

A CDO's guide on Critical Data - The What, Why & How of it!

CRITICAL DATA ELEMENT





Table of Contents

EXECUTIVE SUMMARY	2
WHAT MAKES A DATA CRITICAL	5
WHAT IS A CRITICAL DATA?	6
GROUPING OF CDE → DATASET/DATA CONCEPT	7
WHY IDENTIFY CRITICAL DATA ELEMENTS	9
REDUCE COMPLEXITY	9
MANAGING DATA QUALITY	9
MAKES LINEAGE CREATION EASY	10
EASILY IMPLEMENT CONTROLS TO ENSURE DATA INTEGRITY & FITNESS-FOR-PURPOSE.....	10
THE MULTIPLIER EFFECT OF CDE CLEANSING.....	11
ASSIGN OWNERSHIPS / DATA STEWARDS.....	11
MULTIDIMENSIONAL VIEW OF CRITICAL DATA	11
TAILORMADE POLICIES TO GOVERN CRITICAL DATA	12
MARGINAL BENEFITS OF EXPANDING BEYOND CDE	13
HOW TO SHORTLIST CDE	14
THE PHASED APPROACH	15
<i>Phase 1: Stop the Bleeding</i>	15
<i>Phase 2: Towards CDE Maturity</i>	17
CONCLUSION	23
REFERENCES	25
CONTACTS	27



EXECUTIVE SUMMARY

Since the onset of *information age*, data invariably has been a pain point for every organization. The deeper a firm goes into its data to derive value, higher the number of data issues it encounters in the process. While data quality as a concept has been around for a period as long as half-a-century; yet, the efforts never bore fruits to the fullest due to lack of effective data strategy required to string together a wholistic data management approach that can be implemented through-out organization.

With the advent of data governance guidelines in recent periods and the accompanying widely accepted requirement of focusing only the bare minimum data that is critical to an organization; data management started to seem like an achievable task and rightly so! Indeed data management is achievable, but not without the inherent challenges posed by the process.

Shortlisting critical data elements and handling them with care because they are key inputs to a firm's critical business processes and reports; are the keys to an effective data governance program. One of the important lessons this paper covers is the factors that differentiate a critical data from non-critical.

A given data/dataset becomes critical when its usage is deemed critical by the organization for its fully efficient and compliant functioning. When it comes to levels of criticality, the critical data follows hierarchical levels i.e. what is key/critical data for an organization (as a whole) may not be critical to its various functional departments and vice-versa. Hence, critical data needs to be defined at multiple levels within an organization.

Shortlisting critical data essentially means arriving at a list of physical data attributes that will receive special treatment. A critical data element (CDE) may comprise of more than one such attributes. For instance, *customer name* as data element will have at least three attributes i.e. *first name, middle name & last name*. The list of CDEs must



be logically grouped to make it easier to manage and to keep a multi-dimensional view on the critical data that is relatable to every business user.

Shortlisting CDEs is not only a regulatory requirement in multiple industries but it also serves numerous benefits by reducing complexity involved in managing data. It makes easier for firms to manage quality of smaller number of key data and also ensure data integrity & fitness-for-purpose with implementation of control policies, security measures etc. Creating lineage on critical data further helps in removing manual interventions throughout data lifecycle. The shortlisted CDEs must be assigned ownership (usually under a data officer) at organization or group levels.

There are no set guidelines on the process of identifying critical data. Policy makers mandate that the CDEs be shortlisted by firms under their regulatory oversight however, refrain from prescribing any standards or guidelines. This is because they expect the data owner must fully understand their data and its usage to be able to identify the critical elements out of it.

This paper advises firms to adopt phased approach in shortlisting their critical data elements. In first phase, start from the areas that require immediate attention to *stop the bleeding* and fix the glaring issues with respect to data. The areas that can be focused at this stage are regulatory reports, bottleneck processes and even crucial projects.

In the next phase that focuses on expanding the areas of identifying critical data, stakeholders (including business & data users, report owners, data officers) are required to vote/rate the criticality of data on various criteria such as legal, regulatory financial, reputational, profitability etc. The ratings against each criterion can be allocated in the form of weight-age assignment or simple voting process by stakeholders that captures individual inputs on likely implication of each data element could induce towards every aforementioned criteria.

