Epilogue: The Data Quality Profession

Elizabeth Pierce, John Talburt, and C. Lwanga Yonke

Abstract In this final chapter, we will discuss four significant topics concerning the data quality profession. First, we will examine how the data quality profession has evolved. Second, we will explore what it means to be a data quality professional. Third, we will review the training opportunities currently available to those interested in becoming a data quality professional, and finally, we will assess the outlook for the future of the data quality profession. Throughout this chapter we will use the terms "data" and "information" interchangeably.

1   A Brief History of the Data Quality Discipline

The information and data quality discipline has had a relatively short but rapidly evolving history that can be thought of in five phases:

- Problem Recognition: The Data Cleansing Phase
- Root Cause Detection: The Prevention Phase
- Manufacturing Analog: The Information Product and Process Management Phase
- Information Architecture: The Quality by Design Phase
- Enterprise View: Information as an Organizational Asset Phase

E. Pierce (✉) · J. Talburt
University of Arkansas at Little Rock, Little Rock, AR, USA
e-mail: expierce@ualr.edu; jrtalburt@ualr.edu

C.L. Yonke
International Association for Information and Data Quality (IAIDQ), Baltimore, MD, USA
e-mail: lwanga.yonke@iaidq.org

1.1   Problem Recognition: The Data Cleaning Phase

Many of the more reactive practices of data and information quality still popular today emerged as a by-product of the data warehousing movement advocated by [1,2], and others. Data warehousing was a compelling idea, but like some many great concepts, it was easier to design than to implement. The biggest impediment to data warehouse implementation turned out to be the discovery by most organizations that the data in their operational data stores was in terrible condition. It was incomplete, inconsistent, inaccurate, out of date, unreliable, and plagued by all of the other problems that we now recognize as the symptoms of poor data quality. Because these data resided in many different systems across the organization, these problems mostly lay undiscovered until there was an attempt to integrate them into a single data warehouse.

Even though Total Quality Management (TQM) was in full swing in the manufacturing and services arenas, there seemed to be a different attitude toward data. Companies that were seeking six sigma bounds on product defects seemed to be satisfied with 10 %, 20 %, or even higher levels of defects in their data stores. Redman in [3] was one of the first to expose the extent of the problem and to quantify the impact that poor data quality was having on the organization in terms of operational cost and strategic planning.

As a result, the industry of data cleansing was born, actively pursued by many organizations. The limited benefits and high costs of that approach were documented by early data quality pioneers such as English [4] and Brackett [5]. Sometimes called data cleaning or data hygiene, data cleansing focused on the use of extract-transform-load (ETL) processes to standardize the data from different sources so that it could be merged into a single data warehouse and so that queries against the data would be meaningful. To facilitate the data inspection that precedes data correction, Lindsey, Olson [6, 7], and others began developing and promoting data profiling and other techniques as tools for performing data quality assessment.

1.2   Root Cause Detection: The Prevention Phase

In the next phase, data quality practitioners began to look toward the success of Total Quality Management (TQM) and to adopt some of its best practices. One of the first was root cause analysis. After the immediate need to cleanse data before building a data warehouse, organizations also began to see the value of preventing the same problems from reoccurring. Improvement projects for data quality were encouraged to put a greater emphasis on preventing future data errors first and then correcting existing data errors second [8].

1.3   Manufacturing Analog: The Information Product and Process Management Phase

Practitioners quickly realized that rather than simply adopting certain aspects of TQM, there was value in adapting the entire TQM paradigm to information, by applying manufacturing concepts to information systems [9–11] and to information processes [3, 4, 12–18].

The approach developed by Wang and his colleagues first focused on information systems, based on the view that information is the product of an information system, not a by-product [19]. By viewing data sources as raw materials, the software applications as the manufacturing process, and the final outputs as the products, then the full range of TQM principles could be applied to information systems. The result was the formulation of the Total Data Quality Management (TDQM) process [9, 19].

The approaches developed by Larry English and Tom Redman focused on defining, managing, and improving the business and IT processes through which data is created, captured, stored, delivered, used, and retired. The result was the formulation of Total Information Quality Management (TIQM) [4, 12] and the Second-Generation Data Quality Systems framework, with its specific focus on information chain management and data supplier management [17, 18].

Perhaps the most important consequence of applying the disciplines of product and process management to information is that they brought into consideration the uses and users (customers) of information. Whereas data cleansing and root cause analysis internally focused only on the data itself, with the product and process approaches, it became important to understand the information customer's perspective on the value and usefulness of the information.

1.4   Information Architecture: The Quality by Design Phase

By the time that data and information quality practices began to emerge, software development was a relatively mature process. A well-known principle of software development is that the earlier in the development process that a problem is discovered, the less effort is required to correct it. It is also reflected in Deming's 14-point plan for TQM that quality must be built in from the beginning and not inspected out at the end [20].

Through the first three phases of data cleansing, prevention, and product view, data quality practices were primarily reactive, dealing with problems and issues that were already in place in the organization. The fourth phase represents a more active role of practitioners and researchers in which they have an influence on the initial design of data models and information architectures by keeping data quality in mind from the start [4].

1.5   Enterprise View: Information as Organizational Asset Phase

In the fifth and current phase, there is a growing recognition of data and information as an organizational asset and resource and that data and information quality principles and practices are a critical part of maximizing the value of that asset. As a result, the focus of data quality efforts is progressively shifting from the cost side of financial statements to the revenue side [16]. As adoption of this enterprise view of data quality increases, concepts and practices identified several years ago are becoming more broadly accepted and are being refined. Perhaps one of these fundamental concepts is data stewardship, the recognition that the data are not owned by individuals or departments in an organization but that everyone is entrusted with specific responsibilities for its care and keeping (stewardship) for the good of the entire organization [4]. Closely related to data stewardship is data governance (DG), the rules and policies for making decisions about data management made by the members of a data governance council whose members represent the data stakeholders in the organization [15, 17]. A relatively newer concept is master data management (MDM), the processes, policies, and procedures around the management of the data describing employees, customers, products, facilities, equipment, and other entities that are most critical to an organization: its master data [21].

