



DTCC Data Quality Survey

Industry Report

November 2013

DTCC

Contents



1. Introduction	3
2. Approach and participants	4
3. Summary findings	5
4. Findings by topic	6
a. Data Quality Strategy	6
b. Data Quality Measurement/Analysis	8
c. Data Profiling	10
d. Data Quality Assessment	11
e. Data Cleansing	12
f. Data Quality Technology	13

Introduction



Context

- DTCC commissioned Element22 to conduct a limited and informal data quality survey of voluntary participants in order to gain insights into emerging data quality industry practices which DTCC will consider for implementation and share with survey participants.
- Element22 designed a market survey based on Enterprise Data Management Council's Data Management Maturity Model.

Objectives

- Provide indicative view of industry practices
- Identify emerging standards if they exist
- Share findings with participants and stakeholders
- Determine practices that DTCC should implement

Approach and participants



Participant profile

- 35 firms were invited (on 08/23/13) to participate in the short time frame (09/17/2013 – 11/08/13)
- 20 firms participated and provided responses, representing the following businesses:
 - (3) Asset Management
 - (3) Asset Servicing
 - (5) Data Vendors and Data Service Vendors
 - (3) Infrastructure & Other
 - (2) Investment Bank/Broker
 - (4) Universal Banks

Approach

- Structured interviews
 - Discussion based
 - Open ended
- Analysis
 - Synthesis of responses
 - Tabulation where appropriate
- Findings
 - Examples of common practices
 - Indication of the range of practices
 - Indicative profile of responses based on limited sample (not statistically significant)

Summary findings

Overview

- All of the Survey participants perform Data Quality work and are addressing problems with data
- Financial services firms are in the early stages of implementing Enterprise Data Management Programs
- Many of the respondents still do not have formal Data Quality programs in place even though they do perform Data Quality work

Strategy

- More than half of the respondents do not have DQ Strategy defined and are dealing with data quality on a reactive and ad-hoc basis without a clear understanding of the data quality impact to the business of the firm
- Those with a DQ Strategy have Data Governance and Stewardship established in line with their DQ Strategy

Approaches

- Data Quality work in firms without DQ Strategy is an accumulation of the processes that have been established through time as a reaction to different business problems caused by data quality issues. Most of these firms are planning to start building a formalized DQ programs with established objectives
- Critical Data Elements are key to DQ Programs and for them data quality expectations are first captured and defined. These expectations are then used to establish DQ metrics, measurements and rules
- Vast majority of respondents have DQ measurement processes in place but it is mostly hardcoded in proprietary systems or is manually executed

Use of technology

- There is no consistency in which tools are used by respondents and none of the tools dominate any aspect of Data Quality work
- Most of the respondents use home grown systems for Data Quality or are using DBMS inquiry

Variation/similarity across organization types

- Largest Financial Services Firms have implemented Federated DQ programs with minimal standardization at enterprise level
- Almost all of the respondents have established metrics and use some of EDM Council's Core Dimensions of Data Quality

Opportunities

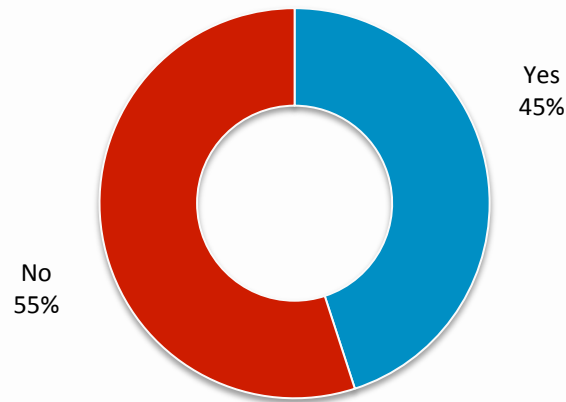
- One of the participants has a defined Data Quality Index to measure overall Data Quality within firm. If a Data Quality Index could be defined in a such a generic way that it could be used across the industry it would be beneficial to all firms within Financial Services Industry

ENTERPRISE DATA MANAGEMENT EDM COUNCIL		Data Quality Dimensions
Accuracy	Affinity of content with original intent; veracity as compared to an authoritative (original or 3rd party) source; measurement precision	
Completeness	Availability of required data attributes	
Conformity	Alignment of content with the required standards	
Consistency	Compliance with required formats or definitions	
Coverage	Availability of required data records	
Duplication	Redundancy of records and or attributes	
Timeliness	Currency of content representation; how well the data represents current market conditions; the data is available and can be used when needed	