This approach relies heavily on deep understanding of stakeholders about their data, its value & criticality in terms of usage. Although, it is a majorly voting process, yet the final decision on shortlisted CDEs should be subjected to approval of chief or



group/function data officer who in-turn must take into account the majority of votes/weight-age received by each data element.

The data elements shortlisted as CDE through the simple voting/rating process should further be refined by applying statistical approach that acts as a funnel for further trimming the list. By using correlation analysis, the highly correlated pairs of CDEs can be identified such that by assessing only one of them can the predictions about other element be made.

By further applying regression analysis on highly correlated CDE pairs, the degree of change induced in one CDE can be deduced when the other CDE changes by a unit. The regression analysis approach offers the vital ability to assess the impact on business by only focusing on very limited number of the input CDEs as many of the critical elements can be represented as function of their highly correlated CDE peer.

By sequentially applying both Correlation & Regression on pairs of CDEs, businesses can substantially reduce the number of critical data elements to manage while confidently trusting that the list (although smaller) holistically captures the health of the data that goes into their business critical reports, processes & projects.



WHAT MAKES A DATA CRITICAL

Even if you are remotely related to the world of data; chances are, you must have (at least once) heard someone saying, “not all data are created equal and some are important than others”! Irony is, this widely spoken statement does not tell you the whole story. Because, it is not the data that is critical – but the very ‘*usage of data*’ that makes it critical.

What is important for one department in a firm may not necessarily be *equally* important to other department. Even more interestingly, what is important to a firm as a whole may not necessary be important to some of its integral business functions. And it is this very property of data that makes job of a CDO challenging and exciting at the same time.

For instance, *marital status* of a customer may not mean anything to a data user whose job is to report overall exposure against the said customer. But it makes all the difference for a marketing head to know their customer’s *marital status* so they can incorporate this information in their market segmentation and promotions. Furthermore, the sales personnel with access to such information can accordingly pitch right products to this customer, such as a honeymoon package, a joint-home loan, credit card or savings account. Therefore, the usage of same data attribute can be plenty for certain business functions but none for others. And therefore, data found to be critical in one function may not be critical for another.

Having said that, even though *marital status* is an important data element for sales & marketing function – can it really be considered a “Critical” data element for the bank as an entity”? Not at all!

Critical data are hierarchical in nature – what is critical for the enterprise may not be critical for its business function/s and vice-versa. A bank may consider only those data elements as critical which are capable of inducing legal, financial or compliance implications and have substantial impact over business processes or are key inputs to



their regulatory/other reports. At the same time, a department within bank may shortlist only those elements as CDEs that help them smoothen their processes, meet their targets and also support in their compliance obligations.

Because a critical data element must have direct implication on a business' smooth functioning & efficiency. Therefore, it must be consequential and vital enough for a bank to go through the efforts to - track, preserve and protect it from defects to avoid undesirable regulatory, financial or even legal consequences. Thus, *marital status* does not qualify to be impactful enough to be considered a critical data for a bank.

Furthermore, as mentioned earlier – it is the usage that makes the data critical. Therefore, an industry that relies heavily on collecting *marital status* of their customers may find it critical. For instance insurance; consider picking up a dataset of customers who purchased fixed-term life insurance product for their spouse in last one year but your database shows that 2% and 3% of these customers have their marital status as 'unmarried' and 'divorced', respectively. Not to mention the possibility of finding out the information totally missing for the said attribute.

What does it point to? A data quality defect! Yes. But not just that, it also points out at the initiation of a popular first step in the process of money laundering (called placement) where illegally earned cash is channeled towards financial products involving lesser scrutiny on source of funds.

Realize how a simple looking data quality defect can actually give rise to implications that are serious enough for the insurer to consider '*marital status*' as a critical data element! And hence, it's the usage that makes the data critical and the usage varies from one industry/business function to another.

What is a Critical Data?