It is important to remember that these phases only represent the evolution in the understanding of data and information quality; it doesn't mean that everyone is at the same level. Quite the contrary, just as the Capability and Maturity Model (CMM) for software development defines five levels of maturity, most organizations are still operating at Levels 1 or 2. Similarly, the data quality programs in most organizations are still focused on basic data cleansing with perhaps some root cause analysis. Few have yet to put into place a comprehensive enterprise-wide data quality program incorporating the components that comprise the higher levels of data and information quality [12, 16, 22].

1.6   What's Next?

The next logical step in the evolution of the data and information quality is an expansion of the current paradigm from a single enterprise to multiple organizations. Movement in this direction has already been signaled by attempts at creating various data standards such as XBRL (eXtensible Business Reporting Language) for the finance industry and the APCD (All-Payer Claims Database) and NCPDP (National Council for Prescription Drug Programs) standards for healthcare. Another is the ISO 8000–110: 2009 on Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification. All are focused on either standardizing or establishing rules for how to standardize information shared between organizations. Data quality will also need to evolve to ensure it

appropriately addresses issues arising from new data create and use processes, such as social media.

2 The Job of the Data Quality Professional

As the data quality field has evolved, so too have the roles, responsibilities, and training of individuals engaged in the pursuit of better data for their organizations. As part of its efforts to better understand the data quality profession and to develop standards for data quality professionals, the International Association for Information and Data Quality (IAIDQ) has sponsored several important studies. The first one was a role delineation/job analysis study conducted between October 2008 and March 2009, to serve as the basis for the Information Quality Certified Professional (IQCP[SM]) credential developed by the International Association for Information and Data Quality, IAIDQ (http://iaidq.org) [23,24]. This study extended work previously done by several others [13,25,26]. The second study was the very first salary and job satisfaction survey for the data quality profession, conducted in 2009 by a team of investigators from IAIDQ and the University of Arkansas at Little Rock [27]. That 2009 survey also included questions about how data quality professionals spent their time at work, what characterized their work environment, and where they obtained their educational backgrounds. Much of the material for this section of the chapter originates from those two studies.

2.1 Recognizing the Data Quality Professional

Unlike more mature fields such as accounting, law, or medicine, there are no consistent degrees, job titles, or state license boards to rely upon when it comes to identifying a data quality professional. Of the 120 job titles self-reported by individuals engaged in data quality activities in IAIDQ's 2009 survey, fewer than 30 % contained either the phrase "Information Quality" or "Data Quality." About 40 % of the titles contained only the word "Information" or "Data" but no reference to "Quality." Of the remaining titles, 2.5 % included the word "Quality" (but not "Information" or "Data"), and 27.5 % have no reference to any of the words "Information," "Data," or "Quality."

To complicate matters further, activities associated with the creation, management, consumption, and improvement of data are performed by many individuals in an organization, not just a few dedicated specialists. Moreover, even those charged with data quality-related duties may have other job responsibilities not related to data quality.

As a result of these difficulties, for certification purposes, IAIDQ defined data quality professionals as people who "hold any of a wide range of positions in their organizations, as individual contributors or as managers. They conduct, lead, champion or participate in information quality projects. They work in any of

the functions or disciplines within their organization or are part of a specialized information quality team; yet all perform information quality activities as part of their job responsibilities. This information quality work is either part-time within a broader organizational role, or on a full-time basis" [23].

The 2009 salary and job satisfaction survey provided further insights on how data quality professionals allocate their time across the six performance domains. The six domains are listed here in order of decreasing time spent.

2.1.1 Information Quality Measurement and Improvement Domain

The Information Quality Measurement and Improvement domain covers the steps involved in executing data quality projects. Activities include gathering and analyzing business requirements for data, assessing the quality of data, determining the root causes of data quality issues, developing and implementing information quality improvement plans, preventing and correcting data errors, and implementing information quality controls. Of all the domains, this one was cited as the set of work most frequently performed by individuals responding to IAIDQ 2009 salary and job satisfaction survey [27].

2.1.2 Information Architecture Quality Domain

The Information Architecture Quality domain is comprised of the tasks that assure the quality of the data blueprint for an organization. Activities include participating in the establishment of data definitions, standards, and business rules; testing the quality of the information architecture to identify concerns; leading improvement efforts to increase the stability, flexibility, and reuse of the information architecture; and coordinating the management of metadata and reference data. This domain along with the next (Sustaining Information Quality domain) was the second set of work most frequently performed by individuals responding to IAIDQ 2009 salary and job satisfaction survey [27].

2.1.3 Sustaining Information Quality Domain

The Sustaining Information Quality domain focuses on putting in place processes and management systems that ensure ongoing information quality. Duties include acting as an information quality consultant for integrating data quality activities into other projects and processes (e.g., data conversion and migration projects, business intelligence projects, customer data integration projects, enterprise resource planning initiatives, or system development life cycle processes) and continuously monitoring and reporting data quality levels. This domain along with the previous (Information Quality Architecture domain) ranked second in frequency of performance by individuals responding to IAIDQ 2009 salary and job satisfaction survey [27].