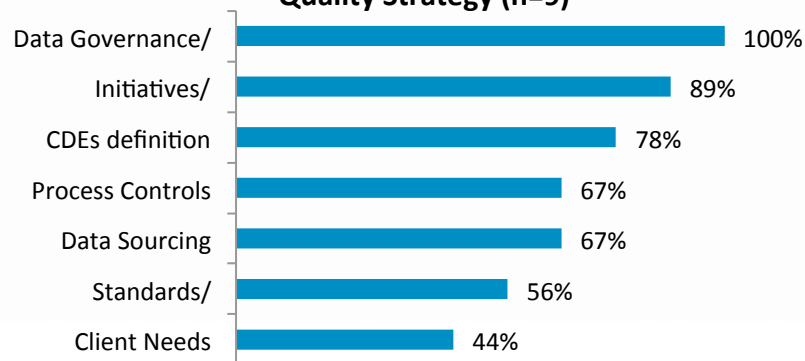
Data Quality Strategy

Do you have a Data Quality strategy defined and what does it cover?

Responses Indicating The Presence Of A Data Quality Strategy (n=20)



The Most Common Elements Covered By The Data Quality Strategy (n=9)



Notes: For all charts, “n” represents the number of respondents for the sample.

Insights

- More than half of the respondents do not have Data Quality Strategy defined and are dealing with data quality on a reactive & ad-hoc basis
- Data Quality work in these firms is an accumulation of the processes that have been established through time as a reaction to different business problems caused by data quality issues. Many of them focus Data Quality work on Critical Data Elements

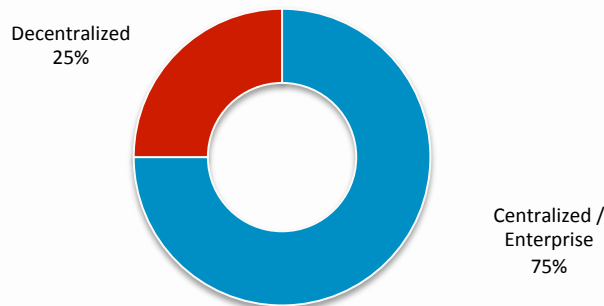
Selected Practices

- We do not have Data Quality Strategy defined. Today within existing projects data quality requirements are captured and then checks and measurements are implemented using common data quality tool.
- Data Quality Strategy aligns critical programs such as BASEL Risk Data Aggregation and other critical business efforts (e.g. creation of single transactional platform).
- Core data is key to our Data Quality Strategy and is selected based on the criteria approved by Data Governance Council. We also have promotion process in place for data to become core. Owners of critical processes or business lines own core data. They have to define core data and data quality rules for the firm.
- Data Quality Strategy categorizes data into 3 tiers based on business importance: 1) Data that must be right, 2) Data that should be right, 3) Data that we would like to be right. Business is encouraged to identify first tier data and this data is documented, metered and reported on. Data Quality Strategy defines the management of data quality, metering, remediation and escalation based on tiers.

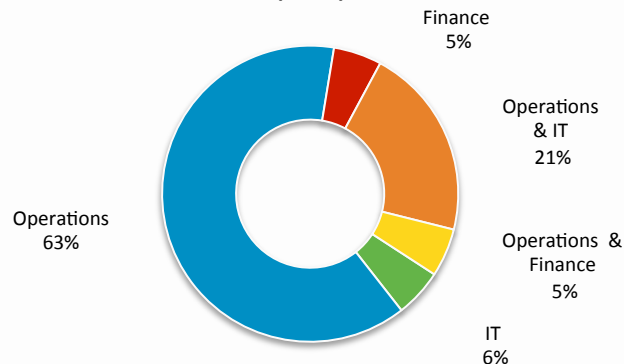
Data Quality Strategy

Location and Organization of Data Quality Program

Responses Indicating A Centralized vs. Decentralized Approach To Data Quality Program (n=20)



Responses Indicating Where The Data Quality Program Resides Within The Organization (n=19)