Before we agree on the definition of critical data – let's settle on the usage & differences between the terms 'critical data' (or critical dataset) and 'critical data



element'. Both the terms are highly subjective and loosely used as synonyms of each other. It is important for any firm implementing data governance framework to clearly define and differentiate between the two.

A Critical Data is the data that has impactful business implications for an organization. Those implications may come from multiple aspects such as:

1. Financial (profitability, revenue, debt, costs, taxations, leverage etc.)
2. Legal & Risk – (local & international laws, compliance, regulations etc.)
3. Stakeholders / reputation (stock performance, suppliers & vendor relations, customer experience, media etc.)
4. Key Projects & Business Processes (regulatory reforms, change management projects, standard BAU processes etc.)

While Critical Data Element (CDE) was originally proposed by BCBS in its data governance guidelines for banks; the term 'critical dataset' became more prevalent with the onset of legally enforced personal data protection regulation in EU, GDPR (General Data Protection Regulation).

As per GPDR Article 4, the personal data (elements) must be collected under similar categories and then treated as critical dataset with similar privacy standards and controls implemented on top of each. GPPR recommends management and protection of critical datasets as a whole instead of focusing on individual CDEs.


Critical data elements, on the other hand, are the attributes of data that represent one or more aspects of business. For example, *customer name* as a data element may comprise multiple fields such as first, last & middle name – possibly captured in 3 different columns of a table but treated as one CDE i.e. *Customer Name*.

Grouping of CDE → Dataset/Data Concept



A notion somewhat similar to *dataset* adopted in the data governance space is ‘data concepts’. It stands for group of CDEs that belong to similar business dimensions. Example, regulatory reporting as a data concept will contain all CDEs that go into preparation and generation of returns submitted to supervisory institutions.

Some businesses prefer to go one step further and define ‘business terms’ closely associated with the critical data elements. It serves dual advantage, firstly, it helps defining a uniform business taxonomy across the enterprise and secondly, it also acts as a key business term for the business to be aware of because of its association with one or more critical data elements.



CDE FEATURES

- **Approved Definition & Other Characteristics**
- **Grouped together in Critical Datasets/Data Concepts**
- **Atomic with No Duplicates**
- **Have Owners, Stewards & Custodians Assigned**
- **Subjected to Data Quality Rules**
- **Controlled & Secured with Policies & Guidelines**
- **Have Life-cycle Captured with Lineage Diagram**
- **Defined at Enterprise or Business Function/Group level**
- **Key Inputs for Major Business Processes/Projects/Reports**



WHY IDENTIFY CRITICAL DATA ELEMENTS

Reduce Complexity

When it comes to data management, all the businesses are complex in nature. Complex because of their data architecture (with or without schema), size of data attributes they manage, the number of systems involved as well as the siloed data operations taking place simultaneously throughout the organization.

Organizations can massively reduce the complexity involved by identifying a short and manageable list of critical data elements that they can keep track of.

Managing Data Quality

Ensuring data quality of hundreds & thousands of data attributes is by no means an easily achievable task. Should you really be doing that? Till the dawn of 21st century, businesses operated under the impression that they are responsible for the quality of each and every data attribute that they own.

Next came the phase where businesses tried focusing on limited number of data elements – possibly following the Pareto Principle (formulated in 1895 & commonly known as 80 : 20 rule).

As per the rule, firms started focusing on ‘vital few’ (20%) data while ignoring ‘trivial many’ (80%). Yet, even 20% of the data was found to be humongous enough to effectively implement a data quality framework and be assured that their data (overall) is good enough for the executives to sleep peacefully at night.

Post GFC and with the introduction of data governance guidelines (BCBS 239), came the concept of identifying critical data elements that stretched beyond 80:20 rule (to



possibly 99.9 : 0.1). The recommendations pertaining to shortlisting ‘most’ critical data elements helped put methodology & explainable logic behind shortlisting certain data elements as key or critical that require special attention to keep data risks in check.

One of the major advantages of identifying critical data elements is the focused efforts that can be put in carrying out cleansing of these elements with the help of data quality framework. The board members can stay assured that with the right treatment applied upon critical data elements – their data is safe, secure and fit-for-purpose.

