

विश्वास के लिए अनुपालन: केंद्रीय बैंकों के लिए एक डेटा गुणवत्ता मॉडल

देबाशीष नंदी और सुजीश कुमार[^] द्वारा

एक जटिल डेटा पारिस्थितिकी तंत्र में, डेटा की गुणवत्ता सुनिश्चित करना केंद्रीय बैंकों के लिए, विशेषरूप से विनियामकीय परिदृश्य में, चुनौतीपूर्ण हो जाता है। चूंकि डेटा की गुणवत्ता को मापना प्रासंगिक और व्यक्तिपरक है, इसलिए इस बात पर जोर दिया जाता है कि डेटा की गुणवत्ता को प्रभावी ढंग से मापने के लिए विभिन्न डेटा गुणवत्ता आयामों को विकसित करने के लिए अनुरूप माप रणनीतियों की आवश्यकता है। यह आलेख विनियमित संस्थाओं द्वारा प्रस्तुत डेटा की गुणवत्ता का मूल्यांकन करने के लिए डेटा गुणवत्ता सूचकांक (डीक्यूआई) बनाने के लिए एक दृष्टिकोण प्रदान करता है। आलेख विभिन्न स्तरों पर डेटा गुणवत्ता सूचकांक के निर्माण के लिए एक चरणबद्ध दृष्टिकोण की रूपरेखा तैयार करता है। प्रस्तावित डीक्यूआई संरचना केंद्रीय बैंकों और विनियामकों को संस्थागत विश्वसनीयता, नियामकीय और पर्यवेक्षी दक्षता और सार्वजनिक विश्वास को व्यवस्थित रूप से बढ़ाते हुए डेटा गुणवत्ता की निगरानी और सुधार करने में सक्षम बनाता है।

परिचय

केंद्रीय बैंक, अन्य बातों के साथ-साथ, वित्तीय स्थिरता बनाए रखने और बैंकिंग क्षेत्र की सुदृढ़ स्थिति को सुनिश्चित करने में महत्वपूर्ण भूमिका निभाते हैं। बैंकिंग प्रणाली के दैनिक संचालन से बड़ी मात्रा में डेटा का उत्पादन होता है। इस तरह का डेटा केंद्रीय बैंकों, बहुपक्षीय निकायों और कई अन्य संगठनों जैसे संस्थानों के लिए एक महत्वपूर्ण संपत्ति है, जिन्हें विशेष रूप से डेटा संग्रह और डेटा संचालित नीति निर्माण का समर्थन करने के लिए इसके रखरखाव का काम सौंपा गया है। डेटा प्रबंधन का अंतिम उद्देश्य गुणवत्ता वाले डेटा का उत्पादन और प्रसार है - जो वर्तमान विश्व में एक मूल्यवान संपत्ति है।

केंद्रीय बैंकों के कामकाज के एक महत्वपूर्ण हिस्से में विनियमित संस्थाओं (आरई) से बैंकिंग और वित्तीय क्षेत्र के आंकड़ों

[^] लेखक सांख्यिकी और सूचना प्रबंधन विभाग, भारतीय रिजर्व बैंक से हैं। इस आलेख में व्यक्त किए गए विचार लेखकों के हैं और भारतीय रिजर्व बैंक के विचारों का प्रतिनिधित्व नहीं करते हैं।

की बड़ी मात्रा का संग्रह शामिल है। एक जटिल और गतिशील डेटा पारिस्थितिकी तंत्र में, डेटा गुणवत्ता को मापना भी चुनौतीपूर्ण है। केंद्रीय बैंक के डेटा रिपॉजिटरी में एकत्रित, संसाधित और बनाए रखे जा रहे डेटा की गुणवत्ता सुनिश्चित करने में स्वचालन की महत्वपूर्ण भूमिका है। यह विभिन्न आर्थिक संकेतकों की प्रभावी ढंग से निगरानी करने, नियामकीय और पर्यवेक्षी अंतर्दृष्टि प्राप्त करने और नागरिकों के हित के लिए डेटा संचालित नीतियों को सुविधाजनक बनाने में सक्षम होगा।

पिछले कुछ वर्षों में, पूरे विश्व में तेजी से देखी गई तकनीकी प्रगति के कारण डेटा संग्रह और प्रसार प्रक्रिया में कई परिवर्तनकारी बदलाव हुए हैं। संगठनों, विशेष रूप से केंद्रीय बैंकों ने डेटा निर्माण प्रक्रिया को मान्य करने के लिए उन्नत सांख्यिकीय तकनीकों और प्रौद्योगिकी उपकरणों को अपनाया है और उभरते डेटा अंतराल और चुनौतियों का समाधान करने के लिए सर्वोत्तम प्रयास किए गए हैं। चुनौतियों का समाधान करते हुए, गैर-पारंपरिक डेटा की एक विकसित विशाल डेटा, पारिस्थितिकी तंत्र में और अधिक जटिलता जोड़ रही है। गिरार्ड (2020) ने कहा कि इस (एआई) युग में डेटा के प्रबंधन के लिए संगठनात्मक चुनौतियों में से एक कर्मचारियों को नवीनतम उपकरणों और प्रौद्योगिकियों से लैस करना है।

डेटा की गुणवत्ता को मापना प्रकृति में व्यक्तिपरक है क्योंकि माप में विभिन्न तकनीकें शामिल होती हैं और यह उत्पादित डेटा के प्रकार पर निर्भर करता है। उपयुक्त आयामों के साथ डेटा गुणवत्ता ढांचे की एक अच्छी तरह से परिभाषित संरचना डेटा की गुणवत्ता को मापने का एक उपयुक्त तरीका है (वैन जीबी 2023)। इस तथ्य से प्रेरित होकर, डेटा गुणवत्ता के विभिन्न आयामों को मापने और डेटा निर्माण और संग्रह प्रक्रिया के लिए डेटा गुणवत्ता सूचकांक प्राप्त करने के लिए एक संरचनात्मक दृष्टिकोण का प्रयास किया गया है। इस पृष्ठभूमि के साथ, इस आलेख का प्राथमिक उद्देश्य साहित्य में परिभाषित विभिन्न डेटा गुणवत्ता आयामों पर विचार करते हुए डेटा गुणवत्ता को मापने के लिए एक दृष्टिकोण प्रदान करना है। विभिन्न संगठनों द्वारा प्रयुक्त मौजूदा डेटा गुणवत्ता संरचना, डेटा गुणवत्ता आयामों के विभिन्न मापकों की गणना के लिए विशिष्ट सूत्र प्रदान नहीं करता है, जबकि यह विभिन्न गुणवत्ता आयामों का आकलन करने के लिए व्यापक दिशानिर्देश

निर्दिष्ट करता है। इसके अलावा, उक्त संरचनाएँ डेटा संग्रह प्रक्रिया और प्रसार प्रक्रिया के लिए अलग-अलग डेटा गुणवत्ता सूचकांक (डीक्यूआई) माप का सुझाव नहीं देते हैं।

भारतीय संदर्भ में, भारतीय रिज़र्व बैंक (आरबीआई) ने हाल ही में पर्यवेक्षित संस्थाओं के लिए पर्यवेक्षी डेटा गुणवत्ता सूचकांक (एसडीक्यूआई) स्कोर प्रकाशित किया है, जो चार डेटा गुणवत्ता आयामों के आधार पर है, सटीकता, पूर्णता, सामयिकता और स्थिरता। एसडीक्यूआई पर्यवेक्षी डेटा गुणवत्ता का एक माप प्रदान करता है, जो पर्यवेक्षी परीक्षणों (आरबीआई, 2025) के लिए आधार बनाता है। एसडीक्यूआई का उद्देश्य पर्यवेक्षी डेटा संग्रह परिप्रेक्ष्य में चुनिंदा पर्यवेक्षी विवरणियों की डेटा गुणवत्ता को मापना है।