2.1.4 Information Quality Strategy and Governance Domain

The Information Quality Strategy and Governance domain includes the efforts that provide the structures and processes for making decisions about an organization's data as well as ensuring that the appropriate people are engaged to manage information throughout its life cycle. Activities include working with key stakeholders to define and implement information quality principles, policies, and strategies; organizing data governance by naming key roles and responsibilities; establishing decision rights; and building essential relationships with senior leaders in order to improve information quality. This domain along with the next (Information Quality Value and Business Impact domain) was the third set of work most frequently performed by individuals responding to IAIDQ 2009 salary and job satisfaction survey [27].

2.1.5 Information Quality Value and Business Impact Domain

The Information Quality Value and Business Impact domain consists of the techniques used to determine the effects of data quality on the business as well as the methods for prioritizing information quality projects. Activities include evaluating information quality and business issues, prioritizing information quality initiatives, obtaining decisions on information quality projects, and reporting results to demonstrate the value of information quality improvement to the organization. This domain along with the previous (Information Quality Strategy and Governance domain) ranked third in frequency of performance by individuals responding to IAIDQ 2009 salary and job satisfaction survey [27].

2.1.6 Information Quality Environment and Culture Domain

The Information Quality Environment and Culture domain provides the background that enables an organization's employees to continuously identify, design, develop, produce, deliver, and support the information quality that customers need. Activities include designing information quality education and training programs, identifying career paths, establishing incentives and controls, promoting information quality as part of business operations, and fostering collaborations across the organization for the purpose of engaging people at all levels in information quality strategies, principles, and practices. This domain was the least frequently performed by individuals responding to IAIDQ 2009 salary and job satisfaction survey [27].

2.1.7 Career Paths for Data Quality Professionals

The 2009 salary and job satisfaction survey also shed light on career paths for data quality professionals. In terms of their career experiences, data quality professionals can be found around the globe, working in nearly every industry with

Table 1 Industries employing data quality professionals

Which of the following best describes your organization's primary industry? ($n = 115$)

Category	Number of responses	Percentage of responses
Banking/financial services	18	16.5 %
Consulting/professional services	14	12.8 %
Insurance	11	10.1 %
Retail/wholesale distribution	8	7.3 %
Energy/oil and gas/utilities	8	7.3 %
Healthcare	8	7.3 %
Government: state/local	7	6.4 %
Government: federal/national	6	5.5 %
Manufacturing	6	5.5 %
Telecommunications/communications	6	5.5 %
Aerospace	4	3.7 %
Software/Internet	3	2.8 %
Education	2	1.8 %
Food/beverage/agriculture	2	1.8 %
Biotechnology/pharmaceuticals	1	0.9 %
Advertising/marketing/public relations	1	0.9 %
Chemical	1	0.9 %
Law	1	0.9 %
Logistics/transportation	1	0.9 %
Nonprofit/trade association	1	0.9 %

banking/financial services (16.5 %), consulting/professional services (12.8 %), and insurance (10.1 %) as the most frequently cited. Industries like banking, financial services, and insurance handle huge amounts of data that are subject to stringent regulations regarding the accuracy and privacy of that data. This provides these firms with a strong incentive for hiring individuals with specialized expertise in data quality. While some of this expertise comes from in-house sources, engaging an external consulting/professional services firm is another avenue for organizations to acquire the data quality expertise that they need. In addition to the industries listed in Table 1, other industries employing data quality professionals contributed by IAIDQ survey participants include IT Services (e.g., Systems Integration/Outsourcing, Computing and Services), Audit/Consulting Services, Consumer Product Goods, and Research (e.g., scientific surveys, clinical trials) [27].

Most data quality professionals report working full time for either public companies (43.6 %) or private companies (32.5 %). Another 9.4 % are employed by nonprofit organizations, and 14.5 % have jobs with some government agency. Data quality professionals typically work for larger organizations with nearly a third responding that they work for organizations with 50,000 or more employees (31.6 %). This is further borne out by the revenue size of organizations with almost a third of them estimating their organization's 2008 annual revenues to be more than $10 billion (30.4 %) [27]. Although good-quality data is of value to organizations of any size, it is typically the larger organizations that currently have the resources and

Table 2 Departments housing data quality professionals

In which department is your position located? ($n = 105$)		
Category	Number of responses	Percentage of responses
Information Technology/information Systems (IT/IS)	60	57.1 %
Accounting/finance	11	10.5 %
Marketing/sales	6	5.7 %
Audit/compliance/risk	5	4.8 %
Production/operations/maintenance	5	4.8 %
Quality assurance	4	3.8 %
Design/engineering	2	1.9 %
Research and development	2	1.9 %
Human resource	2	1.9 %
Purchasing/supply chain management	2	1.9 %
Legal	1	1.0 %
Process management	1	1.0 %
Not applicable: I am self-employed	4	3.8 %

economies of scale to hire individuals specifically designated for data governance and stewardship roles.

In terms of where data quality professionals are placed in an organization, over half of the data quality professionals surveyed (57.1 %) said their position is located within an Information Technology/Information Systems department (IT/IS). This in turn means that 42.9 % of the survey respondents do not work in IT/IS. This clearly dispels the myth held by many that data quality is solely a responsibility of the IT/IS discipline.

Among the non-IT/IS departments, accounting/finance (10.5 %), Marketing/Sales (5.7 %), Audit/Compliance/Risk (4.8 %), and Production/Operations/Maintenance (4.8 %) were the most common business areas that employed data quality professionals. As of yet, very few organizations have a unit devoted entirely to data quality activities. A review of Table 2 summarizing the entire collection of survey responses revealed that although the IT/IS function was named most often, the range of areas where data quality professionals are employed is extensive, encompassing nearly all parts of an organization. This is an indication that to be successful, data stewardship must be a shared responsibility between the IT/IS group and all the organization's business units [27].