Makes Lineage Creation Easy

If only the task of creating a data lineage was as easy as its definition. The difficulty involved in creating lineage is found to be directly proportional to the total number of elements that go into it.

The application of lineage diagram extends beyond merely showcasing the end-to-end flow of critical elements. Lineage creation becomes even more difficult with the modern flow & usages of data such as xml & subscription based data services, two-way (canon) communications, data-as-a-service depiction and even capturing of complex transformations stretching beyond simple data massaging.

Considering all these factors – no organization plans on creating lineage on data that are non-critical and hence, having a list of shortlisted CDE is a must for complying with regulator-pushed ‘lineage creation obligation’ as well as deriving value out of it in a long run.

Easily Implement Controls to Ensure Data Integrity & Fitness-For-Purpose

Be it the privacy data protection of individuals as prescribed in GDPR or data controls & data risk management requirements mandated as part of data governance



framework; ensuring data integrity with the help of controls & security measures is binding to organizations. Achieving such incredibly important task on a relatively smaller dataset is much easier as compared to implementing blanket control policies on entire organization data.

The Multiplier Effect of CDE Cleansing

'No data operates in exclusivity', hence, many CDEs are found linked with other non-CDEs in a way that improvement in quality of one CDE automatically leads to improvement in other (correlated) data elements as well.

For instance, knowing that the quality of your CDE is up to mark and still getting incorrect results automatically suggests the issues with underlying non-CDEs thus helps reducing trial & errors and quickly narrowing down to the data records that require attention.

Assign ownerships / data stewards

This is possibly one of the biggest advantage of having a list of critical data elements. You can assign (human) ownership against critical data where data owners assume the responsibility to maintain high quality of their data with the support from data steward and proper controls implementation on critical data with engaged data custodians.