Insights

- Data Management Program and Data Quality is almost always within Operations (almost 90% if Ops&IT and Ops&Finance are included)
- Majority of the firms have centralized enterprise-wide data quality program and the ones that have them within LOBs are the largest financial services firms that had to create a federated model
- Teams that perform Data Quality work are organized by business lines that they support in over half of the respondents

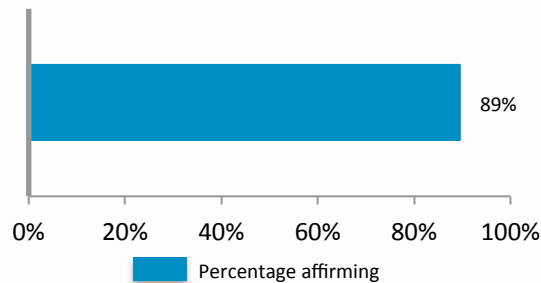
Selected Practices

- Data quality is part of Enterprise Data Management that is within Operations and Technology. Data quality checks are performed on core data across the enterprise and there is a team that is responsible for defining and executing data quality business rules against core data. Separate team is responsible for advancing the data quality strategy and programs, and is focusing on things like data grading and building out future capabilities for data quality platform.
- Data Quality standards are being established, monitored and coordinated at the enterprise level but data quality implementation is done at Business Unit level. In this way, the enterprise group monitors what the different business unit groups are doing for data quality and that they are meeting the needs of the business.
- There is cross-organizational governance around DQ issues to provide visibility to data issues and help prioritize issues. Within a business, issues are raised and potentially remediated. Defect management has owner and resolver. Process has only one owner and the resolver should be at the source, with best case being the root cause resolver. Critical Data Elements are tied to business outcome or business metric and same metric is used as the selection criteria for CDEs.

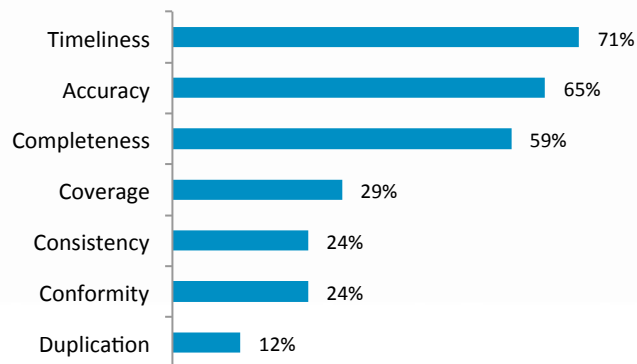
Data Quality Measurement/Analysis

What metrics have you established in the Data Quality program?

**Respondents With Established Data Quality Metrics,
Proportion Of Responses Indicating Use Of Some Or All Of
The EDMC Data Quality Dimensions
(n=19)**



**EDMC Data Quality Metrics Indicated Among Those Using
Some Or All Of The EDMC Data Quality Dimensions
(n=17)**



Insights

- Almost all of the respondents have established DQ metrics and use some of EDM Council's Core Dimensions of Data Quality
- Timeliness, accuracy and completeness are viewed as the most important dimensions of data quality
- Most respondents are tracking quantitative measures such as # of incidents, # of exceptions and # of process failures

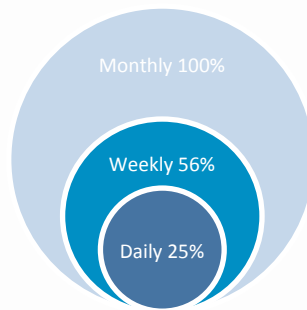
Selected Practices

- We have defined 7 EDM Council's Data Quality dimensions consistently across the firm and Data Quality checks are categorized based on these dimensions. We have wrapped Data Quality program around core data. Because of this, data quality metrics and data quality rules are related to the core data. Where we still need work is on measuring how improvement of data quality increases the business value.
- Main metric we use is the business outcome. It is the only way the business people are able to comprehend the value and business proposition on anything we do within data management including improvement in data quality. Business outcome can also be tied to standards (e.g. 360 view of client; consistent reporting for products across geographical boundaries).
- We measure number of incidents that occur, breach of SLAs, vendor performance metrics, total volumes, number of updates made at product level, pricing changes, number of inquiries we get, number of exceptions performed, and brake them out across locations. We would like to have more enhanced dashboards.

