PwC's data quality capabilities

Providing a holistic DQ framework

March 2021

Table of contents

1

Factors impacting DQ

Why is DQ critical?

Legacy products, systems and fast-changing business environments make it difficult to have a standard DQ process in place, which in turn creates duplication of data or information gaps.

Without clean and standard data, managements across organisations do not get accurate reports and necessary information for tactical and strategic decision making.

Different systems and applications introduce new data formats and related issues.

New technological advancements and availability of data services at lower costs have resulted in the generation of unprecedented volumes of data and lower DQ.

Enterprises create divisions for better management of business processes, which results in duplication of processes and information.

Competitive industries like insurance are driven to provide innovative services to their customers and often produce multiple products and services with overlap and duplication.

Industries like insurance are strongly governed by regulatory authorities, which compel them to maintain different formats and large volumes of data.

Technological advances in systems/applications add to the introduction of complex data and often create in-data integration.

Business impact of poor DQ

Poor-quality data flowing through applications can cause a multitude of negative consequences. A few repercussions of not implementing a holistic DQ framework supported by robust data governance are explained below.

Negative impact on win rate and resultant loss of revenue

The inability to build connections in a customer relationship management (CRM) tool for opportunities within a specific account due to multiple entities that aren't linked.

Inaccurate reporting leading to poor decision making

Inaccurate view of data pipeline, including forecasted margins and demand for staff across an account lead to poor decision making.

Lack of data credibility causing mistrust amongst end users

DQ issues remaining unresolved over time are causing incorrect data elements to be reported through source systems.

Increase in operational costs through poor data-migration practices

The lack of a standardised DQ framework is currently causing ad-hoc migration and increasing operational costs significantly.

Lack of compliance across regulatory responsibilities

The ability to measure and report compliance across a regulatory environment is very important for an organisation to ensure regulatory compliance.

Reactionary DQ remediation

CRM tools are currently creating remediation activities on the fly and using resources to address DQ issues that should otherwise be spent doing other activities without following a standard process.

Violation of business rules

Lack of business rules pertaining to DQ and the ability to monitor compliance against business rules is leading to the generation of redundant data.

Storage costs associated with lack of data-retention policies or monitoring

The lack of data-retention policies is contributing to increased data-storage costs.

2

DQ management

DQ metrics and dimensions

Multiple dimensions of DQ exist and should be considered when defining a client's DQ programme.

- These dimensions are considered industry-leading practices
- The dimensions and terminologies can be customised by clients
- Not all dimensions may be relevant for every field/input
- Key fields/inputs can be measured according to these dimensions

Dimensions	Definition
Existence	• Extent to which the desired data is available on a system – e.g. information related to training hours attended by employees.
Completeness	• Extent to which required data must be populated and the required history exists – e.g. all employees have a location.
Completeness	• Extent to which data adheres to defined business rules, accepted values and accepted formats – e.g. employee gender is F, M or U.
Consistency	• Extent to which identical data must have the same value wherever it is stored or displayed – e.g. aggregated base salary by cost center is consistent between systems.
Accuracy	• Degree to which data should match the agreed source – e.g. initial base salary reflects amount on contract.
Interpretability	• Extent to which data adheres to data management rules and requirements – e.g. all employee-related fields have a definition and data owner.
Uniqueness	• Extent that data should be uniquely stored in one place and not duplicated – e.g. there does not exist multiple records for the same employee.
Availability	• Extent to which current and historic data must be available electronically for analysis – e.g. headcount data can be easily queried to report headcount by region.
Timeliness	• Extent to which the data is refreshed including acceptable systems 'lag' when values change – e.g. base salary updated after promotion within x days.

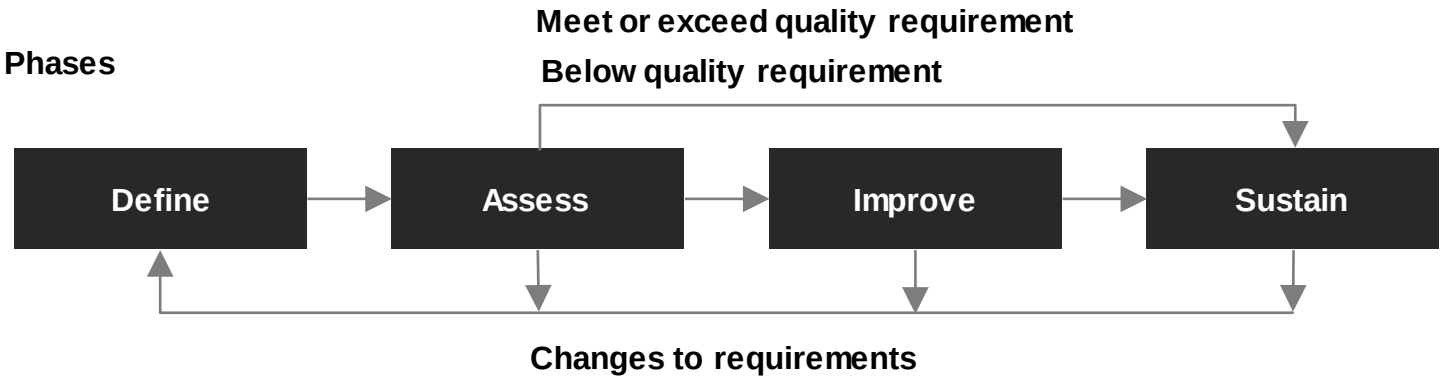
Components of the DQ activity management framework

A DQ framework will measure and monitor the quality of data processed by systems and provide feedback for continuous improvement.

DQ framework:

- **DQF1** Define DQ scope and approach
- **DQF2** Define DQ organisation
- **DQF3** Define DQ process
- **DQF4** Develop DQ technical infrastructure
- **DQF5** Provide training
- **DQF6** Rollout DQ programme

DQF3 DQ process



Key risks addressed

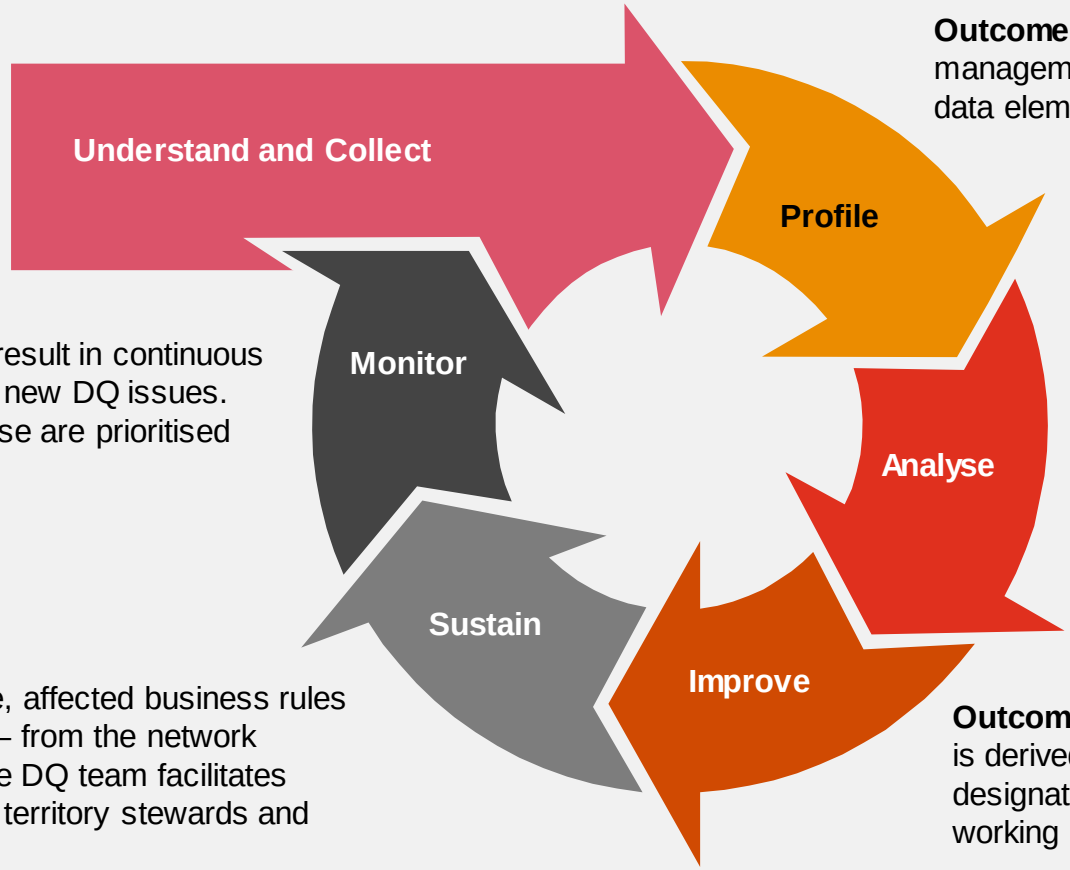
• Quality requirements are unknown or not being addressed	• Businesses perceive DQ levels as higher than they are	• Source data does not meet the increased level of DQ required	• Overall DQ degrades over time • Inadequate control environment in place

DQ objectives

Existence | Completeness | Integrity | Consistency | Accuracy | Interpretability | Uniqueness | Availability | Timeliness

Components of the DQ process framework

Outcome: At the end of this phase, the scopes of DQ checks (business and technical), critical data elements and DQ metric thresholds are defined.



Outcome: At the end of this phase, inconsistencies in data management and DQ policies are identified for defining critical data elements.

Outcome: At the end of this phase, responsibility for identified inconsistencies (DQ tasks) are assigned to key players. Players designated to their relevant roles work with system owners to perform a root-cause analysis and prepare readout decks for their respective inconsistencies.

Outcome: At the end of this phase, a data-cleansing plan is derived using the root-cause analysis results. Players designated to the DQ tasks carry out the improvements by working with system owners and the core DQ team.

Outcome: At the end of this phase, affected business rules are updated in the data dictionary – from the network through the regional level. The core DQ team facilitates socialisation of system changes to territory stewards and data owners.

Outcome: Activities in this phase result in continuous detection, alerting and reporting of new DQ issues. Critical issues identified in this phase are prioritised and assessed.