केंद्रीय बैंकिंग के संबंध में विनियमित संस्थाओं द्वारा प्रदान किए गए आंकड़ों का उपयोग न केवल पर्यवेक्षी उद्देश्य के लिए किया जाता है, जबकि इसका उपयोग नियामकीय, नीति निर्माण, सांख्यिकीय डेटा प्रसार, अनुसंधान और विभिन्न अन्य उद्देश्यों के लिए किया जा रहा है। इस आलेख में बताया गया दृष्टिकोण पर्यवेक्षी आंकड़ों तक सीमित नहीं है; इसमें निर्धारित विवरणी के माध्यम से विनियामक संस्थाओं से एकत्र किए गए सभी प्रकार के आंकड़े शामिल हैं। इसके अलावा, इस आलेख में प्रस्तुत डीक्यूआई चार आंकड़ा गुणवत्ता आयामों से भी आगे जाता है, जिसमें आंकड़ा संग्रह और प्रसार पहलुओं को शामिल किया गया है। इस प्रकार यह आलेख इस पहलू में मौजूदा अंतराल को पाटने में मदद करता है।

आलेख के शेष भाग को पाँच खंडों में विभाजित किया गया है। अगला खंड साहित्य की संक्षिप्त समीक्षा प्रस्तुत करता है। खंड III में विभिन्न डेटा गुणवत्ता आयामों का विवरण दिया गया है, जबकि खंड IV खंड III में चर्चा किए गए कई आयामों का उपयोग करके डेटा गुणवत्ता सूचकांक निर्माण की विधि की रूपरेखा तैयार करता है। अंत में, खंड V आलेख का सारांश प्रस्तुत किया गया है।

2. साहित्य की समीक्षा

डेटा गुणवत्ता को साहित्य जगत में अलग-अलग तरीके से परिभाषित किया गया है। डेटा गुणवत्ता वह स्थिति है जो डेटा उपयोगकर्ताओं की जरूरतों को पूरा करता है (वांग, 1998)। वांग और स्ट्रॉन्ग (1996) द्वारा डेटा गुणवत्ता की व्यापक रूप से स्वीकृत परिभाषाओं में से एक 'उपयोग के लिए उपयुक्त' है। जबकि यह तर्क दिया गया था स्ट्रॉन्ग एवं अन्य (1997), कि उपयोग

के लिए उपयोग अलग-अलग परिस्थितियों में अलग-अलग उपयोगकर्ताओं के लिए अलग-अलग होती है और इसलिए डेटा की गुणवत्ता सापेक्ष होती है और इसका मूल्यांकन उपयोगकर्ताओं से स्वतंत्र होकर नहीं किया जा सकता है। सांख्यिकीय पद्धति पर संघीय समिति (एफसीएसएम) ने डेटा गुणवत्ता को उस डिग्री के रूप में परिभाषित किया है जिस तक डेटा उचित पद्धति का उपयोग करके वांछित जानकारी को इस तरह से ग्रहण करता है जो जनता के विश्वास को बनाए रखता है। डेटा गुणवत्ता के महत्व पर कई लेखकों द्वारा प्रकाश डाला गया था। डेटा की खराब गुणवत्ता गलत निष्कर्ष और घटिया निर्णय लेने की ओर ले जाती है जिससे वित्तीय नुकसान होता है। यह त्रुटिपूर्ण जोखिम आकलन के परिणामस्वरूप हो सकता है और संगठनात्मक प्रदर्शन को नकारात्मक रूप से प्रभावित कर सकता है, डेटा अभिशासन रणनीतियों की मांग कर सकता है (रेडमैन, 2008; खार्ती एंड ब्रोम, 2010; ली एवं अन्य ., 2004)

शोधकर्ताओं ने डेटा की गुणवत्ता को मापने में कई चुनौतियों की ओर इशारा किया है। चूंकि डेटा समय के साथ बदलता है, अतः आज 'उच्च-गुणवत्ता' के रूप में वर्गीकृत डेटा भविष्य में इसी श्रेणी का नहीं रह सकता है। डेटा की इस गतिशील प्रकृति के लिए निरंतर निगरानी और लगातार पुनर्मूल्यांकन की आवश्यकता होती है (बाटिनी एवं अन्य, 2009)। बिग डेटा वातावरण में, डेटा की गुणवत्ता को मापना तब और जटिल तथा थकाऊ हो जाता है, जब डेटा की मात्रा बहुत अधिक हो। डेटा गुणवत्ता के सभी आयामों को ट्रैक करना अधिक कठिन होगा (मुलर एवं अन्य, 2012)। एक गतिशील डेटा वातावरण में, आयामों का चयन प्रासंगिक और उपयुक्त हो सकता है। फदाहुन्सी एवं अन्य (2009) द्वारा विकसित एक ढांचे ने इन चुनौतियों का कुछ हद तक समाधान किया। उदाहरण के लिए, सूचना गुणवत्ता ढांचा आयामों को आंतरिक, प्रासंगिक, प्रतिनिधित्वात्मक और पहुंच पहलुओं में वर्गीकृत करता है जो डेटा गुणवत्ता के मूल्यांकन के लिए एक व्यापक दृष्टिकोण प्रदान करता है। इसके अलावा, किसी विशेष उद्देश्य के लिए 'उपयोग की उपयुक्तता' हो या नहीं, डेटा स्थिरता तब एक बड़ा मुद्दा बन जाता है, जब डेटा स्रोत की संख्या बढ़ जाती है। इसलिए, एक मानकीकृत दृष्टिकोण तब उपयुक्त होता है जब लोगों के विभिन्न डोमेन के बीच डेटा गुणवत्ता के बारे में असहमति होती है। डेटा अभिशासन संरचना के तहत गुणवत्ता प्रक्रिया को लाना डेटा गुणवत्ता सुनिश्चित करने के लिए एक समाधान होगा (स्मॉलवुड, 2014)।

विभिन्न गुणवत्ता मैट्रिक्स के चयन में व्यक्तिपरकता भी डेटा गुणवत्ता को मापने में एक और चुनौती है। ये व्यक्तिपरक उपाय उपयोगकर्ताओं के बीच भिन्न हो सकते हैं, जिससे गुणवत्ता आकलन में विसंगतियां हो सकती हैं (रेडमैन, 2008)। उच्च गुणवत्ता वाला डेटा किसी भी सूचना प्रणाली के लिए एक मूलभूत आवश्यकता है, जबकि गलत डेटा संगठनों को सुधार गतिविधियों, ग्राहकों का दूर जाना, अवसरों को अनदेखा करना और गलत निर्णयों के रूप में बहुत अधिक नुकसानदायक होता है। गलत डेटा प्रविष्टि जैसे मुद्दे डेटा की सटीकता को कम कर सकते हैं। यह गलत जानकारी का कारण भी बन सकता है जब डेटा को अनुचित रूप से रिपोर्ट या उपयोग किया जाता है (ओल्सन, 2003)। गैरेड्ट एवं अन्य, (2014) ने प्रबंधन और वित्तीय रिपोर्टिंग पहलुओं जैसे प्रोद्घवन गुणवत्ता, गलत बयान और आंतरिक नियंत्रण गुणवत्ता में संगठनात्मक विश्वास के बीच संबंध की जांच की। उन्होंने पाया कि विश्वास वित्तीय रिपोर्टिंग गुणवत्ता के साथ महत्वपूर्ण रूप से जुड़ा हुआ है और विकेंद्रीकृत फर्मों में अपेक्षाकृत भिन्न होता है, जबकि केंद्रीकृत डेटा वातावरण में नहीं।