3   Training for Data Quality Professionals

Given the wide range of roles and responsibilities covered by the six domains identified by IAIDQ, how does one prepare for a career devoted to the improvement of data quality? Because data quality concepts and techniques were developed by

industry practitioners, until recently, there has not been a widely accepted body of knowledge and a common vocabulary for describing a skill set for data quality. Whereas a discipline like computer science has a long history of academic research and publication in topics such as algorithm design, the theory of computation, and proof of correctness for algorithms, the same cannot be said for the data quality field. However, this is beginning to change as more conferences and journals are soliciting and publishing articles in this area. Recent papers by Ge and Helfert and Madnick et al. [28,29] describe the emerging framework of data/information quality research. In the practitioner community, English [4] identified typical data quality training topics for various organizational roles. Consequently there is a gradual movement of both academic and professional programs to develop training geared specifically towards providing individuals engaged in data quality activities with the knowledge areas and skills that they need.

Based on the findings of the 2009 IAIDQ salary and job satisfaction report, most data quality professionals start with obtaining a Bachelor's or Master's degree. While the disciplines vary, business (29.8 %) and IT-/IS-related areas (21.8 %) are the most common academic preparation [27] (Fig. 1).

How have today's data quality professionals developed and learned the skills they need for the job? Typically, once an aspiring data quality professional has gained some mastery of a particular subject area (e.g., business, technology, science, or mathematics) along with the IT/IS techniques for managing the data encompassed by that subject, the next step is to augment that knowledge with a combination of self-study, specialized courses, and/or professional training in data quality concepts and best practices. A recent survey indicates that the overwhelming majority of data quality professionals did not receive any formal training in data quality management. Sixty-three percent (63 %) of respondents indicated that they were self-taught, which often was combined with on-the-job training (47 %). Twenty-four percent (24 %) received professional training related to data quality, and 20 % reported having a university training related to data quality [30]. The survey by Sadiq et al. [30] also noted the significant negative impact that the lack of standardized or best-practice data quality education has on the quality, consistency, sustainability, and eventual success of data quality management as currently practiced.

Two recent developments have become significant milestones in the efforts to define standards for data quality knowledge and skills and education:

- The introduction of data quality courses at colleges and universities worldwide, culminating in the establishment of the Information Quality Program at the University of Arkansas at Little Rock
- The introduction of the Information Quality Certified Professional (IQCP) credential, by IAIDQ

We briefly discuss the IQCP credential here and follow with an overview of data quality academic programs and others sources of data quality education and training.
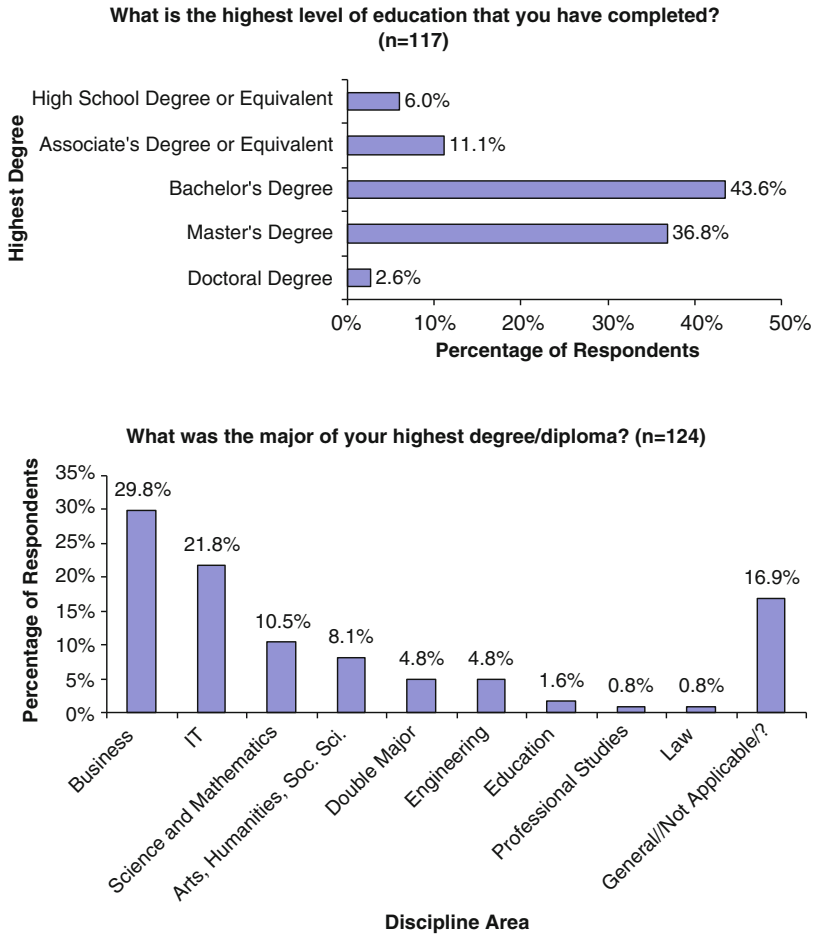
What is the highest level of education that you have completed? (n=117)

High School Degree or Equivalent — 6.0%
Associate's Degree or Equivalent — 11.1%
Bachelor's Degree — 43.6%
Master's Degree — 36.8%
Doctoral Degree — 2.6%

Highest Degree

Percentage of Respondents

What was the major of your highest degree/diploma? (n=124)

Business — 29.8%
IT — 21.8%
Science and Mathematics — 10.5%
Arts, Humanities, Soc. Sci. — 8.1%
Double Major — 4.8%
Engineering — 4.8%
Education — 1.6%
Professional Studies — 0.8%
Law — 0.8%
General/Not Applicable/? — 16.9%

Percentage of Respondents

Discipline Area

Fig. 1 2009 IAIDQ salary and job satisfaction report [27]