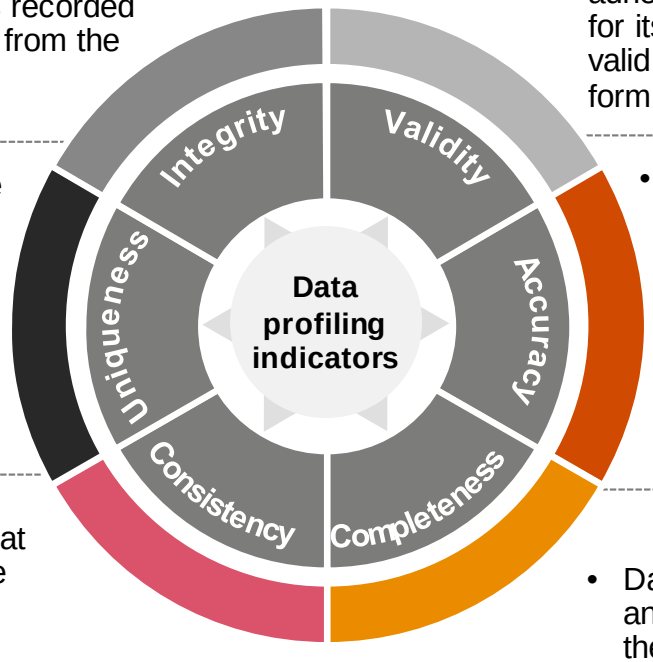
DQ KPI framework

Data gets profiled on the below-mentioned dimensions and weighted average is considered for each dimension for calculating the quality index value.

- Integrity is a measure which ensures that data is recorded exactly as intended from the source location.

- Uniqueness is a measure by which data adheres to unique values without duplication.
- This indicator is the opposite of an assessment of the level of duplication.

- Consistency ensures that any data written into the database must be valid according to all defined rules, including constraints, triggers or any other combination, i.e. the field's definition should be common in all tables.



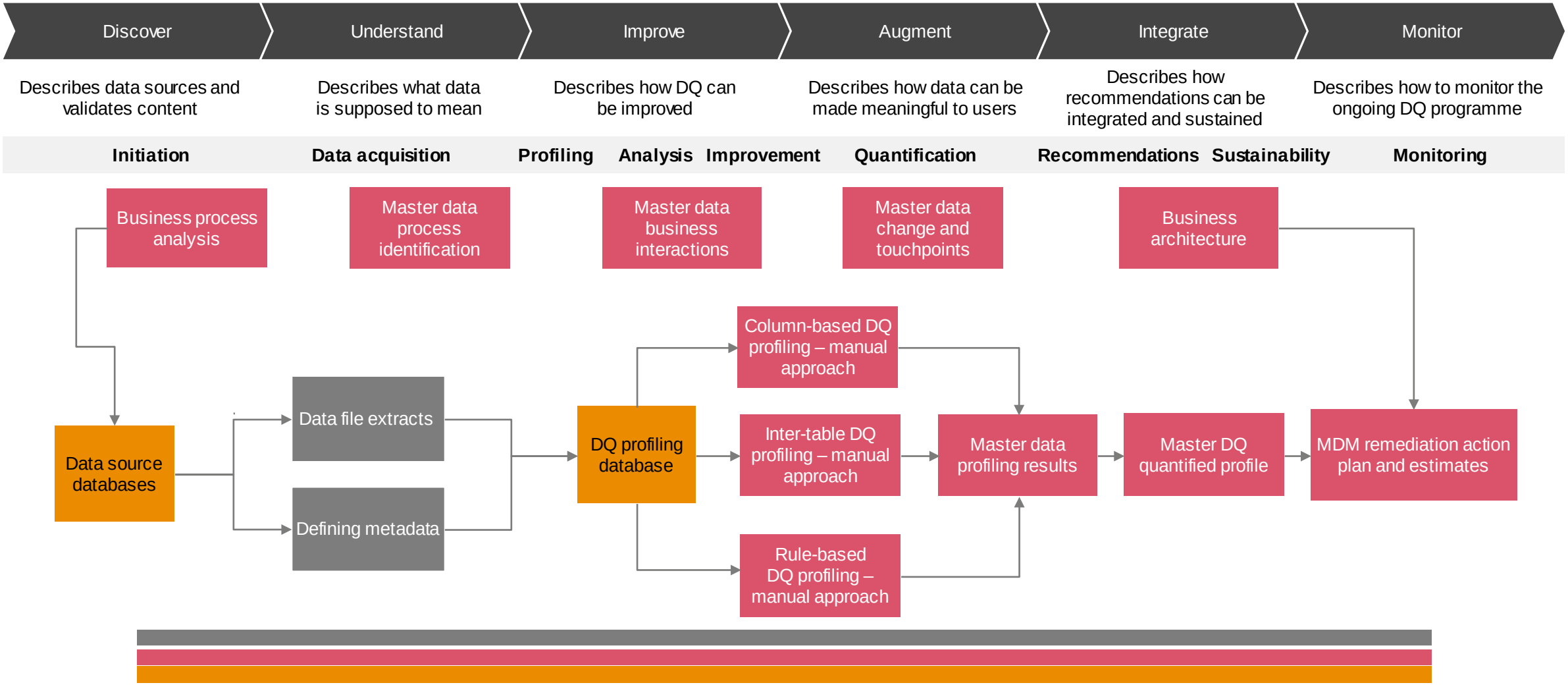
- This is a measure by which data adheres to a common definition for its meaning and use. Data is valid if it does not have any formatting issues.

- This indicator gives a level of assurance that data contains correct values and reflects actuality. Data accuracy is a component of DQ. Data is accurate if it follows defined business rules.

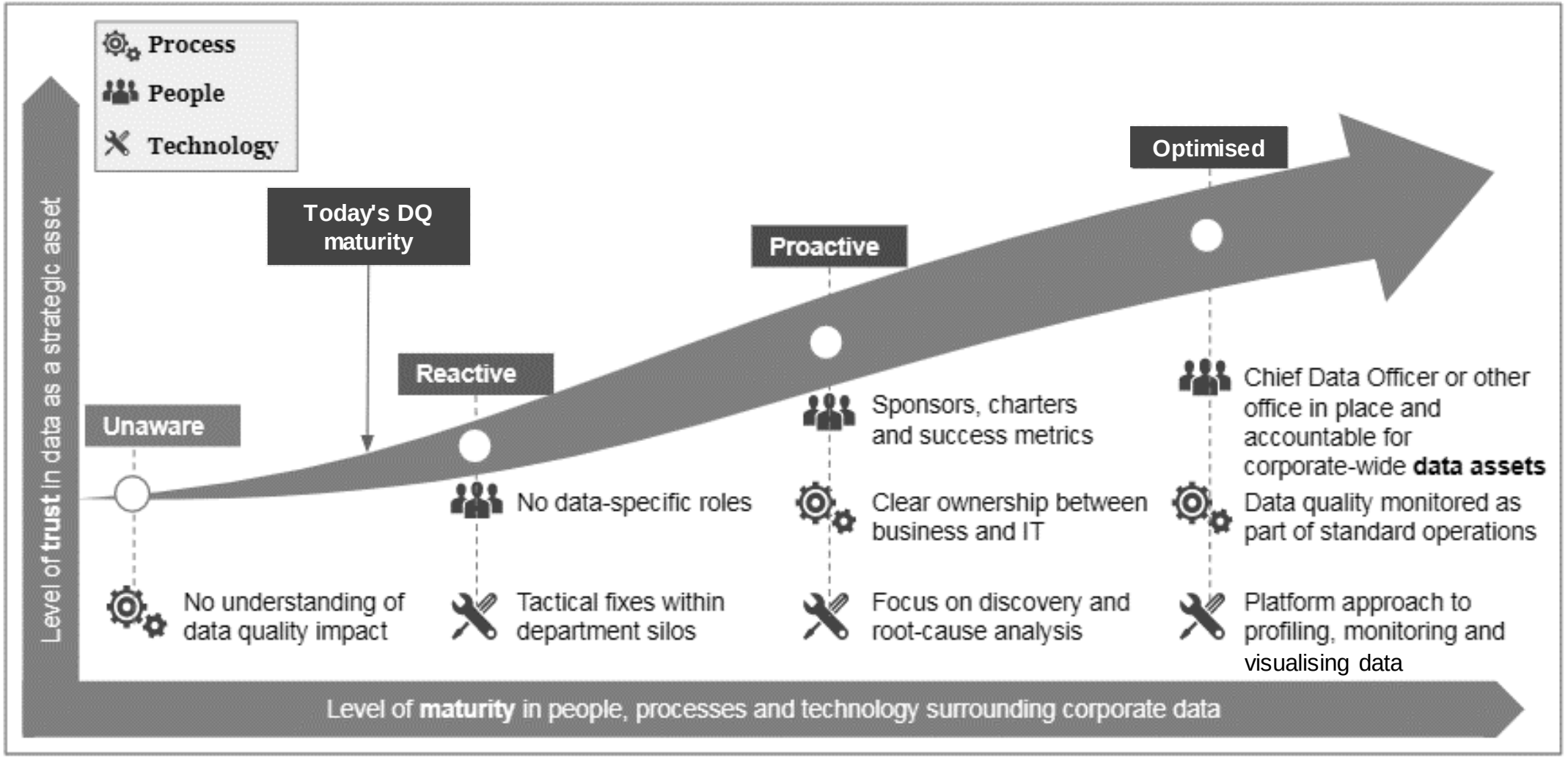
- Data completeness refers to an indication of whether all the data necessary to meet current and future business requirements are available in the data resource.