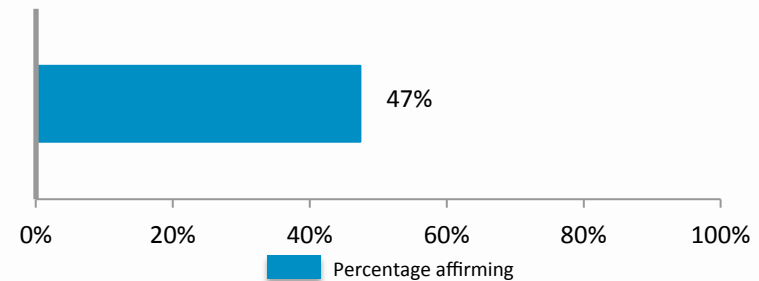
Data Quality Measurement/Analysis

How do you manage, measure and report on data quality (i.e. number of breaks, client satisfaction, cost and ROI, reconciliation required to ensure that data is fit-for-purpose, response time)?

Indicated Frequency Of Data Quality Metrics Reporting (n=13)



Responses Indicating The Use Of Dashboards / Heatmaps For Data Quality Reporting (n=19)



Insights

- All respondents have some kind of Data Quality reporting but tools are not standardized
- Less than half of respondents implemented Dashboards
- Daily reporting is used for operations and 25% have it
- Weekly reporting for tactical management and 56% have it
- Monthly or quarterly reporting for executives and Data Governance Councils and all participants have it

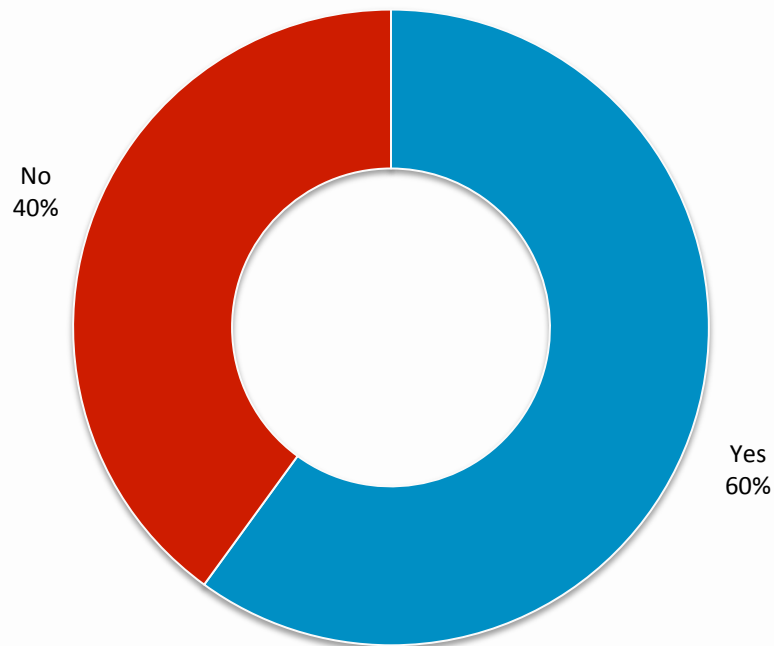
Selected Practices

- User tools allow downloading of the data for interrogation. Some tools have reporting capability and some require IT help. We report daily metrics on key indicators that involve changes e.g. pricing variances; aged tickets on user questions; number of LEI changes or new setups). Weekly reports are more at aggregated level. Monthly reports have additional profiling details.
- We set the time when data is measured every month and also when the measurements and analysis is reported. During the process root cause analysis is done, feedback provided, queries exchanged between independent function that do measurements and the data owners. Every month teams identify problems based on accumulated data and actions are taken (tactical or strategic based on situation). All information is stored in a segmented and defined structure in a repository so it can be retrieved for reporting purposes.
- We just finished a process where we did detailed interview with cross section of critical applications for a cross section of business subject areas by product and asked what they thought about data quality and classify it by 7 Data Quality Dimensions. It is still a collection of subjective views but we came up with a single number that represents Data Quality Index. We have amalgamated all the Data Quality checks and classified them by product. We looked at audit program to see if it could generate a single number for us but found it too difficult. We plan to conduct this process annually.

Data Profiling

Is there a common framework for profiling data within the organization?