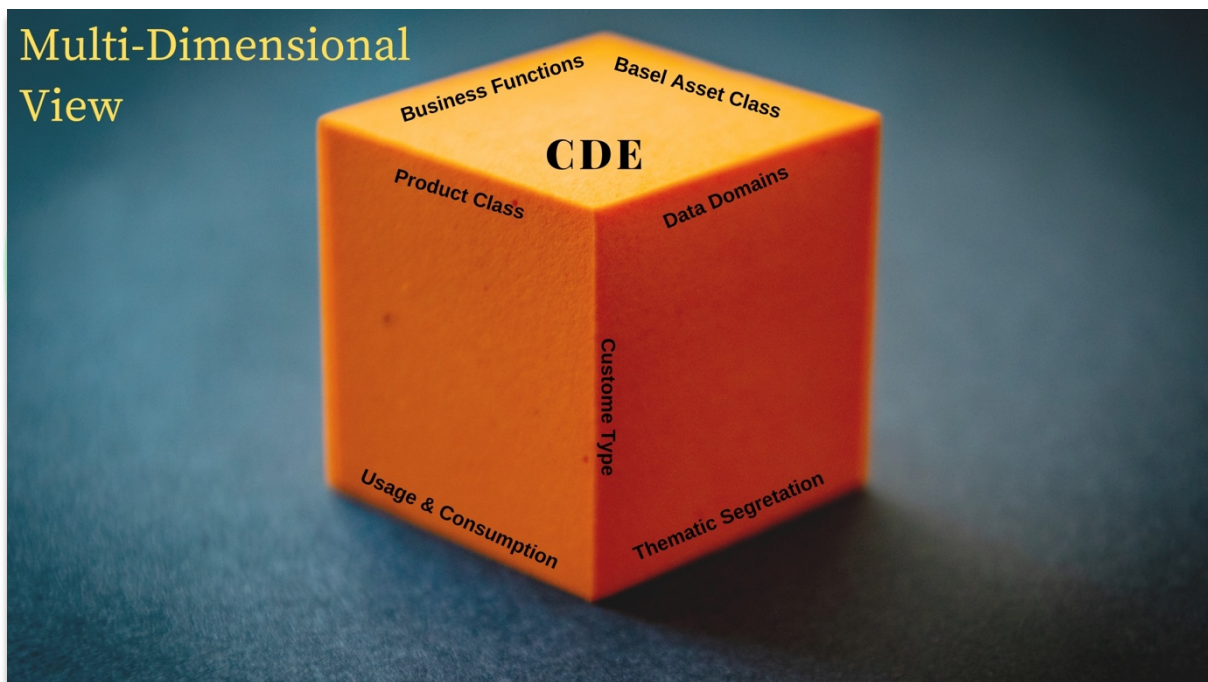
Multidimensional View of Critical Data

Critical data created on the basis of its usage helps looking into data from multiple business dimensions. For instance, same critical dataset can be segregated into multiple subsets (data concepts) on the basis of:

1. Business functions (human resources, operations, sales, marketing)



2. Basel asset class (retail, corporate, banks, SME, specialized lending, sovereign)
3. Data domain and sub-domain (customer, employee, transaction, vendor, product)
4. Product class (loans, trade financing, treasury)
5. Customer type (individual, non-individual, trade counterparties, foreign, local)
6. Thematic segregation (credit risk, operational risk, market risk, deposit & liabilities)
7. Usage & consumption (regulatory reporting, analytical, data modeling, key projects/processes)



Tailormade policies to govern critical data

Policies surrounding preservation, protection and treatment of data can be easily defined if a shortlisted critical dataset exists in the organization. Other policies & guidelines around auditing, data offshoring/outsourcing, data sharing etc. can be seamlessly implemented on critical data.



Marginal Benefits of expanding beyond CDE

Looking after CDEs is a bare-minimum, compliance mandatory yet a colossal task that must be accomplished. However, stretching the similar efforts beyond CDEs may not yield as beneficial results as they do when sticking the efforts with the CDEs.

Yes, it is advisable to stretch the program to other data elements in future when the organization is satisfied with its measures implemented on CDEs but cost-benefit analysis must also be carried-out before undertaking such task. It is better to roll-out such measures outside CDE in a phased manner by further segregating Non-CDEs into Very High, Medium and Low importance data elements.





HOW TO SHORTLIST CDE

CDE shortlisting is a sampling methodology that focuses on bare minimum data attributes but enough in numbers to represent overall health of total (data) population. The CDEs invariably form key inputs in business critical processes and reports. This sampling process not only makes it economically feasible to implement data quality rules and controls on the critical data but is also an operationally less tedious process as compared to focusing on entire data. In the upcoming sections, we will also cover the usage of statistical-applications used in refining and trimming the list of critical data elements.

Unfortunately, there are no one-size fits all methodologies for identifying critical data and there are no text-books recommended by policy makers that a firm can follow and arrive at the list. However, such lack of exhaustive guidance also offers firms immense freedom and flexibility in devising their own ways of shortlisting data elements and even adopting tailor-made approach that fits with the type of data they own. It also not uncommon for businesses to look and follow the best practices adopted by their peers or firms outside their industries.

While there neither are any penalties for shortlisting 'wrong data' nor are there any accolades if you were right the first time. Policy makers are absolutely clear on how they want this process to be followed by entities under their supervision i.e. the firm as the owner of data, must fully know its data to arrive at CDEs and it is not the "regulators' job" to tell them what their CDEs are. If your firm turns out to be wrong with some of the shortlisted CDEs where you left out the key ones and included not so important ones – then repeat the process unless you are satisfied with the list in terms of its size and ability to represent adequate data situation in the firm with nothing else left to be desired.

'Trust only those who tried', so on-boarding data experts from outside is not uncommon for organizations to learn and implement the art & science behind creating CDE list. Having advisors on your side serves multifold purpose; it enhances



confidence levels of data governance team in their critical data endeavors making them believe that they're going in right direction under experts' guidance. Plus, it also offers the firms a peek (through their advisors) into how others in their industry are approaching and achieving success in their critical data undertakings.