डेटा गुणवत्ता को मापना डेटा गुणवत्ता प्रबंधन प्रक्रिया में एक महत्वपूर्ण कदम है। साहित्य में डेटा गुणवत्ता आयामों को मापने के लिए विभिन्न दृष्टिकोणों पर चर्चा की गई है, जबकि गुणवत्ता आयामों पर विचार किया गया है जो साहित्य में सामान्य प्रतीत होते हैं। ये दृष्टिकोण मुख्य रूप से मात्रात्मक हैं, जहां डेटा की गुणवत्ता को सांख्यिकीय तकनीकों या सरल अंकगणितीय अनुपात के आधार पर मापा जाता है या कुछ मॉडलिंग दृष्टिकोण का उपयोग करके जो विभिन्न मात्रात्मक अनुपात और उपायों पर विचार करते हैं (रहम एवं डू, 2000 और चंडोला एवं अन्य., 2009)। हालांकि, कुछ गुणवत्ता आयाम प्रणाली और प्रक्रियाओं से सीधे मापने योग्य नहीं हैं। इस तरह के आयामों को सर्वेक्षण विधियों का उपयोग करके गुणात्मक रूप से मापा जाता है (झांग एवं अन्य., 2005; लेमायर एवं अन्य., 2009)। कुछ लेखकों ने मुख्य रूप से गुणवत्ता तुलना के लिए बेंचमार्किंग विधियों का सुझाव दिया है, जिसमें प्रमुख संगठनों के बेंचमार्क के सापेक्ष डेटा गुणवत्ता का मूल्यांकन किया जाता है। यह संगठन को यह समझने में सक्षम बनाता है कि उनके समकक्षों की तुलना में उनकी डेटा गुणवत्ता कहां है (बाटिनी एवं अन्य., 2009)।

एक तीसरा दृष्टिकोण एक हाइब्रिड दृष्टिकोण होगा, जो मात्रात्मक और गुणात्मक दोनों तरीकों का संयोजन है। यह दृष्टिकोण मानता है कि कोई भी एक विधि डेटा गुणवत्ता माप की सभी जटिलताओं को पूरी तरह से दूर नहीं कर सकती है, विशेष रूप से एक विषम और गतिशील डेटा वातावरण में। हाइब्रिड मॉडल अनुकूलनीय और लोचनीय होते हैं, और वे उपयोगकर्ताओं से चल रही प्रतिक्रिया के साथ वास्तविक समय डेटा गुणवत्ता निगरानी को आसानी से एकीकृत कर सकते हैं।

आईएमएफ, विश्व बैंक और कई अन्य बहुराष्ट्रीय संस्थानों जैसे संगठनों द्वारा कई डेटा गुणवत्ता ढांचे का प्रस्ताव किया गया है। इस तरह के ढांचे अक्सर कई आयामों को एकीकृत करते हैं और कई पहलुओं में डेटा गुणवत्ता का मूल्यांकन करने के लिए एक संरचित दृष्टिकोण प्रदान करते हैं। आईएमएफ द्वारा प्रस्तुत किए गए एक उल्लेखनीय ढांचे को डेटा गुणवत्ता मूल्यांकन फ्रेमवर्क (डीक्यूएफ) के रूप में जाना जाता है, जो डेटा गुणवत्ता के मूल्यांकन के लिए एक व्यापक मॉडल प्रदान करता है। इसमें सटीकता, विश्वसनीयता और समयबद्धता जैसे आयाम शामिल हैं, और इसे विशेष रूप से सरकार और अंतर्राष्ट्रीय संगठनों (आईएमएफ, 2003) में सांख्यिकीय डेटा का आकलन करने के लिए डिज़ाइन किया गया है। वांग और स्ट्रॉंग (1996) द्वारा प्रस्तावित एक अन्य मौलिक ढांचे को सूचना गुणवत्ता फ्रेमवर्क (आईक्यूएफ) के रूप में जाना जाता है, जो मुख्य रूप से इस बात पर जोर देता है कि डेटा गुणवत्ता एक बहुआयामी अवधारणा है जिसमें सटीकता और स्थिरता जैसे तकनीकी पहलू और डेटा गुणवत्ता की उपयोगकर्ता धारणाएं दोनों शामिल हैं। उपयुक्त गुणवत्ता आयामों के आधार पर डेटा गुणवत्ता सूचकांक (डीक्यूआई) की गणना करना एक विशिष्ट दृष्टिकोण है, जिसका उपयोग स्वास्थ्य सेवा, उद्योग और वित्त जैसे विभिन्न क्षेत्रों में किया गया है, जो डेटा गुणवत्ता के लिए एकल समग्र उपाय की सुविधा प्रदान करता है। इस तरह के समग्र उपाय संगठनों को समय के साथ डेटा गुणवत्ता में परिवर्तन को ट्रैक करने में सक्षम बनाता है।

3. डेटा गुणवत्ता आयाम

साहित्य में कई गुणवत्ता आयाम परिभाषित हैं। विभिन्न विषयों में डेटा गुणवत्ता आयामों पर एक हालिया और व्यापक सर्वेक्षण कार्वाल्होआ, एवं अन्य, (2025) द्वारा करवाया गया था।

उन्होंने सर्वेक्षण किया और अपने शोध में लगभग सभी गुणवत्ता आयामों को सूचीबद्ध किया, क्योंकि डेटा गुणवत्ता आयामों के निर्धारण पर कोई सहमति नहीं है। उन्होंने लगभग 66 गुणवत्ता आयामों की पहचान की, जिससे उपयोगकर्ताओं को उपयुक्त रूप से आयामों का चयन करने और डेटा गुणवत्ता ढांचे के विकास की अनुमति मिली। इस क्षेत्र में अनुसंधान से पता चलता है कि विभिन्न आयाम अलग-अलग दृष्टिकोण प्रदान करते हैं या डेटा के उद्देश्य के आधार पर अलग-अलग निर्भरता रखते हैं। केंद्रीय बैंकों और बहुपक्षीय संस्थानों ने अपने डेटा गुणवत्ता ढांचे में लगभग 14 गुणवत्ता आयाम निर्दिष्ट किए हैं। विभिन्न देशों के विभिन्न संगठनों द्वारा आमतौर पर उपयोग किए जाने वाले कुछ डेटा गुणवत्ता आयाम सारणी -1 में सूचीबद्ध किए गए हैं।

'उपयोग के लिए उपयुक्त' डेटा गुणवत्ता की एक व्यापक परिभाषा है जो आवश्यक डेटा के उद्देश्य, जरूरतों और प्राथमिकताओं तथा उपयोगकर्ता के दृष्टिकोण पर निर्भर करती है। ये आवश्यकताएं उपयोगकर्ताओं के समूह में अलग-अलग हो सकती हैं। भले ही डेटा सटीक है, किन्तु यदि वह उपयोगी होने के लिए बहुत देर से निर्मित है, या आसानी से प्राप्त नहीं किया जा

सकता है, या अन्य डेटा के साथ मेल नहीं खाता है, तो जरूरी नहीं है कि वह अच्छी गुणवत्ता का हो। इस प्रकार, गुणवत्ता को एक बहुआयामी अवधारणा (एनरिको और वार्ड, 2004) के रूप में देखा जाता है। इसलिए, संगठनों ने अपने डेटा गुणवत्ता आयामों को परिभाषित किया है जो डेटा संग्रह और प्रसार आवश्यकताओं तथा उपयोग पर निर्भर करता है। गुणवत्ता आयामों की संख्या भी एकत्र या प्रसारित डेटा की प्रकृति और प्रकार के अनुसार भिन्न हो सकती है। ये गुणवत्ता मानदंड या आयाम गुणवत्ता परिभाषा और मूल्यांकन के लिए एक समावेशी दृष्टिकोण को दर्शाते हैं।