Typical gaps in an organisation	Root cause
Missing data validation checkpoints (pre, post load) • Business rules, data accuracy, approvals and reconciliation	• Lack of defined ownership • Business rules not documented
Lack of ownership and accountability from business • Data steward and data governance council	• Lack of governance structures • Missing issue-resolution processes
Siloed -function data check from other stakeholders • Integrated data creation rather than correction at the transaction level	• Lack of downstream data-linkage knowledge • Lack of a standardised process
Missing master data management (MDM) tools and upload mechanism • Failures not captured during uploading activities cause data inconsistency and duplication of effort	• Upload programmes not in sync with business rules
DQ metrics not defined • No checkpoints and metrics to capture failure, reason and accountability • Minimum feedback on process failures	• Absence of metrics, benchmarks, workflow, and audit trail in current process

DQ solution approach



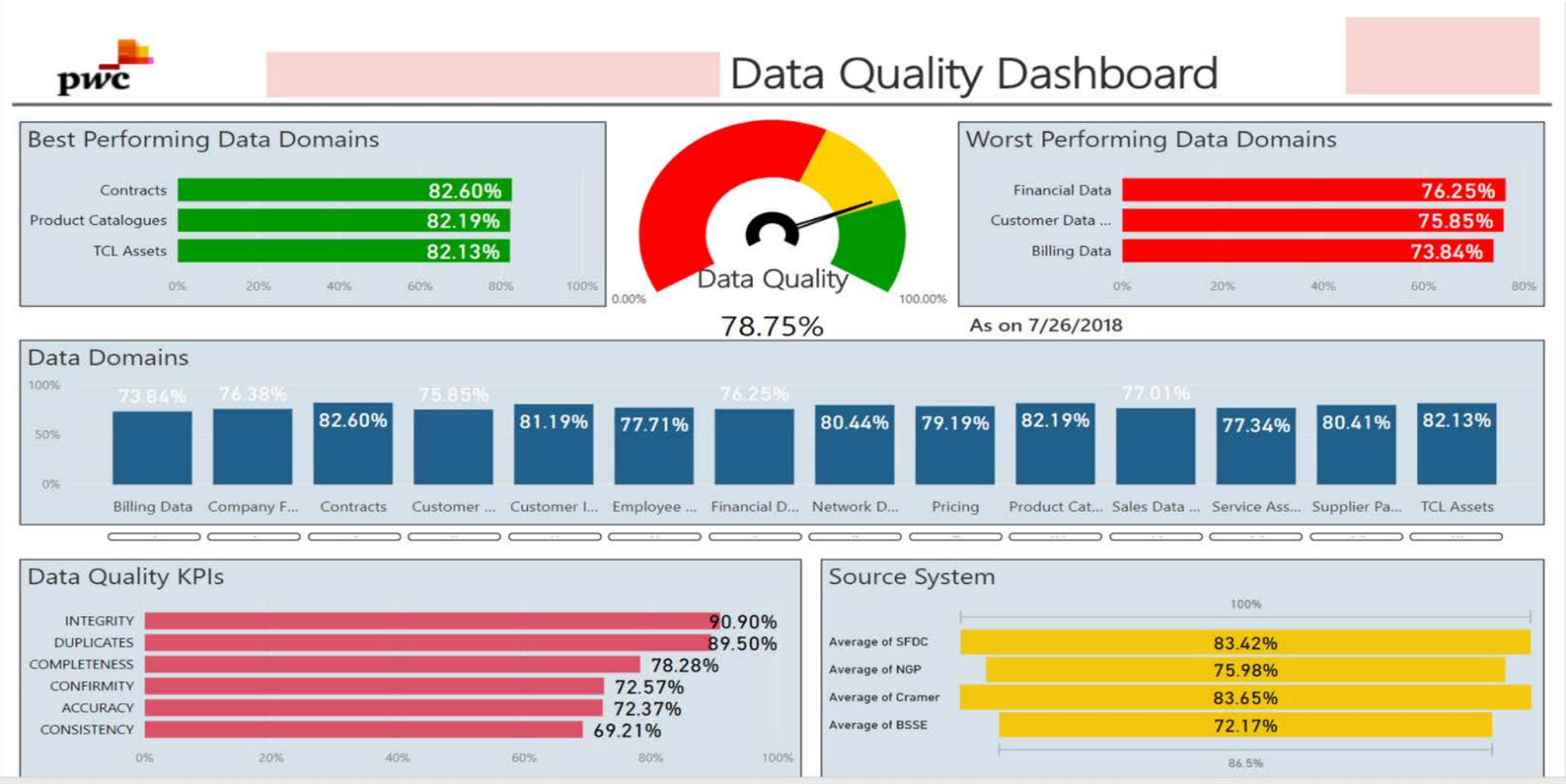
DQ maturity framework to measure success



3

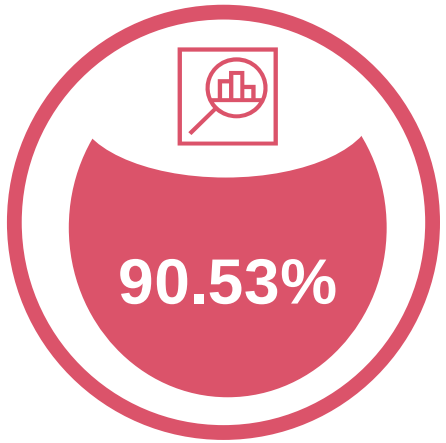
DQ accelerators

DQ accelerators – DQ KPIs dashboard



DQ accelerators – DQ functional index

Overall DQ: Revenue



Total number of records profiled: 11,879,932
Total number of columns profiled: 37

- 4.31% of the revenue-related records are incomplete. Incompleteness of key attributes increases manual efforts and leads to inaccurate reporting of revenue/profits.
- Business rule violations, such as net price being shown as zero when gross price is not zero, could occur.

DQ parameters (in percentage)

Completeness

95.69%

Validity

Need more information

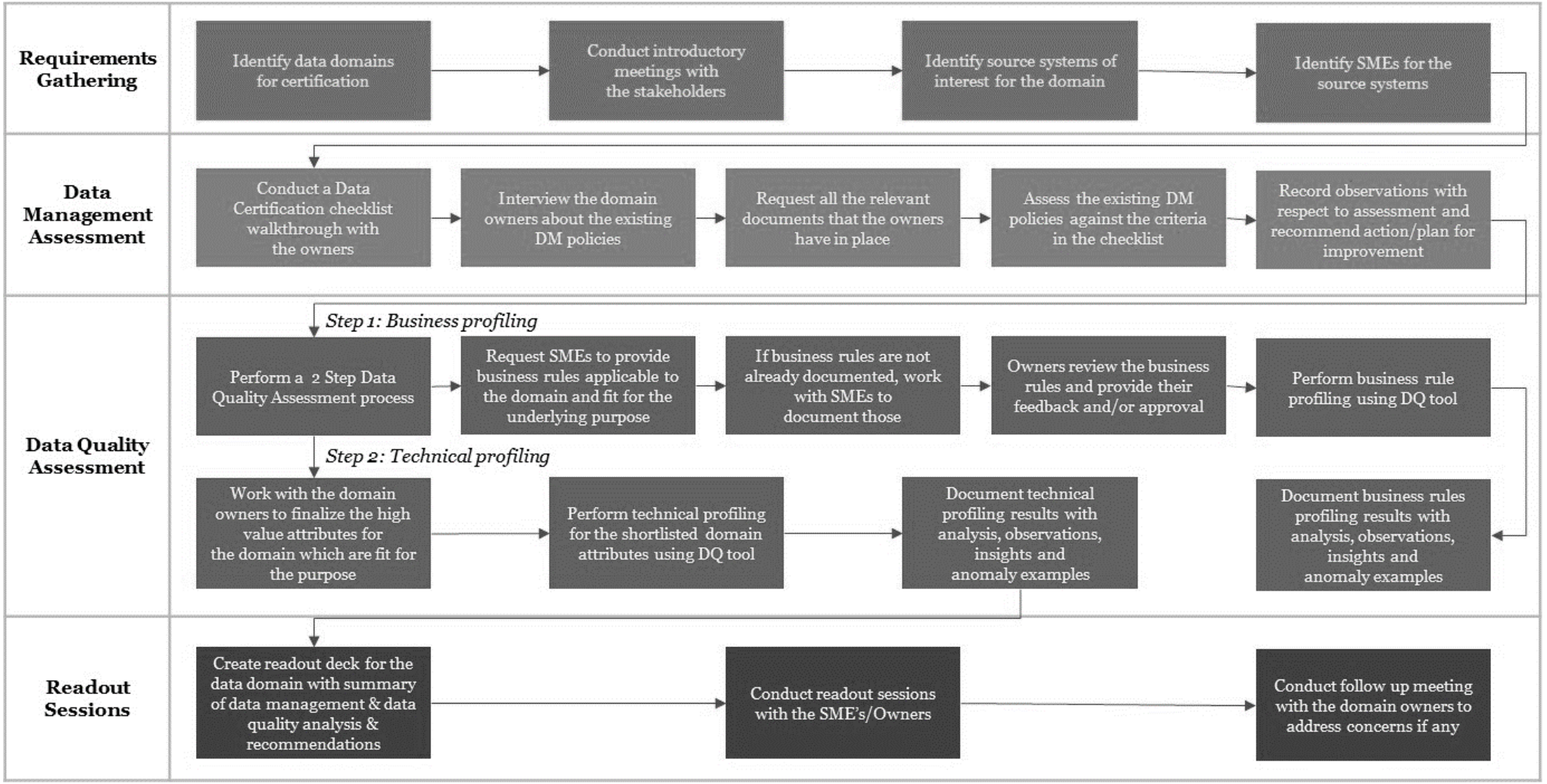
Uniqueness

98.30%

Consistency

89.3%

DQ accelerators – DQ assessment process flow



DQ Accelerators – DQ rules library

Data attributes		Attribute data cleansing rules*				
Field name category	Data attributes/fields	If column/field contains	Rule applied	Updated	Original	Standardised
Address line	CUST_BILL_TO_ADDR_LINE_1, CUST_BILL_TO_ADDR_2, CUST_SHIP_TO_ADDR_LINE_1, CUST_SHIP_TO_ADDR_LINE_2, STORE_ADDR_LINE_1, STORE_ADDR_LINE_2	ALL characters	Change to	UPPERCASE	225 Main St	225 MAIN ST
		P O,P.O.,po	Replace to	PO	P.O Box 225	PO BOX 225
		STREET	Replace to	ST	1484 STATE STREET	1484 STATE ST
		ROAD	Replace to	RD	6195 EAST SAWGRASS ROAD	6195 E SAWGRASS RD
		AVENUE	Replace to	AVE	909 AUSTIN AVENUE	909 AUSTIN AVE
		All punctuation	Replace to	White Space	C,17 L*35 ESTATES	C 17 L 35 ESTATES
		Brackets	Replace to	White Space	(Gus) 315 Pasadena Fwy	GUS 315 PASADENA FWY
		,*/+-..(any special characters)	Replace to	White Space	4565 Great Northern Blvd.,	4565 GREAT NORTHERN BLVD
		Leading & Trailing whitespace	Trim		225 Central Ave	225 CENTRAL AVE
City	CUST_BILL_TO_CITY_NAME, CUST_SHIP_TO_CITY_NAME, STORE_ADDR_CITY_NAME, City	All Records	1. City name CANNOT include ANY special characters AND/OR integer values -- Alphabet characters ONLY 2. IF city name is populated and Address Line and Postal Code / ZIP are not populated recommedation to remove city name as full address cannot be derived.			
		ALL characters	Change to	UPPERCASE	Marietta	MARIETTA
		,*/+-..(any special characters)*	Replace to	White Space	CARLA***	CARLA
		Leading & Trailing whitespace	Trim		Marietta	MARIETTA