Another subtle advantage of having external advisors in your team is the added ease of implementing change in the organization. On-boarding external advisory team exhibits high levels of commitment and involvement from top management. Which further helps in accelerating the pace of reforms as their human talent tends to let-go of their resistance towards change or stop shying away from additional data responsibilities simply to become part of 'data movement' or more so because of their FOMO (fear of missing out) feeling.

The Phased approach

It is always recommended for firms to adopt phased approach in shortlisting their critical data elements. Depending upon size & complexity of data and the firm itself, there are simpler voting & approval ways that can be followed for identifying critical data and if the data is large with huge amount of stakeholders then it is advisable to further undertake rigorous statistical process for pruning of the CDE list prepared from simple approach. Either way, shortlisting critical data is not a one-time exercise but more of a reiterative process that may repeat itself rapidly in initial phase with the frequency slowing down over a period of time before eventually getting stabilized and thereby, reviewed periodically (at most annually) to ensure its currency and relevance with changing business dynamics.

Below are the phases a firm can adopt to arrive at their own critical data list.

Phase 1: Stop the Bleeding

Neither the identification of critical business areas & reports is new to firms nor is the process of identifying & careful handling of data that goes into them. That is, most



firms already know the major data issues they are facing and the implications of not fixing them immediately. They know what data should be subjected to extensive controls, privacy & integrity measures and be assigned to specific people for ownership.

This marks a good beginning of a firm's journey to *critical data land*. Start where you are bleeding the most and plug the flow. Identify the regulatory reports along with legal & financial data intensive processes and key projects that are glaringly important for the firm's compliant and efficient functioning. Make a list of data elements that need immediate attention.

At this stage, focusing on quantity over quality is an entirely forgivable approach. All your key data that go into regulatory reports must already be covered by the end of this exercise, possibly capturing more than required numbers of CDEs which can further be refined & trimmed by following funnel methods explained in Phase 2.

Let the experts comprising business SME and data users provide you adequate feedback and help in narrowing down to the data elements that require instant attention. Be it cleansing, controls implementation or fixing of issues under regulatory oversight – build a reform process and focus on the initially prepared CDE list for results.

Phase 1 serves multiple benefits: as your team works on the first draft of critical data, they will soon mature in their understanding and techniques of carefully crafting ways of dealing with CDEs which in turn will build a brilliant use-case for other functions/business groups to follow. Hence, this is a key step towards building an organization-wide movement surrounding critical data.

Once the team working on critical data has gained momentum & confidence in their tasks during Phase 1, it's time to mature and refine the processes surrounding critical data identification and actions thereof.



Phase 2 comprises of simple approach as the first stage of creating an exhaustive CDE list which can further be trimmed down using statistical analysis that acts as a *funnel step* in arriving at manageable number of CDEs.

Phase 2: Towards CDE Maturity

Stage 1: Simple Approach

This step largely depends upon the understanding of business and data users to identify the elements that are critical. It is a well suited approach for firms that have lesser number of stakeholders with simpler processes & data usages. Considering the cost-benefit mix, this is an ideal approach for smaller firms implementing data governance framework but doing so on relatively tighter budgets.

The process largely comprises of identifying the elements by rating them against multiple business aspects:

1. Financial (profitability, revenue, debt, costs, taxations, leverage etc.)
2. Legal & Risk – (local & international laws, compliance & regulations)
3. Stakeholders / reputation (stock performance, suppliers & vendor relations, customer experience, news & media)
4. Key Projects & Business Processes (regulatory reforms, change management projects, standard BAU processes)

The ratings can be assigned either in the form of weight-age or simple voting depending upon impact of each data element on each of the aforementioned criterions. This approach relies heavily on deep understanding of stakeholders about their data, its value & criticality in terms of usage & compliance. The stakeholders onboarded in the process usually comprise of business & data users, report owners, data officers and so-on.



Although, it is a majorly voting process, yet the final decision on shortlisted CDEs can be subjected to approval of chief or group/function data officer who in-turn must take into account the majority of votes/weight-age received by each data element.

One of the benefits apart from ease of implementation of this approach is the pace at which changes can be brought in and how quickly the resources can be onboarded to create a list.