3.1 International Professional Certification: The IAIDQ Information Quality Certified Professional (IQCP) Credential

Chartered in 2004, IAIDQ is the only international professional organization devoted entirely to information and data quality. It has members in more than 30 countries on five continents. In February 2011, IAIDQ introduced the Information Quality Certified Professional (IQCP) credential. It is rapidly becoming the global standard of competence for data quality practitioners. The IQCP certification has

three components (IAIDQ; http://iaidq.org):

- Work experience and education requirements
- Taking and passing a comprehensive 3-h exam consisting of 150 multiple choice questions with four possible answers each. Following Bloom's Taxonomy, the questions assess three cognitive domains: Recall/Understanding, Application, and Analysis
- Signing the IAIDQ Code of Ethics and Professional Conduct

The IQCP credential must be renewed every 3 years. There are two ways to recertify:

- Submit a recertification journal that documents a minimum level of ongoing professional development
- Or take the exam again

The certification is based on the findings of a job analysis/role delineation study sponsored by IAIDQ between October 2008 and March 2009 and conducted with CASTLE Worldwide, Inc. (CASTLE), a firm that specializes in the development of professional certifications [23, 24]. The purpose of the study was to build IAIDQ's Information Quality Certified Professional credential on a solid foundation validated by practitioners and consistent with best practices. The process followed by IAIDQ and CASTLE complies with widely accepted standards and regulations, such as the ISO/IEC 17024 for Personnel Certification Bodies. The exam content is independent of any specific methodology, vendor, or tool.

The panel of experts assembled for the job analysis/role delineation study and developed a consensus definition of the job of the IQCP which consists of a framework containing six (6) performance domains, twenty-nine (29) tasks, and 207 distinct knowledge and skills. After it was validated by a large international group of information/data quality practitioners, the framework was used to develop the specifications for the IQCP Exam. The six performance domains are:

- Information quality strategy and governance
- Information quality environment and culture
- Information quality value and business impact
- Information architecture quality
- Information quality measurement and improvement
- Sustaining information quality

The 207 distinct knowledge areas and skills are further classified into the following five groups:

- IQ/DQ core
- Quality foundation
- Leading the IQ/DQ effort
- Information management
- People and interpersonal effectiveness

Beyond this primary purpose as the blueprint for the certification exam, the IQCP Framework is also expected to:

- Drive an increase in the quality and consistency of the information/data quality training available in the market place
- Provide a benchmark against which organizations can assess their information/data quality practices

As another indicator of the need for formal data quality training and best practice data quality education standards and benchmarks, Yonke et al. [23] found 43 % of respondents that they surveyed indicated that the most important benefit they expected from a data quality certification was "increased knowledge and mastery of the information/data quality discipline" [23].

3.2 Data Quality Education and Training Sources

In addition to opportunities for self-study through published sources (e.g., social media, books, and journal articles), several formal educational opportunities in data quality are available. These offerings can be categorized as follows:

- The MIT Information Quality Program
- The UALR Information Quality Program
- Data Quality Education at Other Colleges and Universities
- Information Quality Communities of Practice and Certification
- Industry Conferences and Practitioner-Provided Training

3.2.1   The MIT Information Quality Program

Much of the credit for introducing academic rigor into the field of data and information quality belongs to the Massachusetts Institute of Technology (MIT) information quality program. Responding to industry needs for high-quality data and inspired by the success of the Total Quality Management movement in manufacturing, Dr. Stuart Madnick in the MIT Sloan School of Management led a partnership of organizations to create a research program in the early 1990s called Total Data Quality Management (TDQM). While the short-term focus of TDQM was to create a center of excellence among practitioners of data quality techniques, its greatest impact has been to build an academic research community to investigate the theory of data and information quality as well as documenting its best practices.

One of the most successful products of the TDQM program is the MIT information quality program (MITIQ) led by Dr. Richard Wang and housed in the MIT Center for Technology, Policy, and Industrial Development (CTPID). The MITIQ program has been the leader in promoting and disseminating research in

information quality through its sponsorship of the International Conference on Information Quality (ICIQ). The ICIQ has been held annually since 1996 and has created a worldwide community of academicians and practitioners who regularly present at the conference and publish their peer-reviewed papers in its proceedings.

The members of the MITIQ community have been active in developing many of the fundamental principles and tenets of information quality. Examples of these include the studies on impact of poor information quality [3, 31], the dimensions of data quality [11], information as product [10], and information quality assessment and improvement methodologies [32].

Many key initiatives have also had their roots in the MITIQ community as well. One of the most important was the 2009 launch of the Association for Computing Machinery (ACM) Journal of Data and Information Quality (JDIQ) with Dr. Madnick and Dr. Yang Lee (Northeastern University), as its Founding Editors-in-Chief. Another was the organization of the annual Information Quality Industry Symposium (IQIS) established in July 2007. Originally begun to complement the research focus of the ICIQ, the IQIS conference was intended to promote the sharing of best practices and technology among IQ practitioners, IQ tool vendors, and professional organizations promoting IQ. The most recent initiative has been the MIT Chief Data Officer (CDO) Forum started in 2011. Meeting in conjunction with the July IQIS, the new event has been reformulated as the Chief Data Officer and Information Quality (CDOIQ) Symposium to reflect its emphasis on establishing information quality as an organizational function led at the enterprise level by the Chief Data Officer (CDO Forum, 2012).[1]

3.2.2   The UALR Information Quality Graduate Program[2]

Despite the rapid growth in information quality practices that arose out of the data warehousing during 1990s, there was not a corresponding growth in academic programs. In 2000, Craig Fisher introduced an undergraduate course titled Data Quality and Information Systems at Marist College [33]. This subsequently led to the publication of the first college-level textbook on IQ, Introduction to Information Quality [34].