DQ Accelerators – DQ rules library

Data attributes		Attribute data cleansing rules*				
Field name category	Data attributes/fields	If column/field contains	Rule applied	Updated	Original	Standardised
State	CUST_SHIP_TO_STATE_CODE, CUST_BILL_TO_STATE_CODE, STORE_ADDR_STATE_CODE, State	All Records	1. State code CANNOT include ANY special characters AND/OR integer values -- Alphabet characters of length(2) ONLY 2. IF state code is populated and (1) address line and/or city name is left blank AND (2) Postal Code / ZIP is left blank, recommendation to remove state code as full address cannot be derived.			
		ALL characters	Change to	UPPERCASE	ca	CA
		,*/+-..(any special characters)	Replace to	White Space	LA.	LA
		Leading & Trailing whitespace	Trim		CA	CA
Country	Country	ALL characters	Change to	UPPERCASE	us	US
		us, u.s, USA, U.S.A, UNITED STATES	Change to	US	u.s	US
		,*/+-..(any special characters)	Replace to	White Space	*us*	US
		Leading & Trailing whitespace	Trim		us	US
Postal code	CUST_SHIP_TO_POSTAL_CODE, CUST_BILL_TO_POSTAL_CODE, STORE_ADDR_POSTAL_CODE, Zip	All Records	Postal Code / Zip fields MUST follow 1 of 2 VALID patterns (1) integer length(5) e.g. 12345 OR (2) integer length(9) 12345-6789. All other patterns are invalid entries and are recommeded to be left blank.			
		123456789	Add	-	123456789	12345-6789
		,*/+-..(any special characters)	Replace to	White Space	12345-6789!	12345-6789
		Leading & Trailing whitespace	Trim		aa12345-6789	12345-6789

DQ Accelerators – DQ rules library

Data attributes		Attribute data cleansing rules*				
Field name category	Data attributes/fields	If column/field contains	Rule applied	Updated	Original	Standardised
Account / company name	CUSTOMER_SHIP_TO_CUSTOMER_NAME, Store_Name, CUSTOMER_NAME, CUST_NAME, ACCOUNT_NAME	ALL characters	Change to	UPPERCASE	After Hours Auto Care	AFTER HOURS AUTO CARE
		,*/+-..(any special characters)	Replace to	White Space	Lefty's ATV & 4*4	LEFTY S ATV 4 4
		Leading & Trailing whitespace	Trim		River City Tire Service	RIVER CITY TIRE SERVICE
		@	Replace to	AT	Tire @Wheel Masters	TIRE AT WHEEL MASTERS
		&	Replace to	AND	J & J AUTOMOTIVE REPAIR	J AND J AUTOMOTIVE REPAIR
Phone number	STORE_PHONE_NBR, CUST_PHONE_NBR, LoyaltyContactPhoneNumber	All Records	1. There CANNOT be any special characters AND/OR alphabet characters in the phone number fields. 2. Phone Number fields MUST follow ONLY one pattern (1) "(123)456-7899" 3. Phone Number fields MUST be of character length(10), ALL other patterns are invalid phone numbers and are recommeded to be left blank			
		,*/+-..(any special characters)	Add	()	1234567890	(123)456-7890
			Replace to	White Space	(012)345-6789*	(012)345-6789
		Leading & Trailing whitespace	Trim		aa(((012)345-6789	(012)345-6789

DQ Accelerators – DQ rules library

Data attributes		Attribute data cleansing rules*				
Field name category	Data attributes/fields	If column/field contains	Rule applied	Updated	Original	Standardised
Date	Enrollment Date, Cancellation Date, Registration Date	DD/MM/YYYY	Change to	MM/DD/YYYY	25/10/2018	25/10/2018
		Month Day, Year	Change to	MM/DD/YYYY	October 25, 2018	25/10/2018
		All Records	1. ALL records MUST contain the special character "@" to be a valid email. If the special character "@" is not included in the record the record is invalid and MUST be left blank			
Email	STORE_EMAIL_ADDR, CUST_EMAIL_ADDR	ALL characters	Change to	UPPERCASE	After Hours Auto Care	AFTER HOURS AUTO CARE
		Validate character @ exists	Add	@	john.doe.com	john.doe@abc.com
		Validate character ".com" exists	Add	.com	john.doe@abc	john.doe@abc.com
		Leading & Trailing whitespace	Trim		aa((john.doe@abc.com	john.doe@abc.com
Customer account number	CUSTOMER_ACCOUNT_NUMBER	ALL Records	1. IF customer account number field contains ANY special characters remove such special characters from the record			
		ALL characters	Change to	UPPERCASE	e1000	E1000
		Leading & Trailing whitespace	Trim		aa((E1000	E1000

DQ accelerators – DQ impact assessment report

Parameters	S. no	Critical use cases	Table name	Records impacted	Total records	% Failure	Few failing records	Impact description
Logical data error	1	Target should not be a decimal value	Sales_revenue_daily_target	229,083	262,425	87.29%	Product: Residential – Copper	TBC
	2	Should not contain duplicate records	Sales_revenue_daily_target	99,925	262,425	38.08%	Product: Enablement Income	0.75% of profit is wrongly reported
	3	Quantity should not be a decimal value	Sales_revenue_daily_target	217,584	262,425	82.91%	Product: Residential – Copper	TBC
	4	Gross price can't be zero or null when net price is not zero	Sales_all_channels	7,366	651,136	1.13%	Order item: ATL_Sawa_Commission	Incorrect price information leads to wrong financial reporting
	5	Net price can’t be zero or null, especially when gross price is not zero	Sales_all_channels	16,421	651,136	2.52%	Order item: MRYJ2AH/A	6.34% of revenue can be wrongly reported
	6	Order quantity is negative and net amount is zero	Sales_all_channels	1,566	651,136	0.24%	Order item: MRYJ2AH/A	If order quantity is negative, then gross amount is negative which can lead to loss in reporting
	7	Gross price can't be zero or null when net price is not zero	ABC_daily sales_cost_sum	7,249	193,332	3.75%	Order item: 1993010000000	Incorrect price information leads to wrong financial reporting
	8	Net price can't be zero or null, especially when gross price is not zero and its not a reversal product	ABC_daily sales_cost_sum	16,419	193,332	8.49%	Order item: MT9G2AH/A	6.35% of revenue can be wrongly reported
	9	Order quantity is negative and net amount is zero	ABC_daily sales_cost_sum	3,215	193,332	1.66%	Order item: ITM-00016	If order quantity is negative, then gross amount is negative which can lead to loss in reporting
	10	Invoice quantity is null whenever order quantity is zero	ABC_daily sales_cost_sum	61	193,332	0.03%	Order Item: MBHWMATE20PRO128GBTBLK	Null invoice quantity causes reporting errors

DQ accelerators – DQ maturity assessment model

Governance assessment areas	Sub areas	1. Initial/ad hoc/informal	2. Repeatable/reactive/emerging	3. Defined/standardised/ structured	4. Managed/predictable /sophisticated	5. Optimised/improving/strategic
DQ practices	Challenges, issues and gaps	…not possible to identify; no understanding of data quality	…can be potentially identified; need to define data objectives and assign quality measures	…addressed at the LOB level; data quality consistently understood but LOB objectives differ	…centrally managed, aligned; data quality well understood but LOB practices differ	…data quality issues are completely transparent; LOB contribution understood across
	DQ assessment toolset	DQ is done on an ad-hoc basis usually using SQL and Excel.	Basic data-profiling tools are adopted and available for use anywhere in the system development lifecycle.	DQ reporting capabilities are implemented and available to any system.	DQ profiling and issue remediation is integrated into quality reporting platform.	DQ remediation is implemented in the entire data life cycle – i.e. on data ingestion (primary systems), on data at rest (in databases) and data in flight (in ETL and as messages between systems)
	DQ improvements …	…are inconsistent across LOB; data proliferates without control, quality is inconsistent across application silos	…are executed through informal processes with little understanding and usage; data management is reactive in nature	…follow formal processes and procedures in places with <50% of enterprise usage; data management is proactive in nature for critical data only	…follow formal processes and procedures in places with >50% of enterprise usage; data management in general is proactive; quality measures are standardized and published	…follow formal enterprise processes and procedures and used across greater than 90% of the organisation.
	Remediation processes…	…do not exist; no mechanism to measure data quality	…are completed via ad-hoc manual quality evaluations	…rely on informal tools used on an manual basis	…rely on formalised tools used on an automated basis	…rely on automated enterprise tools to measure data quality.
	Reconciliation processes…	…are not possible or are ad hoc; little understanding of the data quality issues – most are undetected	…are random and handled individually by LOBs; data quality measures are standardized at LOB level but inconsistent cross the enterprise	…are executed after data loads are completed; data quality measures are standardised across the enterprise; addressing errors is manually intensive	…completed by a data quality group; errors addressed at the source without interrupting normal business operations	..handled by a data quality COE; quality issues are fixed within the standard process in real time.