Simple approach definitely is not free of drawback, one of them being the excessive dependency on individual stakeholder's understanding of data, its criticality and also their ability to look at the bigger picture (data governance framework) to appreciate the vitality of the critical data shortlisting process. Hence, this approach certainly leaves out a lot to be desired, nonetheless, is an effective one to achieve compliance in a short period of time.





PHASED APPROACH OF SHORTLISTING CDE



STOP THE BLEEDING



SIMPLE APPROACH



FUNNEL/STATISTICAL APPROACH



SHORTLISTED CDE





Stage 2: Quantitative Approach

One of the issues with simpler approach is the amount of subjectivity involved i.e. an element critical to one voter/rater may not be as critical to others. This ‘democratic’ process therefore, may not eventually produce best of the CDE list. Adopting quantitative methods certainly addresses this area as it introduces objectivity and facts in the decision making.

The quantitative approach recommends identifying the criteria against which criticality of data elements can be assessed and then forming a matrix to assign numerical ratings to each CDE against every criterions. Note that, some criterions can be more crucial to a firm than others so assigning ratings to each criterion helps differentiate the levels of importance of criteria.

Assigning a fixed rating scale to both criterions as well as data elements helps in standardizing inputs and removing ambiguities. The final result of this exercise is to arrive at sum-product of ratings assigned which assists in shortlisting only those CDEs which crossed a pre-set threshold.

Critical Data Element Rationalization						
Ranking Criteria (Weights)	10	7	7	7	10	
Proposed CDE/Criteria	Criterion 1	Criterion 2	Criterion 3	Criterion 4	Criterion 5	Total
1 CDE1	7	7	7	10	10	338
2 CDE2	10	10	10	10	10	410
3 CDE3	10	4	10	4	10	326
4 CDE4	10	4	4	7	4	245
5 CDE5	7	7	4	4	4	215

Source: (Jugulum, 2014)

It is important to highlight that the matrix formation also has some element of subjective inputs coming in from those responsible for assigning weights and ratings. However, this gets rectified in the forthcoming steps that refine and shortlist the list further with objective and statistical operations.



The Funnel Approach

Correlation & Regression Analysis

This refinement method aims at further trimming down the CDE list to bare-minimum by employing two sequential statistical steps on the already shortlisted CDEs.

In first step, identifying correlation (extent to which two elements are related to each other and move in same/opposite direction) helps in identifying which pairs of two critical elements are correlated high enough such that only by assessing one of them can the predictions about other element be made. For example, if CDE 'm' lies within acceptable normal range then it can be predicted that its highly correlated CDE 'n' will also lie in normal range. Or if 'm' and 'n' are highly correlated, then one can assume with very high confidence that quality defect in 'm' will lead to defects in 'n' as well.

Hence, identifying highly correlated pair of CDEs helps in segregating them aside for further shortlisting while also ensuring that all the negatively or non-correlated CDEs remain in the finalized shortlisted CDE list.

Correlation helps identifying the degree of relativity between any given pair but does not tell us which CDE (out of the two) should be dropped for achieving goal of further shortening the list. This is where regression analysis comes into picture. Regression analysis tells us the degree of change induced in one highly correlated CDE when the other CDE changes by a unit.

Strictly speaking in data quality terms, degree to which quality of one element improves vis-à-vis the improvement in quality of another (highly correlated) element.

Regression analysis is particularly a useful insight when the two CDEs are a common input to certain critical usage (report or business process); in such scenario – knowing the degree of impact that one CDE will have over another can assist in representing only one of the CDE as a critical input and drop other from the CDE list. The end result



of regression analysis is the functional derivation of one CDE denoted by its highly correlated peer CDE. Thus, regression analysis offers us the vital ability to assess the impact on business by only focusing on very limited number of the input CDEs. Note that regression analysis only bears desired results when performed on the already shortlisted pairs of exceptionally high-correlated CDEs.

One of the advantages of applying phased approach recommended in this paper is that businesses can shortlist their CDEs using both qualitative & quantitative approaches thus reducing subjectivity from the process. The approach is also found to be easily implementable as it breaks-down the process into achievable tasks with minimal disruption in the existing processes & BAU tasks of stakeholders.