डेटा गुणवत्ता ढांचे और साहित्य में समीक्षा किए गए गुणवत्ता आयामों के आधार पर, इस आलेख में आठ डेटा गुणवत्ता आयाम प्रस्तुत किए गए हैं जो व्यापक हैं और डेटा गुणवत्ता के मुद्दों का बड़े पैमाने पर ध्यान रखते हैं, जिसमें डेटा का संग्रह और प्रसार शामिल है। केंद्रीय बैंक या नियामक आमतौर पर आर्थिक और वित्तीय डेटा को इकट्ठा करने, विश्लेषण करने और निगरानी करने के स्पष्ट उद्देश्य के साथ एक डेटा संग्रह प्रारूप तैयार करते हैं। वे प्रासंगिक स्रोतों और चरों की उचित पहचान के साथ डेटा संग्रह के उद्देश्य और लक्ष्यों को भी ध्यान में रखते हैं। डेटा प्रारूपों का

सारणी 1: डेटा गुणवत्ता के अलग-अलग आयाम

क्रम संख्या	गुणवत्ता के आयाम	बैंक ऑफ इंग्लैंड	यूरोपीय केंद्रीय बैंक#	ओईसीडी	ऑस्ट्रेलियाई सांख्यिकी ब्यूरो	जर्मनी का संघीय सांख्यिकी कार्यालय	कनाडा सरकार
1	समय की पाबंदी और समयबद्धता	✓	✓	✓	✓	✓	✓
2	शुद्धता	✓	✓	✓	✓	✓	✓
3	विश्वसनीयता			✓			
4	पहुँच/स्पष्टता	✓		✓	✓	✓	✓
5	स्थिरता						✓
6	विवेचनीयता			✓	✓		✓
7	प्रासंगिकता	✓			✓	✓	✓
8	जुटना	✓		✓	✓		✓
9	संपूर्णता		✓				✓
10	स्थिरता		✓				
11	दिखावट		✓				
12	भरोसा		✓				✓
13	तुलनात्मकता	✓				✓	
14	लागत क्षमता*			✓			

टिप्पणियाँ: i. कुछ संगठन अपने डेटा गुणवत्ता फ्रेमवर्क में समय की पाबंदी और समयबद्धता का एक साथ इस्तेमाल करते हैं और कुछ अलग-अलग इस्तेमाल करते हैं।

ii. * ओईसीडी लागत-दक्षता को गुणवत्ता का एक आयाम नहीं मानता है, जबकि यह गुणवत्ता के किसी भी विश्लेषण में ध्यान में रखा जाने वाला एक कारक है क्योंकि यह सभी आयामों में गुणवत्ता पर असर डाल सकता है।

iii. # विनियामकीय डेटा गुणवत्ता फ्रेमवर्क से जुड़े आयाम।

स्रोत: लेखकों द्वारा अलग-अलग संगठनों की वेबसाइट से एकत्र किया गया।

नामकरण अक्सर अलग-अलग देशों में भिन्न होता है। उदाहरण के लिए, यूरोपीय केंद्रीय बैंक ने अपने डेटा प्रारूप को 'रिपोर्टिंग टेम्पलेट' या 'डेटा टेम्पलेट' नाम दिया। फेडरल रिज़र्व 'कॉल रिपोर्ट' शब्द का उपयोग करता है, और बैंक ऑफ इंग्लैंड का डेटा प्रारूप 'सांख्यिकीय रिटर्न' है। रिज़र्व बैंक ऑफ ऑस्ट्रेलिया आर्थिक और वित्तीय डेटा एकत्र करने के लिए 'सांख्यिकीय फॉर्मर्स' का उपयोग करता है जिसका उपयोग विभिन्न सांख्यिकीय विज्ञप्तियों और सारणियों को तैयार करने के लिए किया जाता है। भारतीय संदर्भ में, रिज़र्व बैंक आरई के लिए सांख्यिकीय या नियामक डेटा एकत्र करने के लिए 'रिटर्न' नामक एक टेम्पलेट का उपयोग करता है। सुविधा के लिए, इस पूरे आलेख में 'रिटर्न' शब्दावली का उपयोग किया गया है। इस पृष्ठभूमि के साथ, निम्नलिखित गुणवत्ता आयाम डेटा गुणवत्ता का एक समावेशी मूल्यांकन प्रदान करते हैं।

3.1 समयबद्धता

समयबद्धता आयाम को कभी-कभी समय की पाबंदी के साथ एक दूसरे के स्थान पर उपयोग किया जाता है, या दोनों को 'समय की पाबंदी और समयबद्धता' के रूप में एक साथ उपयोग किया जाता है। किसी भी मामले में, डेटा गुणवत्ता का महत्वपूर्ण पहलू डेटा की समय पर उपलब्धता है। समयबद्धता आयाम मुख्य रूप से इस बात का मूल्यांकन करता है कि संगठन द्वारा एकत्र किए जाने वाले आंकड़े निर्धारित समय-सीमा के अनुसार समय पर प्राप्त हुए हैं या नहीं। आरई द्वारा डेटा प्रस्तुत करने की समय सीमा का पालन करना महत्वपूर्ण है, क्योंकि डेटा की समय पर उपलब्धता महत्वपूर्ण है – विशेषरूप से जब कोई विशेष डेटा सेट किसी अन्य डेटा से संबंधित हो, या इसे डेटा के किसी अन्य सेट के साथ देखा जाना हो।

कभी-कभी समयबद्धता को संदर्भित किया जाता है कि डेटा कितना अद्यतित है या डेटा का निर्माण या रिपोर्ट किए जाने पर डेटा कितना नवीन है - डेटा की प्रासंगिकता से जुड़ा हुआ है। दोनों दृष्टिकोणों का उपयोग यह आकलन करने के लिए किया जाता है कि अपेक्षित समय पर डेटा प्रदान किया गया है या रिपोर्ट किया गया है। आमतौर पर, इस उपाय को एकत्र किए जा रहे कुल डेटा के सापेक्ष डेटा की समय पर उपलब्धता को ध्यान में रखते हुए अनुपात के रूप में व्यक्त किया जाता है। संगठन आमतौर पर रिटर्न जमा करने के लिए समयसीमा निर्धारित करते हैं। रिटर्न किसी भी रूप का हो सकता है - पर्यवेक्षी, नियामकीय या सांख्यिकीय।

एक मानक डेटा गुणवत्ता ढांचे में, समयबद्धता आयाम को मापने के लिए निम्न सूत्र का उपयोग किया जा सकता है:

$$\text{समीकरण (T)} = \frac{n_r}{n_d + n_r} \times 100 \quad (1)$$

जहां, n_r निर्धारित समय में प्रस्तुत रिटर्न की संख्या है और n_d देरी से अर्थात् निर्धारित समय के बाद, प्रस्तुत रिटर्न की संख्या है। डेटा गुणवत्ता सूचकांक संगणन में इस प्रतिशत माप को सही तरीके से भाहित किया जाना चाहिए।