4

Resources and experience

PwC is one of the largest providers of data and analytics services in India, supported by global expertise

Our data and analytics experience

250+ data and analytics implementations in India (last 10 years)	**50+** implementations of reporting solutions in India
170 countries in our global delivery network	**16** industries (29 sub-industries) serving clients with strategy through execution capabilities
9,700+ technology consultants worldwide	**950+** data and analytics consultants in India

PwC India's Technology Consulting practice is certified at CMMI Level 5

PwC in Gartner's Leaders Quadrant for Service Providers

Our DQ experience

20+ DQ solution implementations (last 10 years)	Platform-based toolkit consisting of templates and accelerators and latest DQ solutions with regular technology updates
15+ resources part of PwC's DQ CoE	DQ rules inventory, assessment process flow, impact assessment report, DQ maturity assessment model, DQ KPI dashboards
10+ industries with a focus on retail and consumer, healthcare, manufacturing and BFSI	Design and development guidelines, blueprint documentation, gap analysis document, Industry-specific business rules

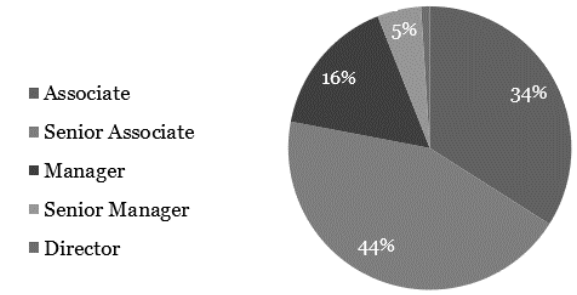
PwC is a member of the ASG, Erwin and Infogix Alliance Program.

End-to-end consulting capability for data and analytics across industries

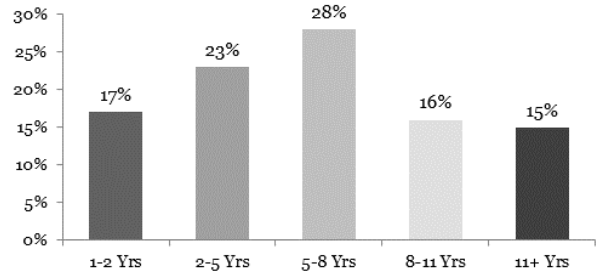
Leadership: 9 Partners, 3 Executive Directors and 14 Directors

A pool of 1,000+ D&A consultants and 200+ specialised in advanced analytics

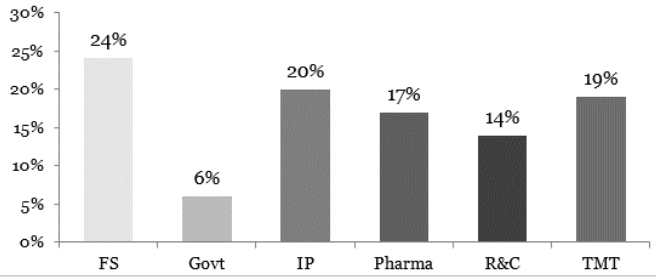
Designation-wise distribution of resources



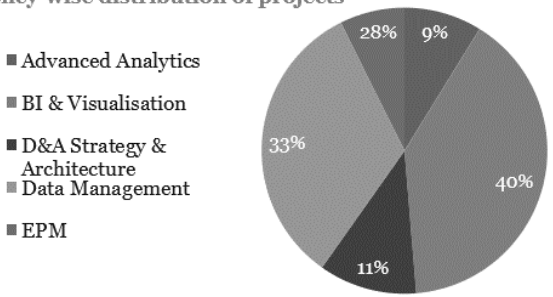
Experience-wise distribution of resources



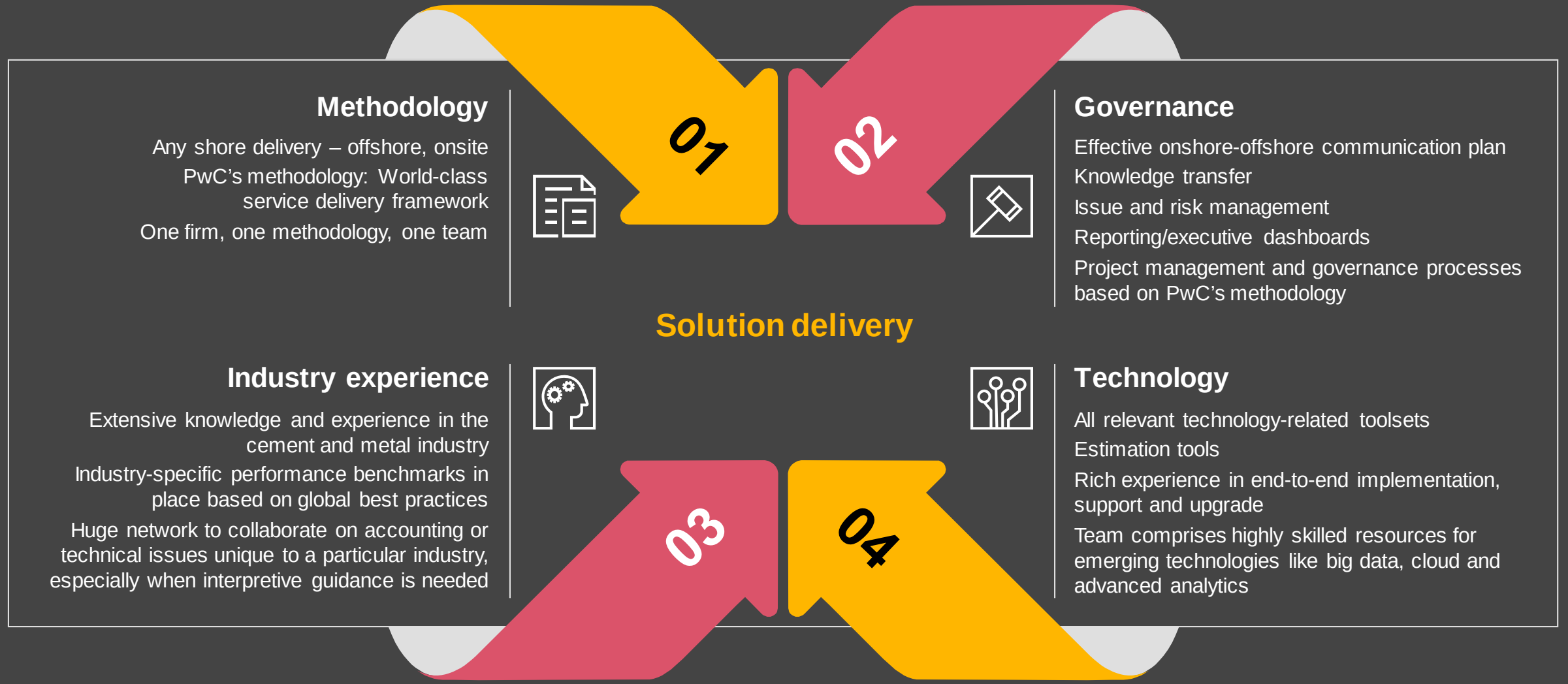
Industry-wise distribution of projects



Competency-wise distribution of projects



How we operate: Solution delivery framework



Methodology

Any shore delivery – offshore, onsite

PwC's methodology: World-class service delivery framework

One firm, one methodology, one team

Governance

Effective onshore-offshore communication plan

Knowledge transfer

Issue and risk management

Reporting/executive dashboards

Project management and governance processes based on PwC's methodology

Industry experience

Extensive knowledge and experience in the cement and metal industry

Industry-specific performance benchmarks in place based on global best practices

Huge network to collaborate on accounting or technical issues unique to a particular industry, especially when interpretive guidance is needed

Technology

All relevant technology-related toolsets

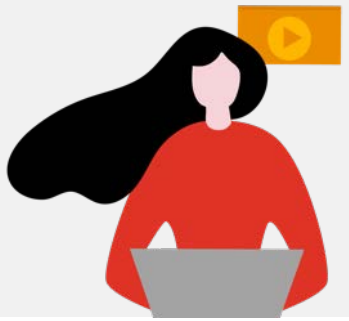
Estimation tools

Rich experience in end-to-end implementation, support and upgrade

Team comprises highly skilled resources for emerging technologies like big data, cloud and advanced analytics

Our data and analytics capabilities across different competencies

The Data and Analytics competency helps clients increase the value of their data. We help clients define their information strategy, architecture and governance; implement enterprise content and data management solutions; and get the most from business intelligence and analytics by transforming business information into better, more timely decision making.



Information strategy, architecture and governance

Helping clients drive architecture, organisation, and alignment of information assets to deliver value, make better decisions and manage risk.

Service offerings

- Information strategy and roadmap
- Information architecture
- Data governance strategy

Business intelligence

Helping clients address the technical, process, and business dimensions of transforming data into actionable information

Service offerings

- Assessment and health checks
- CoE
- Tool selection and rationalisation
- Implementation

Enterprise content management

Helping clients turn unstructured and semi-structured content into meaningful and usable information

Service offerings

- Web content management
- Document management
- Digital/media asset management
- Record management
- Knowledge/idea management
- Contract management

Enterprise data management

Helping clients organise, integrate, and retrieve information assets to reduce cost and complexity, increase trust and integrity, and improve operational effectiveness

Service offerings

- Reference and master data management
- Data integration
- Metadata management
- Data warehousing
- Data migration and conversion
- Data modelling
- DQ

Advanced analytics technology

Helping clients with real-time analysis of no-conforming data sources to enable insights across federated and/or complex data sets – often encompassing machine learning techniques

Service offerings

- Big data strategy and assessment
- Big data architecture
- Big data and analytics CoE
- Big data implementation
- AI- and ML-based models

Our select experience across DQ tools and technologies that support our offerings

Tool components	Activities performed	Expertise
Trillium Data Quality	• Installation and configuration of Trillium • Project implementation • Address validation and enrichment	Experts and SMEs available (worked on Trillium v11, v13, v14)
Informatica Data Quality	• Installation and configuration of IDQ • Project implementation (including multiple languages as well) • Integration with ETL tools (like ODI, Informatica) on UNIX and Windows platform	Experts and SMEs available
Data Flux /SAS DQ	• Installation and configuration of DataFlux/SAS DQ • Project implementation with national as well as international data • Integration with ETL tools on UNIX and Windows platform	Experts and SMEs available
Talend DQ	• Tool configuration • Project implementation	Experts and SMEs available
Data Gym	• Installation and configuration of Data Gym • Project implementation	Experts and SMEs available
IBM Information Analyzer	• Installation and configuration of IA • Project implementation	Experts and SMEs available