In 2005, Dr. Richard Wang working in collaboration with Charles Morgan, the company leader of Acxiom Corporation headquartered in Little Rock, Arkansas, and with Dr. Mary Good, Dean of the Donaghey College of Engineering and Information Technology (EIT) at the University of Arkansas at Little Rock (UALR), conceived a plan to create the first graduate degree program in information quality [35]. A Master of Science in Information Quality (MSIQ) was the first program developed

---

[1]CDO Forum (2012) The Second MIT Chief Data Officer Forum (http://mitiq.mit.edu/CDO/2012/).

[2]The University of Arkansas at Little Rock Information Quality Graduate Program, UALR IQ (http://ualr.edu/informationquality/).

Table 3 Information quality course offerings at the University of Arkansas at Little Rock

PhD in Integrated Computing with emphasis in Information Quality	INFQ 7303 Principles of Information Quality
	INFQ 7342 Information Quality Tools and Industry Landscape
	INFQ 7367 Information Quality Policy and Strategy
	7318 Total Quality Management and Statistical Process Control
	INFQ 7337 Project and Change Management
Master's degree in Information Quality	INFQ 7322 Information Quality Theory
	INFQ 7348 Entity Resolution and Information Quality
	IFSC 7360 Data Protection and Privacy
	IFSC 5330 Database Security
Graduate Certificate in Information Quality	IFSC 5345 Information Visualization
	IFSC 7320 Database Systems and Information Architecture
	IFSC 7310 Information Systems Analysis

and approved, and the first cohort of 24 students was enrolled in the fall of 2006. The UALR Information Quality Graduate Program (UALR IQ, 2012) is housed in the Department of Information Science in the Donaghey College of Engineering and Information Technology and has expanded to include a Graduate Certificate in Information Quality and an information quality emphasis track in the Integrated Computing PhD program (Table 3).

The curriculum development for the program was a joint effort among several collaborators [36]. Much of the course content has to be developed from the ground up including the Principles of Information Quality course [37], the Information Quality Tools course [38], and Entity Resolution and Information Quality [36].

As the only university program granting graduate degrees in information quality in the United States, it was decided from its inception that the program should have a distance education component. Beginning in 2007, the program was made available online by live webcasting of the classes required for the MSIQ program. Unlike traditional online classes where students work at their own pace, the UALR IQ online program is synchronized with the on-campus course offerings. Each course has an on-campus classroom and normal meeting schedule. The classrooms are specially equipped so that as lectures are being delivered to the local students, they are also webcast to remote students in realtime. The webcast system allows remote students to see whatever is displayed on the instructor's desktop and to have both chat and audio interactions with the instructor and other members of the class. The webcasts are also recorded for later viewing. Remote students are required to take their major examinations through a test proctoring service [39].

Students across the USA in states such as California, Texas, New York, Georgia, and North Carolina, and in several foreign countries such as Brazil and South Africa have successfully completed their degrees online. As of May 2012, the UALR IQ

Table 4 Colleges and universities offering IQ/DQ courses

School	Program name	Courses
Marist College, Poughkeepsie, New York, USA	BS w/ Major in Information Technology and Systems MS: MSIS degree	ITS 428 Data Quality in Information Systems MSIS 557 Data Quality in Information Systems
University of Nebraska at Omaha, Omaha, Nebraska, USA	Information Systems Quantitative Analysis	Information Quality and Data Management (ISQA 8206/4200)
Northeastern University, Boston, Massachusetts, USA	MBA MIS/Honors Undergraduates	Information Quality: Technology and Philosophy Information Quality for Global Managers Information Resource Management
The University Of Michigan-Dearborn, Dearborn, Michigan, USA	MSE in Manufacturing Systems Engineering	IMSE 532 - Information for Manufacturing
Dublin City University, Dublin, Ireland	MSc in Business Informatics	Research at the Business Informatics Group
Nyenrode Business University, Breukelen, Netherlands	Strategic Business Planning	Masterclass Data Quality Management
University of South Australia, Adelaide, SA, Australia	Doctor of Information Science	Research Program
University of St. Gallen, St. Gallen, Switzerland	Competence Center Corporate Data Quality	Consortium Research Program
University of Westminster, London, UK	MSc Information Quality	Course modules in Information Quality, Information Security, Data Warehousing and Data Mining, Enterprise Applications, Statistical Modeling, Postgraduate Project

program has graduated 58 students with the MSIQ degree and 7 students with the Integrated Computing PhD degree with an Information Quality emphasis.

3.2.3   Data Quality Education at Other Colleges and Universities

Since 2006, several other colleges and universities have developed graduate-level programs in data quality (Table 4). These include a Master of Science in Information Quality at the University of Westminster in London, UK, and Graduate Studies in Information Quality at the University of South Australia in Adelaide. In addition, although not offering full degree programs, many schools have introduced courses in data quality into existing business or IT degree plans. IAIDQ maintains an updated list on its website (IAIDQ; http://iaidq.org).

3.2.4 Information Quality Communities of Practice

Information Quality Communities of Practice are groups organized on a volunteer basis by data quality professionals to promote awareness of the discipline and to better share data and information quality best practices. The major Information Quality Communities of Practice include the international association IAIDQ and several national organizations such as DGIQ in Germany, ExIQ in France, AECDI in Spain, ArgIQ in Argentina, and QIBRAS in Brazil. These communities of practice are a valuable resource for sharing experiences and disseminating knowledge regarding data quality techniques and best practices.

In November 2010 at the 15th International Conference for Information Quality held at the University of Arkansas at Little Rock, representatives of the major national Information Quality Communities of Practice from around the world agreed to upgrade their current course materials so they could help their members prepare for the Information Quality Certified Professional exam. In addition to generating these education materials, they also plan to develop a special version for universities in order to provide a worldwide education base of high-quality instructional units.