3.2 सटीकता

डेटा गुणवत्ता का सटीकता आयाम आम तौर पर आरई द्वारा प्रस्तुत डेटा की शुद्धता या सटीकता से मापा जाता है। डेटा फाइलिंग प्रक्रिया में, डेटा की सटीकता संस्थाओं द्वारा दायर किए गए समग्र डेटा की गुणवत्ता निर्धारित करती है। सटीकता वास्तविक डेटा को दर्शाती है जिसमें वांछनीय विशेषताएं होनी चाहिए जैसे कि त्रुटियों और विचलन से मुक्त होना, सही मूल्य से निकटता और उच्च परिशुद्धता। हालांकि, इसे मापना मुश्किल है, क्योंकि इसे सैद्धांतिक रूप से अनुमानित मूल्यों और सही (अज्ञात) मूल्यों के बीच अंतर के रूप में परिभाषित किया गया है। डेटा संशोधन सटीकता का एक अच्छा मूल्यांकन दे सकते हैं क्योंकि वे यह निर्धारित करने के लिए एक तंत्र प्रदान करते हैं कि अपने 'अंतिम' मूल्य (ओईसीडी, 2003) तक पहुंचते हैं तो समय के साथ अनुमान कैसे बदलते हैं। यह दृष्टिकोण आरई द्वारा प्रस्तुत डेटा की सटीकता को कैप्चर करने के लिए विशेष रूप से उपयुक्त है, क्योंकि डेटा रिपोर्टिंग में संशोधन आम हैं, खासकर बैंकिंग या वित्तीय क्षेत्र में।

संशोधन की सीमा डेटा की गुणवत्ता को निर्धारित करती है - चाहे परिवर्तन न्यूनतम हो या पर्याप्त। यदि परिवर्तन उसी डेटा की प्रारंभिक प्रस्तुति से महत्वपूर्ण है, तो यह निश्चित रूप से एक गुणवत्ता का मुद्दा है। यह निर्धारित करना भी महत्वपूर्ण है कि क्या परिवर्तन वास्तविक है या डेटा त्रुटि के कारण है। इसके अलावा, डेटा को फिर से भरना या फिर से जमा करना आवश्यक रूप से वास्तविक संशोधनों के कारण नहीं है। सत्यापन विफलता कभी-कभी डेटा प्रस्तुति में विफलता का कारण बन सकती है, जिसके लिए आरई को पुनः प्रस्तुत करने की आवश्यकता होती है।

डेटा गुणवत्ता ढांचे में, सटीकता आयाम को एक प्रमुख संकेतक के आधार पर रिपोर्टिंग अवधि के दौरान डेटा को संशोधित करने की संख्या और संशोधन के परिमाण की जांच करनी चाहिए। एक संशोधन को एक ही कुंजी आइटम के बाद और पहले के अनुमान के बीच के अंतर के रूप में परिभाषित किया गया है। इन पहलुओं को ध्यान में रखते हुए, सटीकता आयाम के लिए एक सूत्र को निम्नानुसार परिभाषित किया जा सकता है:

मान लीजिए कि 'u' किसी विशेष रिपोर्टिंग इकाई को किसी दिए गए रिपोर्टिंग अवधि में एक विशिष्ट रिटर्न को फिर से सबमिट करने की कुल संख्या को दर्शाता है, 'v' किसी विशेष रिपोर्टिंग इकाई के लिए किसी विशेष रिटर्न के लिए सत्यापन विफलताओं की संख्या को दर्शाता है, और 'w' सत्यापन विफलताओं के कारण नहीं होने वाले पुनः सबमिशन की संख्या को दर्शाता है, जैसे कि $u = v + w$ ।

सापेक्ष माध्य पूर्ण संशोधन (आरएमएआर) की गणना नीचे दिए गए सूत्र का उपयोग करके सभी पुनः सबमिशन u (सत्यापन विफलता के कारण पुनः सबमिशन, v और सत्यापन विफलता के अलावा, w) के लिए की जाती है:

$$RMAR_u = \frac{\sum_u |Z_f - Z_i|}{\sum_u |Z_f|} \quad (2)$$

जहां, Z_f पहली प्रस्तुति में रिपोर्ट किया गया मान है और Z_i अंत मान है (की एग्रीगेट Z के लिए प्रस्तुत अंतिम मान)

$$\text{Then, Accuracy (Ac)} = \begin{cases} 0 & \text{if } RMAR_u > 1 \\ 100 - (RMAR_u \times 100) & \text{otherwise} \end{cases} \quad (3)$$

3.3 विश्वसनीयता

विश्वसनीयता संपूर्ण डेटा निर्माण और रिटर्न जमा करने की प्रक्रिया की विश्वसनीयता की डिग्री को मापती है। यह आकलन करता है कि क्या सभी डेटा मैनुअल हस्तक्षेप के बिना स्वचालित तरीके से तैयार किए गए हैं। आरई द्वारा प्रदान किए गए डेटा की विश्वसनीयता मुख्य रूप से डेटा प्रक्रिया के तीन पहलुओं पर निर्भर करती है।

- ग्रेनूलर स्तर के डाटा कैचरिंग तंत्र में स्वचालन की सीमा,
- विनियामकीय आवश्यकताओं को पूरा करने के लिए आंकड़ों के एकत्रीकरण और गणना प्रक्रिया का स्वचालन; तथा
- डाटा पारिषण प्रक्रिया का स्वचालन।

आम तौर पर, नियामकों का तीसरी प्रक्रिया पर नियंत्रण होता है, क्योंकि डेटा सबमिशन चैनल नियामक द्वारा प्रदान किए जाते हैं। हालांकि, पहली दो प्रक्रियाएं अक्सर नियामक को दिखाई नहीं देती हैं। उनका आकलन करने के लिए, डेटा ऑडिटर संस्थाओं का दौरा करते हैं या सर्वेक्षणों के माध्यम से जानकारी का अनुरोध करते हैं।

ग्रेनूलर डेटा को ऑनलाइन ट्रांजैक्शन प्रसंस्करण प्रणाली (ओएलटीपी) या अन्य स्वचालित प्रणालियों जैसे कोर बैंकिंग सिस्टम (सीबीएस), ट्रेजरी ऑपरेशंस सिस्टम (टीओएस), आदि के माध्यम से प्राप्त किया जाता है जो आरई के डेटा वेयरहाउस (डीडब्ल्यू) से जुड़े होते हैं। डेटा एकत्रीकरण या रिटर्न जनरेशन प्रक्रिया या तो डीडब्ल्यू में या प्रबंधन सूचना प्रणाली (एमआईएस) के माध्यम से होती है, जिसमें डेटा निकालने के लिए व्यावसायिक तर्क के साथ विभिन्न प्रोग्रामों का उपयोग किया जाता है। एकत्रीकरण का एक हिस्सा कभी-कभी पूर्वनिर्धारित डेटा टेम्पलेट्स में डेटा को विभेदित करके मैनुअल रूप से किया जाता है। डेटा संग्रह तंत्र की विश्वसनीयता बढ़ाने के लिए इन प्रक्रियाओं के स्वचालित तरीके से होने की उम्मीद है। डेटा प्रक्रिया का तीसरा स्तर डेटा ट्रांसमिशन स्तर है जहां रिटर्न प्रस्तुत करने के लिए विभिन्न चैनलों का उपयोग किया जा रहा है। इनमें सिस्टम-टू-सिस्टम चैनल, फाइल अपलोड चैनल, एप्लिकेशन प्रोग्रामिंग इंटरफेस (एपीआई) आधारित चैनल और वेब आधारित या स्क्रीन-आधारित सबमिशन चैनल शामिल हैं। इनमें से, सिस्टम-टू-सिस्टम और एपीआई आधारित चैनल डेटा जमा करने के लिए सबसे विश्वसनीय साधन प्रदान करने की पूरी तरह से स्वचालित डेटा सबमिशन प्रक्रिया सुनिश्चित करते हैं। विश्वसनीयता का माप प्रकृति में गुणात्मक है और ऊपर वर्णित रिटर्न प्रस्तुत करने की प्रक्रिया के लिए आरई को दिए गए अंकों के आधार पर प्राप्त किया जाता है। विश्वसनीयता को मापने के लिए सुझाया गया एक स्कोरिंग मैट्रिक्स सारणी -2 में दिया गया है।