Our engagement delivery across DQ tools and technologies

Client/industry information	Trillium	IDQ	IBM Information Analyzer	DataFlux/ SAS DQ	Siperian	Quality Stage	Data Gym
Leading banking/insurance client in India	✓	✓		✓			✓
Fuel industry client in the USA	✓						
Healthcare client in the USA	✓	✓					
Multinational banking client		✓		✓			
Large Indian FMCG company		✓			✓		
Large telecom major		✓		✓		✓	
Real-time assurance	✓			✓			
Hospitality major in the USA			✓				

5

Case studies and citations

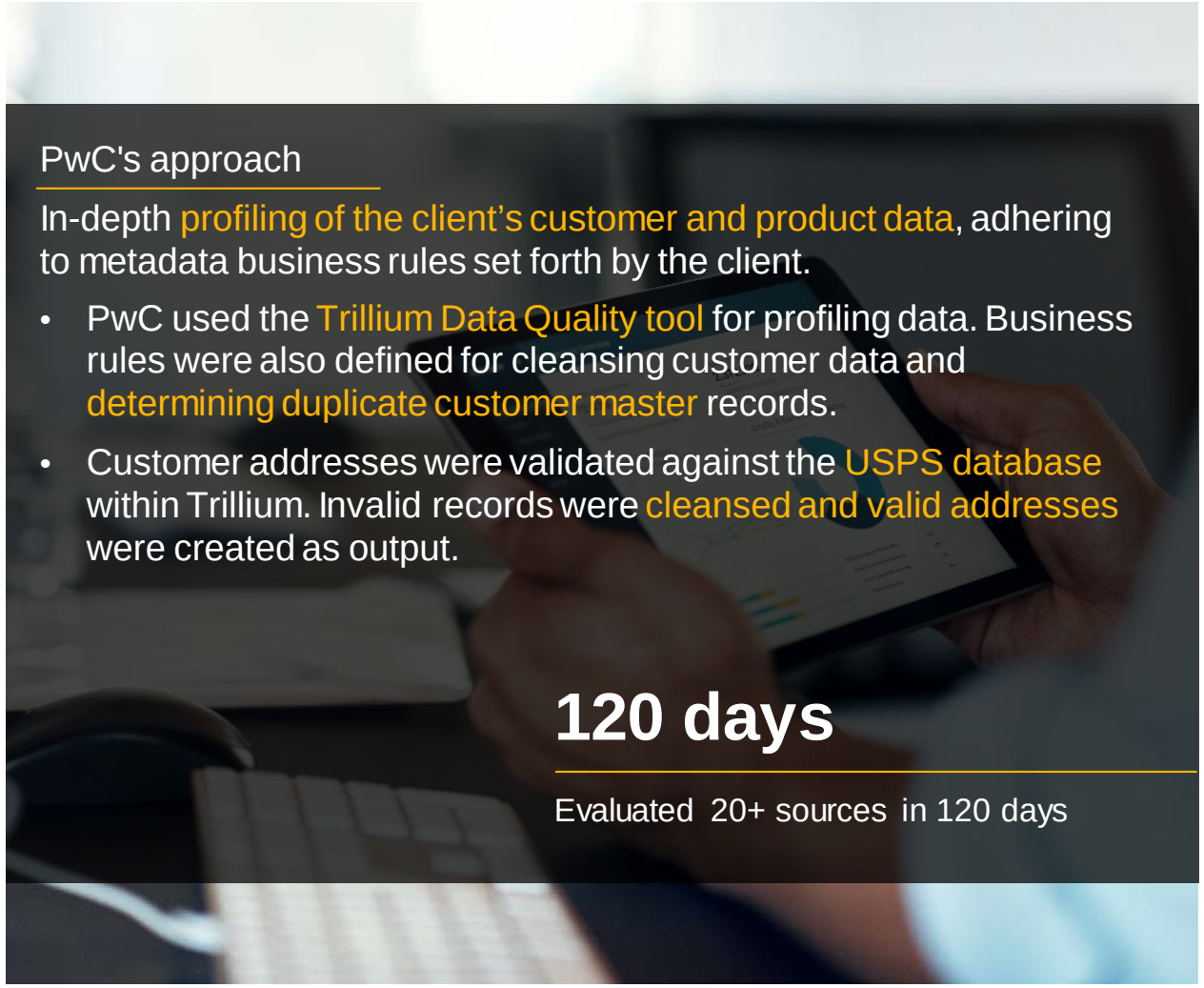
Case study 1

DQ assessment for a large healthcare corporation in the USA

The healthcare client distributes healthcare systems, medical supplies and pharmaceutical products. Additionally, it provides extensive network infrastructure for the healthcare industry.
The client has actively used master data such as vendor, customer and material masters in various enterprise resource planning (ERP) and CRM systems for its master data management operations. The client was facing issues related to incorrect details, data inaccuracy, duplicate data, inconsistent data and issues related to integration of various database interfaces.

Client benefits

- The results delivered by PwC helped the client to determine the appropriateness of source system data in satisfying the data requirements, and to define cross-system update rules
- The client's data was found to be in line with industry standards, except for a single source system that produced a high volume of duplicates.
- It was also determined that the client's previous DQ initiatives had paid off.



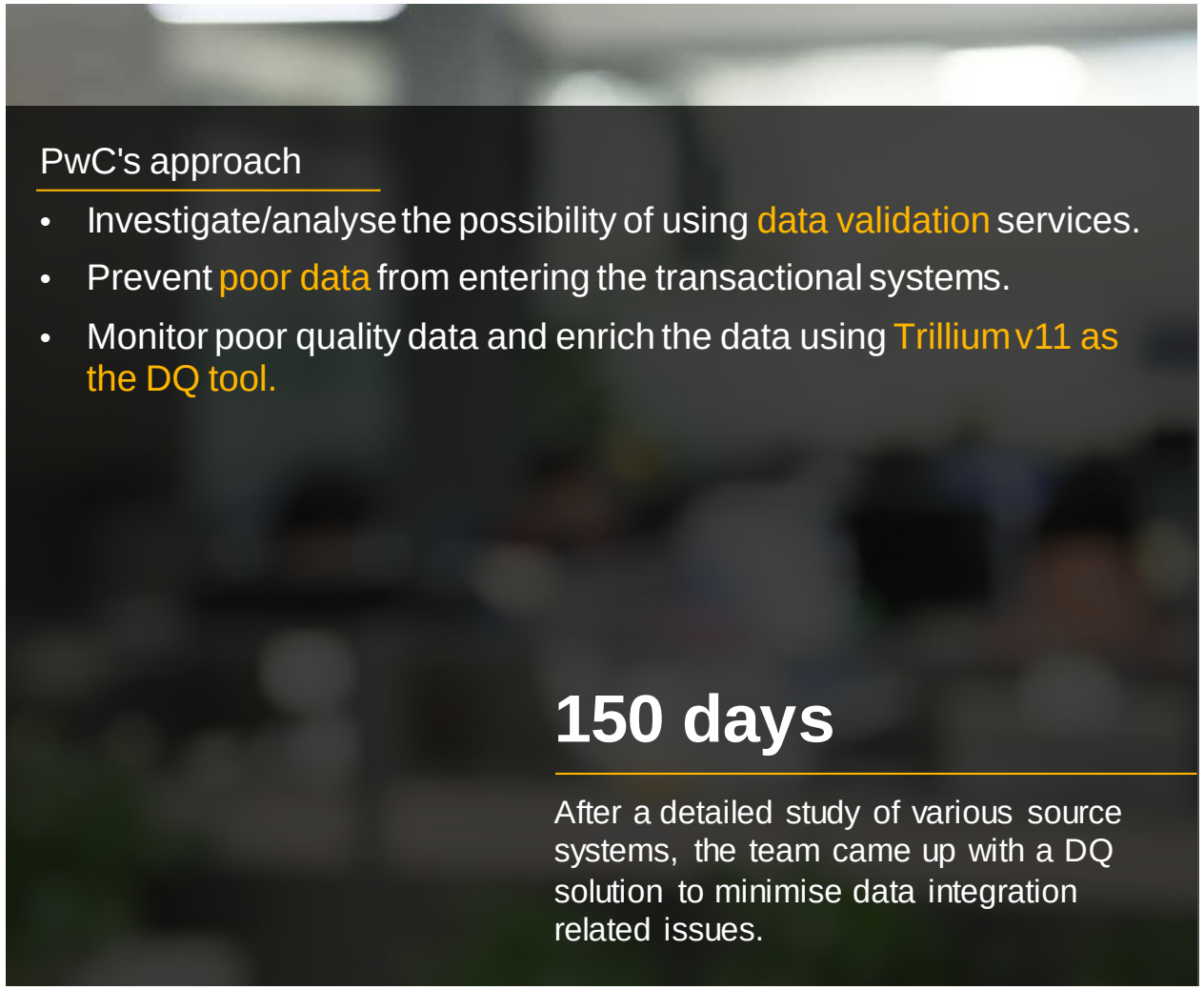
Case study 2

Customer DQ assessment for an international professional services firm

The DQ project was undertaken in order to establish a set of firm-wide DQ metrics and analyse/profile data from core transactional applications which included the CRM, financial and HR applications. The project also aimed at improving DQ awareness within the firm and establishing a set of operating procedures for fixing and preventing DQ problems.