3.2.5 Industry Conferences and Practitioner-Provided Training

This last group is probably responsible for the bulk of data quality training available thus far. Combining half-day or full-day tutorials with conference presentations and case studies, practitioner-led international conferences held around the world have played a key role in developing data quality professionals. We note in particular:

- The Information Quality Conference organized by Larry English and held in the USA between 1999 and 2005.
- The conferences organized by IAIDQ and its partners and held in the USA since 2006 (www.idq-conference.com).
- The Data Management and Information Quality Conference Europe held in London since 1999 and organized by IRM UK and its partners including Larry English and DAMA (www.irmuk.co.uk).
- Data Quality conference in Sydney, Australia, organized by Ark Group Australia since 2006, and under the brand name Data Quality Asia Pacific Congress since 2008 (www.dqasiapacific.com).

Practitioner training focusing on data and/or information quality is also offered from various consultants, professional organizations, software vendors, training institutes, and government agencies. Software vendors often link their courses closely to their own products that they offer for resolving data quality problems, while training institutes use their data quality modules to enhance their education program portfolios. Some examples of organizations offering data quality training are listed in Table 5.

Table 5 Examples of organizations offering data quality training and consulting

Organization	URL
Information Impact International, Inc.	http://www.infoimpact.com/
Knowledge Integrity, Inc.	http://knowledge-integrity.com/
Granite Falls Consulting	http://www.gfalls.com/index.html
Navesink Consulting Group	http://www.dataqualitysolutions.com/
The Data Governance Institute	http://www.datagovernance.com/
Institute for Certification and Training of Information Professionals	http://ictip.org
Castlebridge Associates	http://castlebridge.ie
eLearningCurve	http://ecm.elearningcurve.com
BI Community	http://www.bi-community.org

4 The Future for the Data Quality Profession

The data quality profession continues to grow. Although costs and revenue issues associated with poor-quality data have always been a factor, it appears that compliance and regulatory issues have become a major driver for organizations to invest in treating their information as a strategic asset [23, 40]. This motivation stems from regulation specifically aimed at improving the quality of data as well as consequences of privacy legislation. The Data Protection Act in 1998, the Federal Data Quality Act in 2001, the Sarbanes-Oxley Act in 2002, and Basel II in 2004 all contain language regarding better data management so as to ensure the accuracy of the data being reported to the Federal Government. Privacy legislation such as the National Do Not Call Registry and HIPAA has made it imperative that organizations maintain good-quality data regarding their customers in order to meet their legal obligations. Issues such as these along with the growth of data warehousing, business intelligence, and master data management initiatives have spurred organizations' desire for better data.

Today more organizations than ever are moving towards treating their data as critical assets that must be effectively governed for maximum value. To do this, organizations need everyone who works with data whether they are a data creator, data processor, or data consumer to appreciate the basic tenets of the six domains that IAIDQ has defined as key data quality knowledge areas. In addition, organizations require individuals with more in-depth training and understanding to take on leadership roles in implementing data quality best practices as part of the organization's day-to-day data management and stewardship activities. Universities, communities of practice, and professional trainers/consultants all have a role to play in providing educational opportunities. Web resources in the form of websites, Web training videos, white papers, wikis, and electronic serial collections make it economically feasible to disseminate tools, techniques, and lessons learned on a global scale.

Looking forward, the main challenge for data quality professionals will be overcoming the obstacles that prevent organizations from maximizing the value of their information. A 2012 survey conducted by UALR and IAIDQ revealed that data

quality professionals continue to face numerous challenges in their organizations. These obstacles include:

- Lack of accountability and responsibility for data quality
- Too many information silos
- Lack of awareness or communication of the magnitude of data quality problems
- Lack of common understanding of what data quality means
- Lack of awareness or communication of the opportunities associated with high-quality data.
- Lack of senior leadership in tackling data quality issues.
- Lack of data quality policies, plans, and procedures.
- Perception that data quality is an IT issue only rather than an organization-wide issue, and in some organizations, there may be a reverse perception that data quality is a business issue only and cannot be helped with IT support.
- Lack of data quality goal setting and measurement.
- Lack of data quality skills and expertise.
- Lack of data quality tools and automation.
- Lack of resources including limited staff to manage data issues and promote data quality, cost to build a good data quality program, and 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.

If data and information quality is to make progress as a discipline, these obstacles must be alleviated. Many of these problems have at their root a lack of awareness by organizational leaders of either the negative impacts of poor data quality or a lack of awareness that there exist effective strategies and practices for eliminating poor data quality [12, 16, 41]. Communities of practice, academic centers, and leading practitioners must continue their efforts to spread the message while they build upon the data quality knowledge base. By combining traditional learning avenues such as university and college courses and degrees, conferences and workshops with the power of online courses and social media such as webinars and wikis, individuals in the data quality field can help provide a rich infrastructure for those trying to establish and grow a culture in their organization that supports and promotes data quality improvement.