डीजीपी और जीएपी के लिए रिटर्न वार स्कोर आरई से प्राप्त किए जा सकते हैं जबकि डीटीपी स्कोर नियामकों द्वारा प्रदान की गई डेटा सबमिशन प्रणाली से प्राप्त किया जा सकता है। अंत में, विश्वसनीयता आयाम निर्धारित करने के लिए एक भारत औसत स्कोर प्राप्त किया जा सकता है।

3.4 सुसंगतता

डेटा गुणवत्ता का स्थिरता आयाम व्यवसाय सत्यापन सहित विभिन्न सत्यापन नियमों के उल्लंघन की जांच करता है। डेटा

सारणी 2: विश्वसनीयता स्कोरिंग मैट्रिक्स

स्वचालन का स्तर

श्रेणी	< 30 प्रतिशत	30-50 प्रतिशत	50-80 प्रतिशत	> 80 प्रतिशत
डेटा निर्माण प्रक्रिया (डीजीपी)	30	50	80	100
डेटा एकत्रीकरण प्रक्रिया (डीएपी)	30	50	80	100
डेटा प्रस्तुतीकरण के माध्यम				
डेटा आदान-प्रदान प्रक्रिया (डीटीपी)	प्रणाली से प्रणाली	एपीआई	फ़ाइल अपलोड	वेब आधारित/अन्य
स्कोर	100	100	80	60

टिप्पणी: i. यह आवश्यक नहीं है कि स्कोर निश्चित हो और ये संगठन द्वारा स्वचालन सेट के महत्व/लेवल के हिसाब से अलग-अलग हो सकते हैं।

ii. स्वचालन का प्रतिशत प्रत्येक प्रक्रिया (डीजीपी और डीएपी) में स्वचालित रिटर्न की संख्या के आधार पर प्राप्त होना चाहिए।

आइटम डेटा फ़ाइल में संबंधपरक या स्थिर हो सकते हैं (बाटिनी और स्कैनापिका, 2006)। एक डेटा टेम्पलेट आमतौर पर प्रकृति में संबंधपरक¹ होता है, जिसका अर्थ है कि कई संख्याएं/कोशिकाएं आपस में जुड़ी हुई हैं और इसमें गणना शामिल है। स्थिरता आयाम यह जांचता है कि प्रारूप में दिखाई देने वाला डेटा तार्किक और अंकगणितीय संचालन का अनुसरण करता है या नहीं और क्या अपेक्षित डेटा बिंदु कई शीट या रिटर्न में रिपोर्ट किया गया है। इन्हें अखंडता बाधा कहा जाता है, जो ऐसे गुण हैं जिन्हें डेटाबेस स्कीमा के सभी उदाहरणों से संतुष्ट होना चाहिए।

डेटा प्रस्तुतीकरण प्रक्रिया में-जब एक ही डेटा बिंदु या डेटा तत्व को अलग-अलग रिटर्न में प्रस्तुत करने की आवश्यकता होती है, और डेटा एक ही रिपोर्टिंग अवधि से संबंधित होता है, तो यह उम्मीद की जाती है कि सभी रिटर्न में एक ही मूल्य की सूचना दी जाती है। यहां रिपोर्ट किया गया डेटा बिंदु रिटर्न में सुसंगत हो सकता है।

मान लीजिए कि c_e उन डेटापॉइंट्स की संख्या है जो कई रिटर्न में रिपोर्ट किए जाते हैं। यदि c_t ऐसे डेटा बिंदुओं की संख्या (c_e में से) है जो डेटा प्रस्तुत करने के दौरान रिपोर्ट किए जाते हैं जो रिटर्न से मेल नहीं खाते हैं।

तब रिटर्न सुसंगतता (C_o) निम्नानुसार प्राप्त की जा सकती है:

$$C_o = \frac{c_e}{c_t} \times 100 \quad (4)$$

¹ भले ही डेटा संबंधपरक न हो, स्थिरता नियमों को परिभाषित किया जा सकता है। उदाहरण के लिए, एक प्रभावली प्रारूप के मामले में, शब्दार्थ नियमों को संबंधपरक बाधाओं के समान तरीके से परिभाषित किया जाता है (एटजेनी, और डी एंटोनेलिस, 1993; बाटिनी और स्कैनापिका, 2006)।

3.5 पूर्णता

पूर्णता डेटा गुणवत्ता का एक गुणात्मक माप है जो यह बताता है कि दी गई जानकारी/दिशानिर्देशों/परिभाषाओं आदि का उपयोग करके डेटा मान किस हद तक पर्याप्त रूप से भरे हुए हैं। डेटा पॉप्युलेटिंग प्रक्रिया में डेटा एकत्रीकरण शामिल होता है जिसमें अंकगणितीय या तार्किक गणना शामिल होती है। मुख्य रूप से, आरई एक स्वचालित प्रक्रिया के माध्यम से या कुछ मैनुअल हस्तक्षेप के साथ आवश्यक डेटा को पॉप्युलेट करता है। चूंकि आरई तकनीकी वातावरण के विभिन्न स्तरों पर काम करते हैं, इसलिए रिटर्न निर्माण के लिए उचित मार्गदर्शन उनके लिए बहुत आवश्यक है ताकि उनकी रिटर्न निर्माण गतिविधियों को सुव्यवस्थित किया जा सके। आम तौर पर, नियामक परिपत्रों और प्रेस विज्ञप्तियों के माध्यम से आवश्यक दिशानिर्देश, डेटा परिभाषाएं, संकलन मैनुअल, नियामक परिवर्तनों और डेटा आवश्यकताओं में परिवर्तन आदि पर अपडेट प्रदान करता है। आरई से यह भी अपेक्षा की जाती है कि वे ऐसे दस्तावेजों को बनाए रखें, और विभिन्न रिटर्न पर उन्हें प्रदान की गई जानकारी को ट्रैक करें। प्रत्येकरिटर्न निर्माण प्रक्रिया के लिए एक संकलन मैनुअल या प्रक्रियात्मक दस्तावेज बनाए रखने का सुझाव दिया जाता है। यह दस्तावेज आरई के लिए एक व्यावसायिक निरंतरता दस्तावेज के रूप में भी कार्य करता है। इन सभी पहलुओं और आरई में अपेक्षित दस्तावेजों की उपलब्धता को ध्यान में रखते हुए, आरई को उचित अंक प्रदान करते हुए, पूर्णता आयाम का एक गुणात्मक माप विकसित किया जा सकता है।

डेटा अंतराल, लापता अवलोकन, गणना त्रुटियों आदि पर विचार करते हुए पूर्णता को मात्रात्मक रूप से मापा जा सकता है। यहां यह उस हद तक संदर्भित करता है जिस तक उपयोगकर्ता

लापता टेम्पलेट्स और लापता मानों के बिना सभी डेटा प्राप्त करते हैं और डेटा संबंधित मेटाडेटा के साथ होता है। इसमें डेटासेट और अतिरिक्त जानकारी दोनों शामिल हैं जो उपयोगकर्ताओं को उनके विशिष्ट संदर्भों में डेटासेट को समझने में मदद करती है। गुणात्मक पूर्णता आयाम डेटा प्रसार प्रक्रिया की गुणवत्ता के मूल्यांकन के लिए उपयुक्त है।