Client benefits

- High quality of data
- Data consistency across the systems
- Sanctity of data is maintained



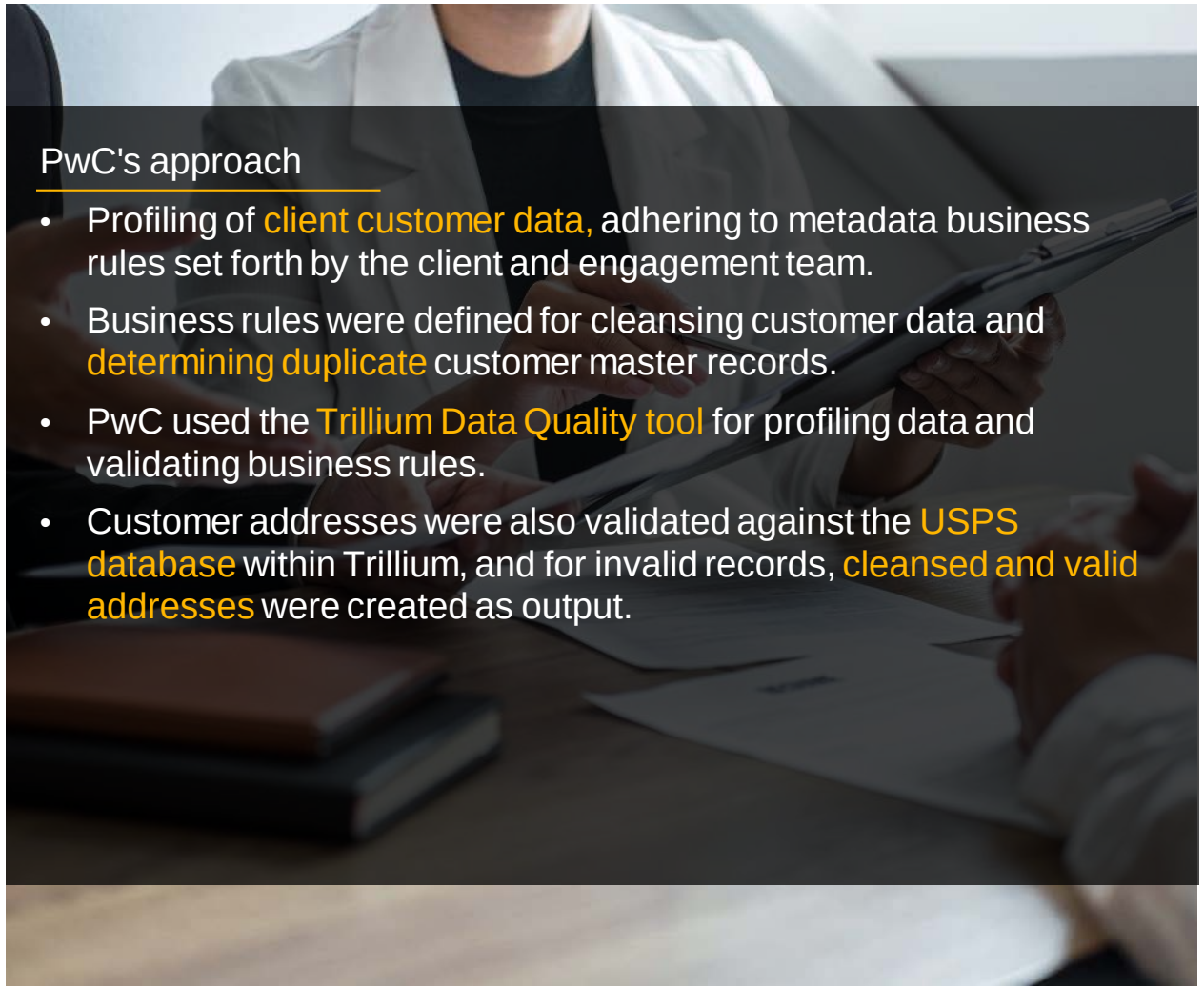
Case study 3

DQ assessment engagement for a petroleum corporation in the USA

Client is a USA-based refiner, transporter and marketer of transportation fuels, lubricants, petrochemicals and other industrial products. The client was seeking to profile customer data to obtain a DQ baseline for critical fields, identify DQ issues, and validate relationships between relevant tables.

Client benefits

- High-level statistics and detailed results were provided by technical profiling.
- Customer addresses were cleansed and provided accurate address records.
- Merge candidates were identified and records were merged accordingly, assisting the client in identifying disparate data among different source systems.



PwC's approach

- Profiling of client customer data, adhering to metadata business rules set forth by the client and engagement team.
- Business rules were defined for cleansing customer data and determining duplicate customer master records.
- PwC used the Trillium Data Quality tool for profiling data and validating business rules.
- Customer addresses were also validated against the USPS database within Trillium, and for invalid records, cleansed and valid addresses were created as output.

Case study 4

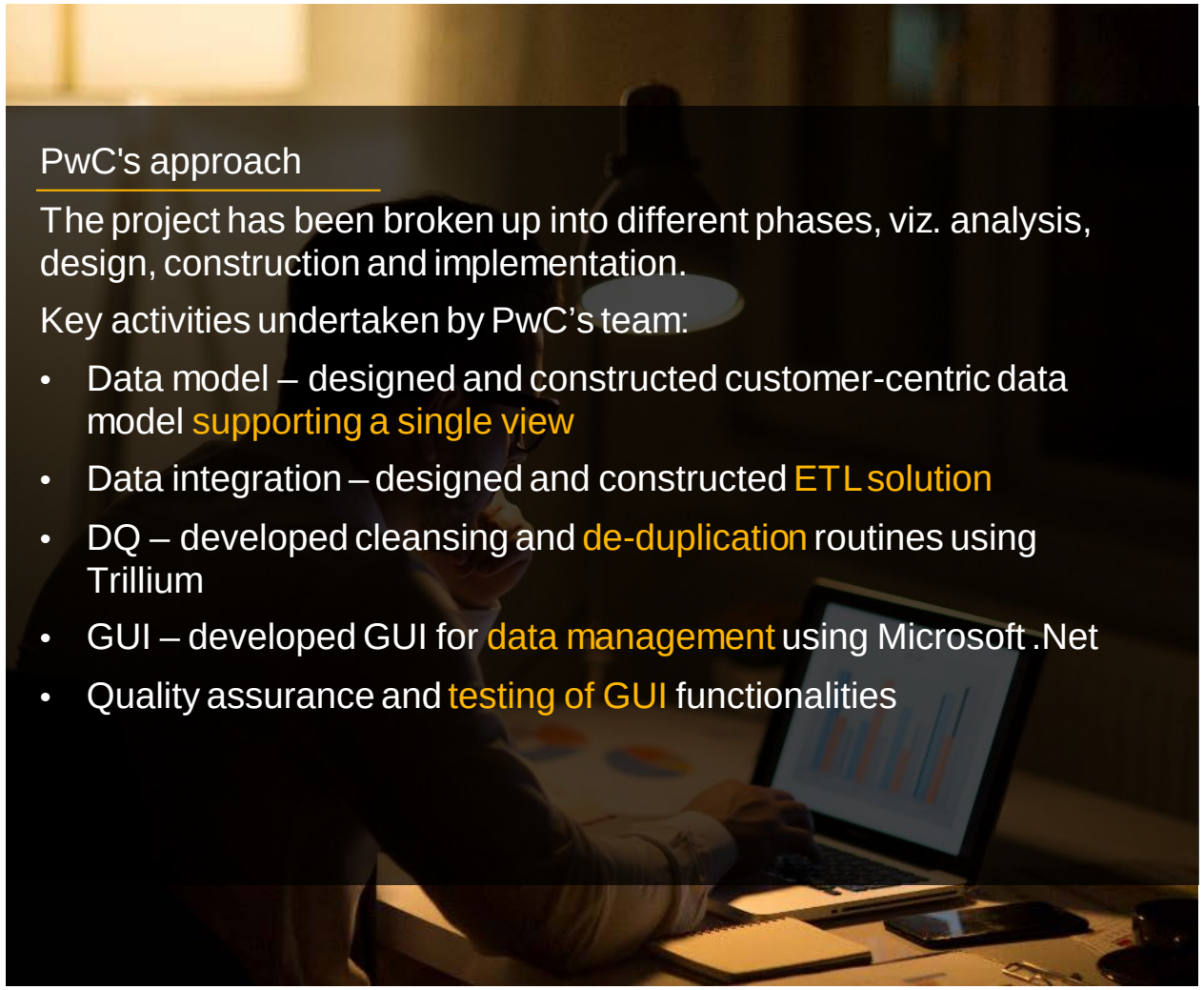
Customer data integration and GUI implementation for a large bank in Ukraine

The client is a diversified financial institution providing a full range of banking services throughout Ukraine. Key business challenges and needs (across retail, corporate and wealth) included:

- Architectural limitations of applications leading to client record duplication – a 'single view of clients' (including accurate contact information and information on the banking products used) is not available
- Low quality of client information leading to incorrect management decisions, impact on risk management (including anti-money laundering, collection) and insufficient information for marketing campaigns

Client benefits

- Single view of customer and GUI for user-defined fields management
- Cleansed data ensuring a single version of truth across customer information reporting
- Facilitating customer data analytics



PwC's approach

The project has been broken up into different phases, viz. analysis, design, construction and implementation.

Key activities undertaken by PwC's team:

- Data model – designed and constructed customer-centric data model supporting a single view
- Data integration – designed and constructed ETL solution
- DQ – developed cleansing and de-duplication routines using Trillium
- GUI – developed GUI for data management using Microsoft .Net
- Quality assurance and testing of GUI functionalities