References

1. Inmon WH (1992) Building the data warehouse. Wiley, New York
2. Kimball R, Ross M, Thornthwaite W, Mundy J, Becker B (1998) The data warehouse lifecycle toolkit. Wiley, New York

3. Redman TC (1998) The impact of poor data quality on the typical enterprise. Commun ACM 41(2):79–82
4. English LP (1999) Improving data warehouse and business information quality: methods for reducing costs and increasing profits. Wiley, New York
5. Brackett M (1996) The data warehouse challenge: taming the chaos. Wiley Computer Publishing, New York
6. Lindsey E (2008) Three-dimensional analysis: data profiling techniques examining data content, structure, and quality. Data Profiling LLC, USA
7. Olson J (2003) Data quality: the accuracy dimension. Morgan Kaufmann, Burlington
8. McGilvray D (2008) Executing data quality projects: ten steps to quality data and trusted information. Morgan Kaufmann, Burlington
9. Wang RY, Kon H-B (1993) Toward Total Data Quality Management (TDQM). In: Wang RY (ed) Information technology in action: trends and perspectives, 1st edn, Prentice Hall, Englewood Cliffs
10. Wang RY, Lee YW, Pipino LL, Strong DM (1998) Manage your information as a product. Sloan Manag Rev 1998(Summer):95–105
11. Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to consumers. J Manag Inf Syst 12(4):5–34
12. English L (2009) Information quality applied: best practices for improving business information, processes, and systems. Wiley, New York
13. English LP (2008) Information quality management: job position roles. IDQ Newsletter 4(2), International Association for Information and Data Quality, IAIDQ. http://iaidq.org/publications/english-2008-04.shtml
14. Redman TC (2007) The body has a heart and soul: roles and responsibilities of the Chief Data Officer. IDQ Newsletter 3(1), International Association for Information and Data Quality, IAIDQ. http://iaidq.org/publications/redman-2007-01.shtml
15. Redman TC (2005) A comprehensive approach to data quality governance. In: A Navesink Consulting Group White Paper. International Association for Information and Data Quality, IAIDQ. http://iaidq.org/publications/redman-2005-06.shtml
16. Redman TC (2008) Data driven: profiting from your most important business asset. Harvard Business Press, Boston
17. Redman TC (2001) Data quality: the field guide. Digital Press, Surry Hills
18. Redman TC (1997) Data quality for the information age. Artech House, Norwood
19. Huang K-T, Lee YW, Wang RY (1999) Quality information and knowledge. Prentice Hall, Upper Saddle River
20. Deming WE (1986) Out of the crisis. MIT Press, Cambridge
21. Loshin D (2008) Master data management. Morgan Kaufmann, Burlington
22. Baskarada S (2009) Information quality management capability maturity model. Vieweg+Teubner, Germany
23. Yonke CL, Walenta C, Talburt JR (2011) The job of the information/data quality professional. Published by IAIDQ. http://iaidq.org/publications/yonke-2011-02.shtml
24. Yonke CL, Walenta C, Talburt JR (2011) The skills of the information/data quality professional. International Association for Information and Data Quality, IAIDQ. http://iaidq.org/publications/yonke-2011-11.shtml
25. Chung WY, Fisher CW, Wang R (2002) What skills matter in data quality? In: Proceedings of the 7th international conference on information quality, MIT, Cambridge, 8–10 November 2002. MIT IQ Publishing, Cambridge
26. Pierce EM (2003) Pursuing a career in information quality: the job of the data quality analyst. In: Proceedings of the 8th international conference on information quality, MIT, Cambridge, 7–9 November 2003. MIT IQ Publishing, Cambridge
27. Pierce EM, Yonke CL, Lintag A (2009) 2009 Information/Data Quality Salary and Job Satisfaction Report: Understanding the Compensation and Outlook of Information/Data Quality Professionals. International Association for Information and Data Quality, IAIDQ. http://iaidq.org/publications/pierce-2009-07.shtml

28. Ge H, Helfert H (2006) A review of information quality research. In: Proceedings of the 11th international conference on information quality, MIT, Cambridge, 10–12 November 2006. MIT IQ Publishing, Cambridge
29. Madnick SE, Wang RY, Lee Y-L, Zhu H (2009) Overview and framework for data and information quality research. J Data Inf Qual 1(1):2.1–2.22
30. Sadiq S, Indulska M, Jayawardene V (2011) Research and industry synergies in data quality management. In: Proceedings of the 16th international conference on information quality, MIT, Cambridge, 18–20 November 2011. MIT IQ Publishing, Cambridge
31. Fisher CW, Kingma BR (2001) Criticality of data quality as exemplified in two disasters. Inf Manag 39:109–116
32. Pipino L, Lee Y, Wang R (2002) Data quality assessment. Commun ACM 45(2):211–218
33. Fisher CW (2001) A college course: data quality in information systems. In: Proceedings of the 6th international conference on information quality, MIT, Cambridge, 2–4 November 2001. MIT IQ Publishing, Cambridge
34. Fisher C, Lauria E, Chengalur-Smith S, Wang R (2006) Introduction to information quality. M.I.T. Information Quality Program, Cambridge
35. Talburt J (2011) Entity resolution and information quality. Morgan Kaufmann, Burlington
36. Lee Y, Pierce E, Talburt J, Wang R, Zhu H (2007) A curriculum for Master's of Science in Information Quality. J Inf Syst Ed 18(2):233–242
37. Zhou Y, Talburt JR (2011) A gateway course for an information quality graduate program. In: Proceedings of the international symposium on information systems and management in Asia Pacific, ISMAP-2011, Xi'an, China, 9–11 December 2011
38. Zhou Y, Talburt JR (2011) Developing a graduate-level course on information quality tools: closing the gap between theory and practice. In: Proceedings of the association for information systems sigIQ Workshop, Shanghai, China, 3 December 2011
39. Zhou Y, Talburt JR (2011) Campus-Centric Distance Education using Wimba Live Classroom. In: Proceedings of the 2011 IEEE international conference on IT in medicine and education, Guangzhou, China, 9–11 December 2011
40. Pierce EM, Yonke CL, Malik P, Kagathur Nargaraj C (2012) The State of Information and Data Quality: Understanding How Organizations Manage the Quality of Their Information and Data Assets. International Association for Information and Data Quality, IAIDQ. http://iaidq.org/publications/pierce-2012-11.shtml
41. Lee Y, Pipino L, Funk J, Wang R (2006) Journey to data quality. MIT Publishing, Cambridge