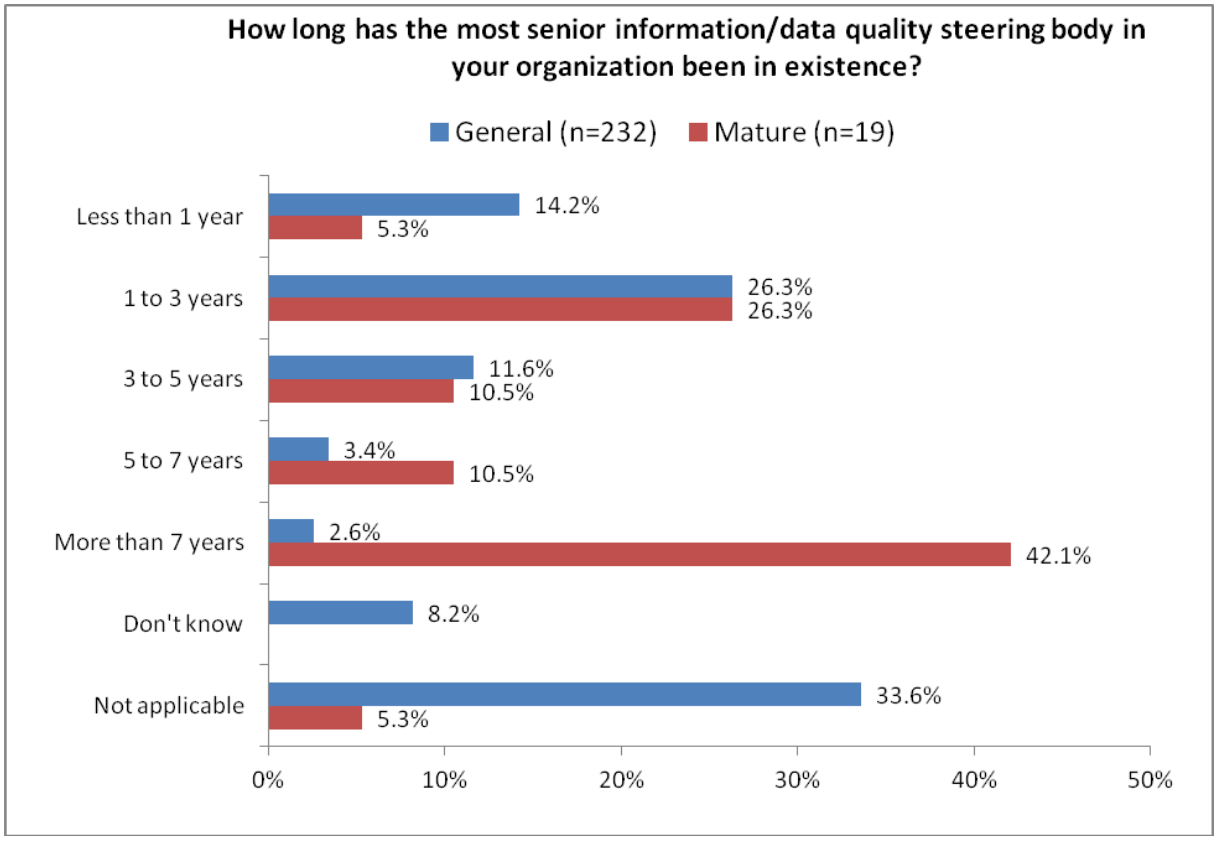












For other reports in this series, please visit

www.iaidq.org/publications