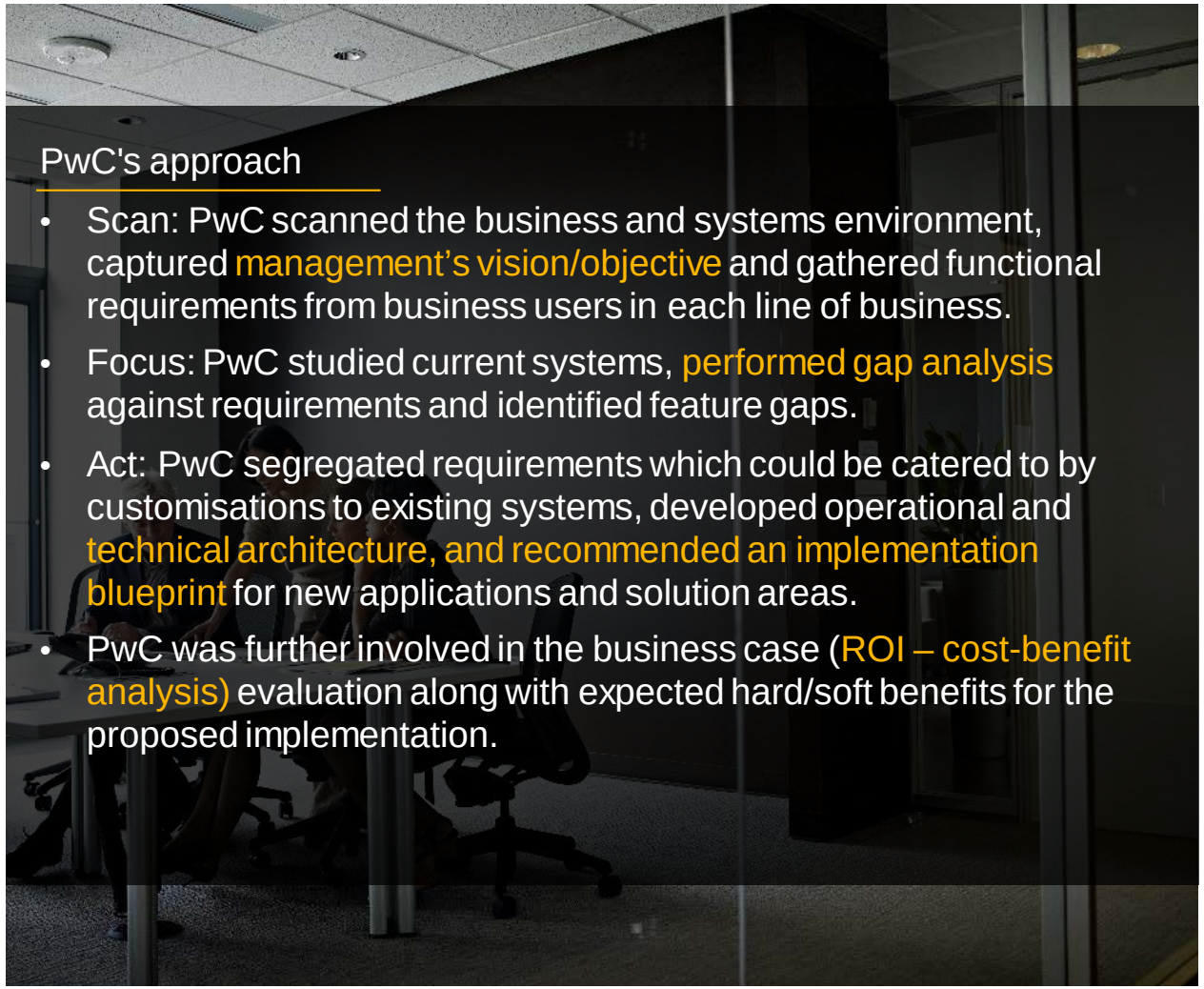
Case study 5

EDW roadmap and DQ tool evaluation for a financial services group

The organisation wishes to review customer accessing and targeting processes and systems across its business lines (AMC, insurance and distribution) with a goal towards revenue uplift and enabling operational efficiencies, especially in front-line processes, i.e. sales, services and marketing.
PwC was engaged to define functional requirements, analyse gaps w.r.t. current application systems, design technical and operational architecture, and recommend tools/solutions (BI, ETL, MDM, CRM and Portal) for implementation in a phased manner.

Client benefits

- Scalable technology platform for growing financial services business
- Standardise deliveries to channel partners and customers
- Measure and improve productivity of teams and distribution
- Differentiate offerings from competition



PwC's approach

- Scan: PwC scanned the business and systems environment, captured management's vision/objective and gathered functional requirements from business users in each line of business.
- Focus: PwC studied current systems, performed gap analysis against requirements and identified feature gaps.
- Act: PwC segregated requirements which could be catered to by customisations to existing systems, developed operational and technical architecture, and recommended an implementation blueprint for new applications and solution areas.
- PwC was further involved in the business case (ROI – cost-benefit analysis) evaluation along with expected hard/soft benefits for the proposed implementation.

Case study 6

DQ solution implementation at a leading bank in India

The bank collects customer information from various sources like core banking, credit cards, loans, demat accounts and third-party MFs.

So, the primary concern was a single customer view in the existing data warehouse after de-duping/clustering. Apart from that, house-holding, standardisation and some augmentation were also part of the solution.

Client benefits

- Single view of customer and GUI for user-defined fields management
- Cleansed data ensuring a single version of truth across customer information reporting
- Facilitating customer data analytics

Case study 7

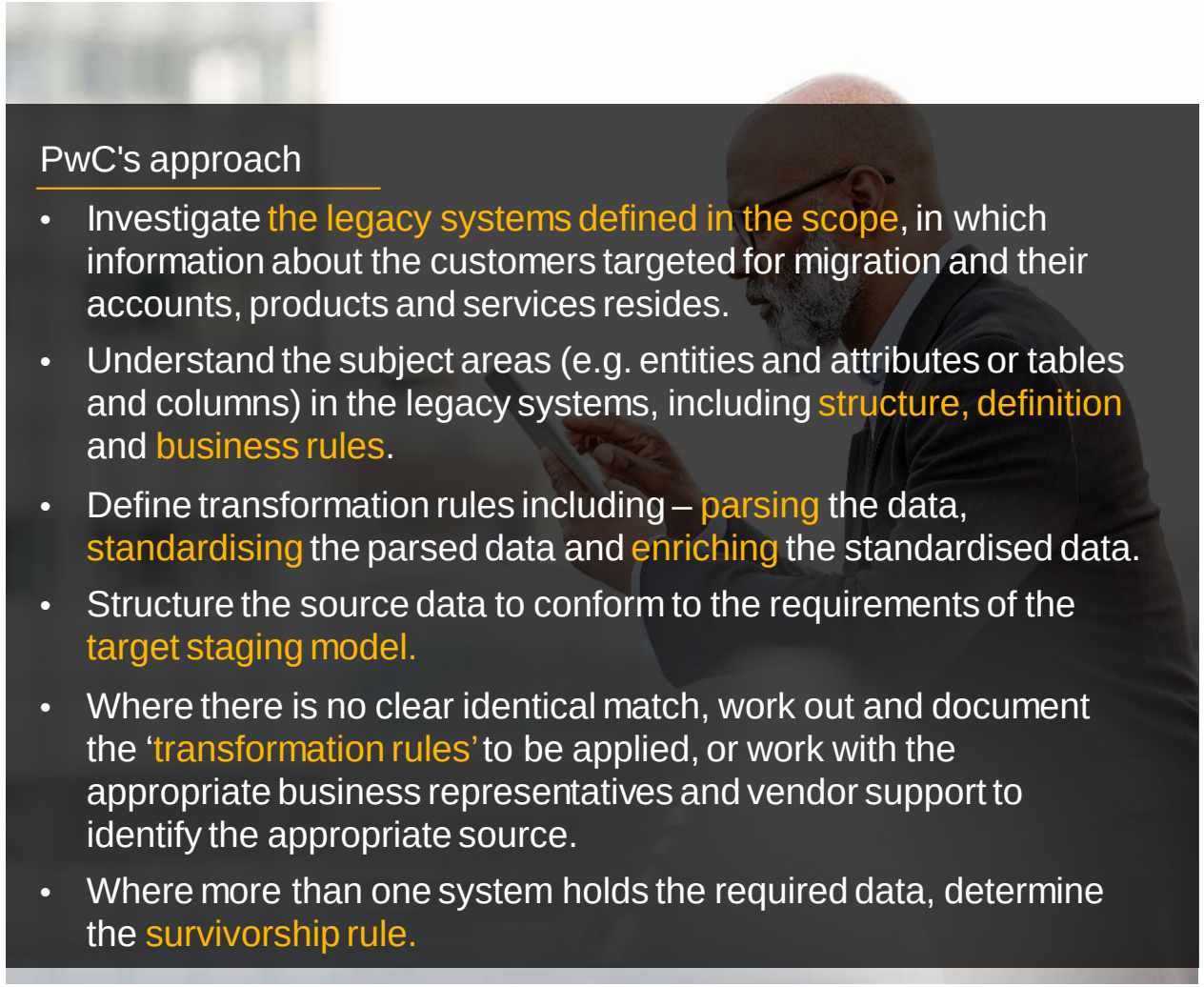
DQ and MDM implementation at a leading telecom major in Australia

The client had undertaken a massive data transformation project. The objective of this project was to migrate data from the numerous (close to 200) existing billing and customer systems to Seibel CRM application and Kenan billing application.

In the process, the customer, service, product and provisioning information had to be cleansed and a master data set had to be maintained.

Client benefits

- Successful migration of all the customers from the various legacy systems
- Reduction in overhead of maintaining multiple legacy systems for various applications
- A master data set for all the customers, billing accounts, services and product information in the target system



PwC's approach

- Investigate the legacy systems defined in the scope, in which information about the customers targeted for migration and their accounts, products and services resides.
- Understand the subject areas (e.g. entities and attributes or tables and columns) in the legacy systems, including structure, definition and business rules.
- Define transformation rules including – parsing the data, standardising the parsed data and enriching the standardised data.
- Structure the source data to conform to the requirements of the target staging model.
- Where there is no clear identical match, work out and document the 'transformation rules' to be applied, or work with the appropriate business representatives and vendor support to identify the appropriate source.
- Where more than one system holds the required data, determine the survivorship rule.

Case study 8

DQ implementation programme for a leading manufacturer of medical devices in the USA

The client is dedicated to transforming lives through innovative medical solutions that improve the health of patients around the world. PwC was appointed to perform a data health assessment to get a DQ baseline for critical data elements, identify DQ issues, and validate relationships between relevant tables across different business functions. Our client was undergoing migration activities and such data migration usually requires data cleansing to achieve the desired DQ level in the destination system. During the migration, the client was facing some risks/challenges related to data corruption and data integrity, semantic issues, and several data anomalies.

Client benefits

- Provided cleansed and accurate address records
- Identified merge candidates and merged records which are critical for data migration activities
- Helped the client to perform more accurate planning, budgeting, and external disclosure

PwC's approach

We performed a DQ assessment on vendor, product, etc., using Informatica Data Quality (IDQ).

- Worked closely with the migration programme team and end users to identify DQ issues and profiled around 200+ attributes.
- Identified opportunities for DQ improvement, DQ control, and necessary data requirements to support the implementation of the new S/4 HANA system.
- Analysed data readiness for the new S/4 platform, and interim needs to develop a data ingestion strategy and roadmap.
- Helped to identify the complex extract-transform-load (ETL) conversion objects.
- Generated a report with all the anomalies based on DQ dimensions – existence, completeness, integrity, consistency, accuracy and uniqueness.

Contact us

Mukesh Deshpande
Leader, Data Management
PwC India
mukesh.deshpande@pwc.com

Amit Lundia
Leader, Data Governance
PwC India
amit.lundia@pwc.com

Mohua Sen
Data Quality Lead
PwC India
mohua.sen@pwc.com

pwc.in

VB/March 2021 - M&C 11502