Responses Indicating A Common Framework For Profiling Data Is Used Within The Organization (n=20)



Insights

- Large (40%) portion of participants does not have a common profiling framework
- Very few participants have formal profiling frameworks in place that are enforced
- Profiling is mostly done on a reactive basis, within projects or initiatives, for vendor feeds, and for all data domains
- SQL queries are dominantly used as the profiling tool

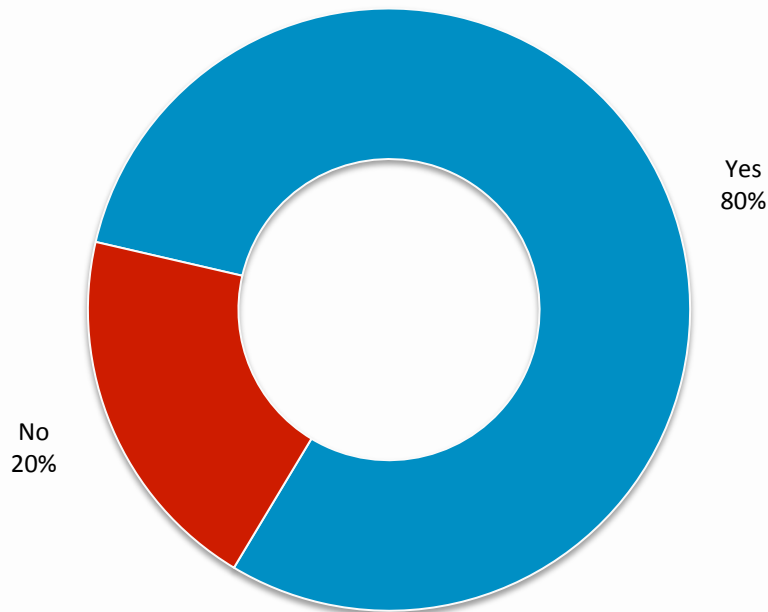
Selected Practices

- We have different LOBs that manage different data sets and require different profiling frameworks. Also, it is technically difficult for us to have a common framework across different data stores.
- We use a common profiling tool that performs correlation analysis to help identify content of and any issues around Critical Data Elements and then focus our monitoring efforts on those CDEs.
- We use a common methodology for profiling new vendor sources to establish data reconciliation processes.
- Profiling is done for new projects but is not yet imbedded in the Data Management day-to-day process. Data Management is a key contributor to the data requirements for all new projects and as such is defining profiling requirements.
- We do profiling during the investigation process. For example if we find an excessive amount of exceptions then in the process to of analysis we may perform profiling. But it is not done in the formal way.

Data Quality Assessment

Are business rules, measurements, and processes based on established data quality metrics?

Responses Indicating Business Rules, Measurements And Processes Based On Data Quality Metrics (n=20)



Insights

- Data Quality Rules and processes are by far based on defined metrics
- Rules are still mainly hardcoded into proprietary systems
- Respondents are finding it difficult to standardize business rules although some are exploring the use of the Semantics of Business Vocabulary and Business Rules (SBVR) standard from OMG
- Business rules and measurements are usually first defined for Critical Data Elements

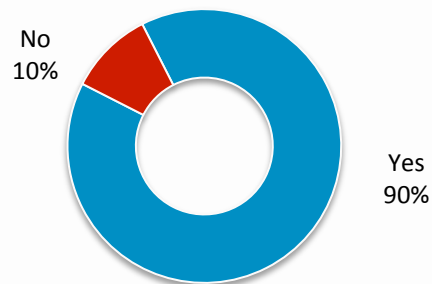
Selected Practices

- We start with data quality dimensions and define metrics on that basis and then business rules, measurements and data quality processes are based on those metrics and KPIs and agreed upon internal standards as well as regulatory requirements.
- We have 1000 validation rules covering internal consistency, mandatory fields, derivation, and symbology cross-reference. Our objective is to clear any exceptions before data goes to consumers. We categorize rules based on the number of records that may be affected by the rule. Data Maturity is constantly evolving and we have periodic review (for process and technology) of BAU to improve maturity. Goal is to achieve zero exceptions.
- We are working with business to understand their needs, capture business requirements, enhance them with Subject Matter Experts, and determine on field basis what rules are based on which categories to properly drive exceptions and manage process.
- Rules are applied based on data element's importance to the business. Timelines dictates how often we run measurements.