दोनों प्रकार के उपाय पूर्णता के विभिन्न पहलुओं का आकलन करते हैं, और वे इस बात की अधिक समग्र समझ प्रदान करने के लिए काम करते हैं कि डेटा कैसे पूर्ण या अधूरा है। मात्रात्मक उपायों के मामले में, पूर्णता की कुछ विशेषताओं को देखना चाहिए खाली रिकॉर्ड, विशेषता पूर्णता और इकाई पूर्णता। भारित पूर्णता प्राप्त करने के लिए पूर्णता की प्रत्येक विशेषता को उचित भार प्रदान किया जा सकता है। पूर्णता के प्रत्येक गुणवत्ता पहलुओं से संबंधित तकनीकी स्कोर का मूल्यांकन संगठन में डेटा लेखा परीक्षकों द्वारा किया जा सकता है।

3.6 प्रासंगिकता

डेटा गुणवत्ता का एक अन्य महत्वपूर्ण आयाम प्रासंगिकता आयाम है, जो उस डिग्री को संदर्भित करता है जिस पर डेटा उपयुक्त, उपयोगी है, और एक विशिष्ट उद्देश्य के लिए इसकी प्रयोज्यता है। यदि केंद्रीय बैंक द्वारा उत्पादित या प्रसारित डेटा इच्छित उपयोगकर्ताओं के लिए प्रासंगिक नहीं है, तो यह नीति निर्माण और विश्लेषण का प्रभावी ढंग से समर्थन नहीं कर सकता है। इस आयाम का उपयोग संगठन द्वारा प्रसारित डेटा की गुणवत्ता का मूल्यांकन करने के लिए किया जाता है। डेटा की प्रासंगिकता इस बात पर निर्भर करती है कि क्या यह उन उपयोगकर्ताओं को उपयोगी अंतर्दृष्टि प्रदान करता है जो अपनी वांछित स्तर की जानकारी प्राप्त करना चाहते हैं। प्रासंगिकता आयाम अक्सर गुणात्मक होता है, और यह मूल्यांकन करता है कि डेटा उपयोगकर्ताओं या हितधारकों की आवश्यकताओं को कितनी अच्छी तरह पूरा करता है। प्रसारित डेटा को संदर्भ के साथ संरेखित करना चाहिए- जिसका अर्थ है कि डेटा को विश्लेषण के डोमेन और उद्देश्य से संबंधित होना चाहिए। निर्णय लेने के लिए डेटा अप-टू-डेट, प्रयोग करने योग्य और कार्रवाई योग्य होना चाहिए। गुणवत्ता के इस आयाम का उपयोग करते हुए, डेटा प्रबंधक या लेखा परीक्षक एकत्र किए जा रहे डेटा का गुणात्मक मूल्यांकन कर सकते हैं और जांच सकते हैं कि स्थिति

और उपयोगकर्ता की जरूरतों के आधार पर डेटा प्रासंगिक है या नहीं। नियमित डेटा उपयोगकर्ता सर्वेक्षण और बातचीत डेटा प्रबंधकों या लेखा परीक्षकों को इनपुट प्रदान करेगी जो उचित स्कोरिंग मैट्रिक्स का निर्माण करके इस आयाम के लिए स्कोर प्रदान कर सकते हैं।

3.7 स्थिरता

डेटा गुणवत्ता का स्थिरता आयाम इंगित करता है कि डेटा समय के साथ कैसे सुसंगत और विश्वसनीय बना रहता है। यह डेटा की क्षमता को दर्शाता है, विभिन्न समय अवधि में इसकी अखंडता और उपयोगिता को बनाए रखता है, यह सुनिश्चित करता है कि डेटासेट में परिवर्तन को इसकी गुणवत्ता को प्रभावित किए बिना ट्रैक और नियंत्रित किया जाता है। स्थिरता आयाम को या तो गुणात्मक रूप से मापा जाता है या संदर्भ और डेटा के प्रकार के आधार पर मात्रात्मक मैट्रिक्स का उपयोग करके मापा जाता है। गुणात्मक उपाय उपयोगकर्ता प्रतिक्रिया या विशेषज्ञ आकलन के आधार पर किए जा सकते हैं। उदाहरण के लिए, प्रयोज्यता, पता लगाने की क्षमता², और डेटा ने वास्तविक अनुप्रयोगों या डेटा के विश्लेषणात्मक अभ्यास में कितनी अच्छी तरह प्रदर्शन या उपयोग किया है, इसका आकलन किया जा सकता है। यदि डेटा उपयोगकर्ता लगातार पाते हैं कि डेटा विश्वसनीय और सुसंगत है, तो डेटा को स्थिर माना जा सकता है।

एक मात्रात्मक पहलू में, डेटा बहाव, स्थिरता अनुपात और परिवर्तन दर जैसे उपायों का उपयोग डेटा के स्थिरता आयाम का आकलन करने के लिए किया जा सकता है। इन सरल उपायों में अक्सर सूत्रों का उपयोग करके संख्यात्मक गणना शामिल होती है। उदाहरण के लिए, कुल्लबैक-लीबलर विचलन या जेन्सेन-शैनन विचलन (सीज़ार, आई 1975; नीलसन, एफ. 2021) का उपयोग करते हुए समय के साथ डेटा में परिवर्तन का संकेत देने वाले डेटा बहाव का मापना इसी तरह, डेटा स्थिरता अनुपात समय के साथ लगातार डेटा बिंदुओं का अनुपात प्रदान करता है। एक वैकल्पिक उपाय परिवर्तन दर होगी जिसे एक अवधि के दौरान डेटा परिवर्तनों की संख्या और उसी अवधि के लिए कुल डेटा बिंदु के अनुपात से मापा जाता है। यह उपाय उपयोगकर्ता को बताएगा कि समय के साथ डेटा कितनी जल्दी बदलता है।

² ट्रेसबिलिटी का अर्थ है समय श्रृंखला डेटा की उपलब्धता जिसका अर्थ है डेटा के इतिहास को उसके मूल से लेकर वर्तमान अवधि तक ट्रैक करने की क्षमता।

3.8 पहुँच

पहुँच डेटा गुणवत्ता का एक और महत्वपूर्ण आयाम है जो उस आसानी को संदर्भित करता है जिसके साथ डेटा तक पहुँचा जा सकता है, पुनर्प्राप्त किया जा सकता है और जरूरत पड़ने पर उपयोगकर्ताओं द्वारा उपयोग किया जा सकता है। यह उपयोगकर्ताओं के लिए मेटाडेटा उपलब्धता को भी संदर्भित करता है, जिसमें वह फॉर्म या माध्यम शामिल है जिसके माध्यम से जानकारी तक पहुँचा जाता है, डेटा सुरक्षा सुविधाएँ और अंतर-परिचालनीयता शामिल हैं। उपयोगकर्ताओं को प्रदान की जाने वाली सहायता डेटा और इसकी पहुँच के बारे में पूरी जानकारी प्राप्त करने के लिए पर्याप्त हो सकती है। यद्यपि डेटा में अन्य गुणवत्ता आयाम जैसे सटीकता, समयबद्धता और पूर्णता उच्च गुणवत्ता वाले डेटा का संकेत देते हैं, यह तब तक मूल्यवान नहीं है जब तक कि आवश्यकता पड़ने पर यह डेटा उपयोगकर्ताओं के लिए आसान तरीके से सुलभ न हो। इस आयाम को आमतौर पर गुणात्मक मानदंडों का उपयोग करके मापा जाता है - डेटा प्रबंधकों या डेटा लेखा परीक्षकों द्वारा आवधिक प्रतिक्रिया सर्वेक्षण का संचालन करना। डेटा गुणवत्ता के पहुँच आयाम तक पहुँचने के लिए डेटा उपयोगकर्ताओं की प्रतिक्रिया बहुत महत्वपूर्ण मानदंड है। डेटा के उपयोगकर्ताओं (प्रसार और संग्रह दोनों) के लिए फीडबैक सर्वेक्षण किए जा सकते हैं। यह आयाम डेटा पोर्टल पर उपयोगकर्ताओं के लिए समर्थन की उपलब्धता, पहुँच में आसानी और डेटा पोर्टल के आसपास नेविगेट करने में आसानी की भी जांच करता है।