CONCLUSION

In a nutshell:

1. Not all data are created equal – it is the usage that makes data critical.
2. Data Criticality is *hierarchical and subjective* in nature. That is, critical data can be defined at enterprise level or at comprising business group level or even at business function level. There may be overlaps between the critical data at multiple hierarchies or usage points. However, what is critical at enterprise level may not be critical at underlying function or group level and vice-versa. Therefore, a single business could be managing multiple list of CDEs depending upon the aforementioned factors.
3. Logical grouping of critical data element helps businesses keep multidimensional view of their critical data. The grouping is commonly known as data concept or (critical) dataset
4. Identifying and shortlisting critical data elements serves multiple benefits for an organization:
 - a. Reduces complexity in managing data
 - b. Easier to implement data quality rules and framework as a whole
 - c. Easier to create data lineage on critical data (to also fulfil compliance requirement in certain industries)
 - d. Controls & Security measures can be implemented seamlessly to ensure data integrity and fitness-for-use
 - e. Appropriate privacy classification helps protecting data from compromising usage
 - f. Helps craft effective policies surrounding critical data
5. Shortlisting Critical Data Elements is an iterative process that must begin with focus on areas requiring immediate attention. The process can further be



expanded to enlist more critical data from multiple key processes and regulatory reporting requirements.

- a. Phase 1: Stop the bleeding
- b. Phase 2: Further enhance list using *simple approach* and lastly trim this list by employing statistical process involving correlation and regression.





REFERENCES

(Jugulum, 2014) - *“Competing with High Quality Data: Concepts, Tools, and Techniques for Building a Successful Approach to Data Quality”*

Albright, S. C., W. L. Winston, and C. J. Zappe. 2009. *Data Analysis and Decision Making*. Revised 3rd ed. Mason, OH: South-Western Cengage Learning.

Batini, C., and M. Scannapieco. 2006. *Data Quality: Concepts, Methodologies and Techniques*. Berlin: Springer.

Chiang, F., and M. J. Renee. 2008. “Discovering Data Quality Rules.” -Proceedings of the 34th International Conference on Very Large Data Bases. Auckland, New Zealand.

(Center, 2017) -
https://www.ibm.com/support/knowledgecenter/en/SSWSR9_11.6.0/com.ibm.mdmhs.dev.party.doc/concepts/c_understandingcriticaldataelements.html

Creveling, C. M., J. L. Slutsky, and D. Antis, Jr. 2002. *Design for Six Sigma in Technology and Product Development*. Upper Saddle River, NJ: -Prentice Hall.

Dasgupta, Somesh. 1993. “The Evolution of the D2-Statistic of -Mahalanobis.” *Sankhya* 55:442–459.

Davenport, Thomas H., and Jeanne G. Harris. 2007. *Competing on -Analytics: The New Science of Winning*. Boston: Harvard Business School Publishing.

(Informatica, What is Critical Data, n.d.) - <https://www.informatica.com/au/services-and-training/glossary-of-terms/critical-data-definition.html#fbid=9igZ-z-0vK3>



(DAMA),2007)<http://www.damany.com/images/meeting/112008/datacontrolsanddqslas.pdf>

Data Elements or Business Elements - That Is the Question Gareth Isaac -
<https://www.datajigsaw.com/articles/dm/DataElementorBusinessElementQuestion>
(Informatica, What is Critical Data, n.d.)

<https://blog.masterdata.co.za/2018/06/12/approaches-for-selecting-critical-data-elements/>

<https://www.colibra.com/blog/4-data-governance-best-practices>

<https://www.ibmbigdatahub.com/blog/governance-over-critical-data-elements>

<https://university.colibra.com/courses/top-down-or-bottom-up-identifying-critical-data/>

REGCENTRIC



Contacts

To learn more, contact:

Thomas Verlaet

Founder | Principal Consultant

thomas.verlaet@regcentric.com

Robert Philipsz

Principal Consultant

robert.philipsz@regcentric.com

Sanjeev Kaushik

Senior Consultant

Sanjeev.kaushik@regcentric.com