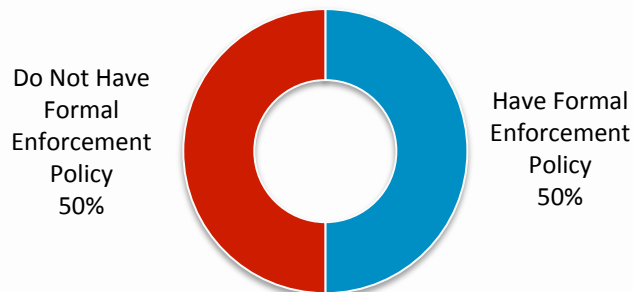
Data Cleansing

Are data cleansing processes part of the overall workflows and are the changes made in systems of record (authoritative systems)? If so how is it enforced?

Responses Indicating Cleansing Processes is done in Systems of Record (n=20)



Responses Indicating Formal vs. Informal Enforcement Policy For Cleansing (n=10)



Insights

- Almost all of the respondents have cleansing processes done in authoritative systems
- That does not apply to overall workflows
- Half of respondents have a formal enforcement policy

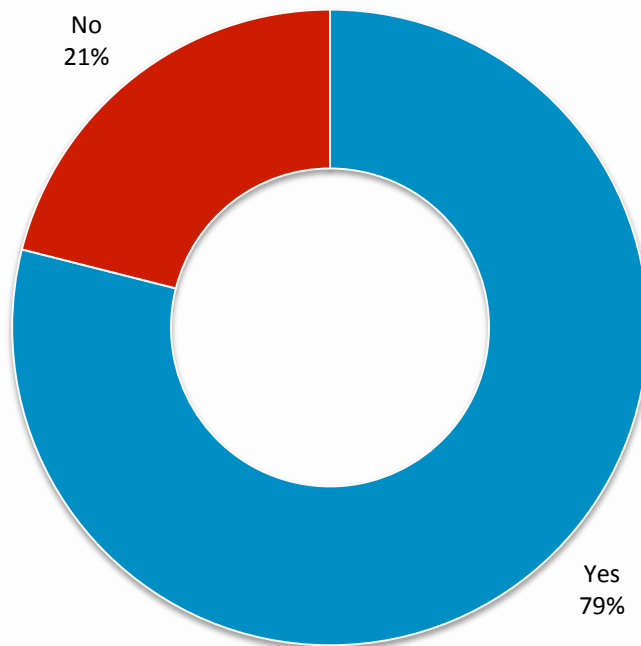
Selected Practices

- We remediate data in the source whenever possible or applicable. Our policy is to cleanse data as close to origination of the data or in system of record as it is possible. Enforcement is done through defined issue remediation process.
- We are trying to push changes as far upstream as it is possible. However in some cases this is limited due to the lack of connectivity to upstream systems. Reports on when reconciliation cannot be done in automated fashion goes to Data Quality or Enterprise Data Management group. Our goal is to master data in one place and distribute throughout the firm. However, we do not have a formal policy and need to get buy in from senior management.
- We always try to solve the problem at source but it depends on root cause. We sometimes filter or overlay data until source fixes are performed.

Data Quality Technology

Which tools if any do you use for Data Quality work and what is your experience?

Responses Indicating Use Of Tools To Define Data Quality Rules (n=19)



Insights

- A wide variety of data quality tools are used across the industry and none of the tools dominate any aspect of Data Quality work

Selected Practices

- Use proprietary tools to build rules and maintain rulebooks, do matching and detect duplicates. We are moving towards SBVR as a standard language to define rules.
- Trillium is used for profiling and Collibra for Critical Data Elements definitions. We use GoldenSource for master data management.
- We don't put rules into the code, instead we use tools that analysts can use to create and run the rules.
- We use Informatica Data Quality (IDQ) to perform data reconciliations between data caches. Rules are embedded into the systems.
- We use MarkitEDM for financial instruments data descriptions, term and conditions; classifications; cross-reference codes; etc. We also use Datanomics and TopBraid for rules and data dictionary.
- Infrastructure is developed in-house with maybe some support from the outside and databases contain rules to check pricing data (tolerance checks) and flag suspect values. Rules are defined by content specialists and are set based on what's expected in the markets.