4. डेटा गुणवत्ता सूचकांक – विधि

अनुमानित विभिन्न डेटा गुणवत्ता आयाम का उपयोग करके, कोई भी सूचकांक के रूप में डेटा गुणवत्ता के भारित माप पर पहुँच सकता है। इस तरह के सूचकांक का निर्माण केंद्रीय बैंकों को डेटा गुणवत्ता प्रगति की निगरानी करने, प्रत्येक रिटर्न/प्रकाशन के लिए सुधार के क्षेत्रों की पहचान करने और विश्वसनीय निर्णय लेने को सुनिश्चित करने में सक्षम बनाता है। गुणवत्ता आयामों को अनुमानित प्रत्येक आयाम के महत्व के अनुसार भारित किया जा सकता है। भारिता निर्धारित करने की आवश्यकता नहीं है और संगठनों द्वारा निर्धारित डेटा गुणवत्ता आयामों के महत्व के अनुसार भिन्न हो सकते हैं। डेटा संग्रह और डेटा प्रसार प्रक्रियाओं के लिए आयामों की प्रयोज्यता अलग है। संगठन डेटा संग्रह और प्रसार प्रक्रियाओं दोनों के लिए डेटा गुणवत्ता सूचकांकों के निर्माण के लिए उपयुक्त गुणवत्ता आयामों का चयन कर सकता है।

आमतौर पर, रिटर्न आरई द्वारा विभिन्न आवृत्तियों पर प्रस्तुत किए जाते हैं, अर्थात्, साप्ताहिक, पाक्षिक, मासिक, अर्ध-वार्षिक, आदि। नियामक को गणना की जाने वाली डीक्यूआई की आवृत्ति अर्थात् मासिक या त्रैमासिक तय करने की आवश्यकता है। वांछित अवधि में आने वाले सभी रिटर्न को डीक्यूआई की गणना के लिए माना जा सकता है। यदि किसी को मासिक या त्रैमासिक आधार पर डीक्यूआई की गणना करनी है, तो उस महीने या तिमाही में आने वाले सभी रिटर्न पर विचार किया जा सकता है, भले ही रिटर्न की आवृत्ति कुछ भी हो।

मान लीजिए (f_1, f_2, \dots, f_p) एक महीने में आने वाले रिटर्न/प्रकाशनों की p विभिन्न आवृत्तियों (साप्ताहिक, पाक्षिक, मासिक, त्रैमासिक, आदि) का समुच्चय है। निर्माण किए जाने वाले डेटा गुणवत्ता सूचकांक में एक महीने में p आवृत्ति के इन सभी रिटर्न को कवर किया जाना चाहिए।

मान लीजिए कि E_i संगठन को रिटर्न जमा करने वाले आरई का सेट है, $i = 1, 2, \dots, l$, R_j आरई द्वारा प्रस्तुत रिटर्न का सेट है, $j = 1, 2, \dots, m$, और D_k विचाराधीन गुणवत्ता आयामों का सेट है, $k = 1, 2, \dots, n$ । डेटा गुणवत्ता सूचकांक की गणना के लिए, डेटा विश्लेषक को ट्रिपलेट निर्धारित करना होगा: (E_i, R_j, D_k) ; $i = 1, 2, \dots, l$; $j = 1, 2, \dots, m$; $k = 1, 2, \dots, n$ ।

ट्रिपलेट (E_i, R_j, D_k) के लिए गुणवत्ता आयाम स्कोर i^{th} इकाई, j^{th} इकाई, एवं k^{th} आयाम के लिए मापा जाता है, जिसे d_{ijk} द्वारा दर्शाया जाता है। इन स्कोरों को तब इकाई-रिटर्न डेटा गुणवत्ता सूचकांक, $DQI(E_i, R_j)$ प्राप्त करने के लिए उपयुक्त भार w_k के साथ एकत्रित किया जाता है और इसे निम्नानुसार परिभाषित किया गया है:

$$DQI(E_i, R_j) = \frac{1}{N} \sum_{k=1}^n w_k d_{ijk} \quad (5)$$

जहाँ $N = \sum w_k$, प्रत्येक आयाम के लिए सही भारिता विनियामक के डेटा संग्रहण प्रक्रिया और प्रणाली तथा उसके संबन्धित महत्व पर निर्भर करता है। भारत के संदर्भ में, वर्मा और नंदी (2017) के अनुसार, सटीकता सबसे ज़रूरी डेटा गुणवत्ता आयाम (31.25%) थी, इसके बाद स्थायित्व (21.25%), समयबद्धता (20%) और पूर्णता (11.25%) थे। अध्ययन में विशिष्टता पर भी विचार किया गया, जिसकी भारिता (16.25%) थी, जो विश्वसनीयता आयाम के साथ से काफी हद तक जुड़ा हुआ है।

4.1 इकाई स्तर डीक्यूआई

इकाई स्तर डीक्यूआई को किसी विशेष रिटर्न के डेटापॉइंट/सेल की संख्या के आधार पर उचित भार के साथ रिटर्न गुणवत्ता सूचकांकों को एकत्रित करके प्राप्त किया जा सकता है³। इकाई स्तर डीक्यूआई को डीक्यूआई (Ei) द्वारा दर्शाया जाता है और इसे निम्नानुसार परिभाषित किया गया है:

$$DQI(E_i) = \frac{1}{N} \sum_{j=1}^m \alpha_j DQI(E_i R_j) \quad (6)$$

जहाँ α_j ; एक रिटर्न R के लिए एक संस्था E द्वारा प्रस्तुत डेटा पॉइंट्स/सेल्स की संख्या के आधार पर भारित हैं। ऐसी संस्थाएं जो अधिक डेटा पॉइंट्स फाइल कर रही हैं, उनका α_j में $N = \sum \alpha_j$.

4.2 रिटर्न स्तर डीक्यूआई

इकाई के समग्र व्यावसायिक प्रोफाइल को भारित करके रिटर्न स्तर डीक्यूआई प्राप्त किया जा सकता है। इकाई द्वारा किए गए कुल बैंकिंग व्यवसाय के हिस्से का उपयोग करके भारिता प्राप्त किया जा सकता है, ई समग्र बैंकिंग व्यवसाय के लिए, इकाई⁴ ई को सापेक्ष महत्व देने के लिए एक प्रमुख संकेतक है।

रिटर्न लेवल $DQI(R_j)$ की गणना इस प्रकार की जा सकती है:

$$DQI(R_j) = \frac{1}{N} \sum_{i=1}^l \beta_i DQI(E_i R_j) \quad (7)$$

जहां β_i ; संस्था के कारोबार पर आधारित भारिता है, $N = \sum \beta_i$

4.3 विनियामक के लिए डीक्यूआई का मापन

एक एंटरप्राइज़ स्तरीय डेटा गुणवत्ता सूचकांक या तो संस्था-स्तरीय डीक्यूआई या रिटर्न स्तरीय-डीक्यूआई को मिलाकर बनाया जा सकता है। इसलिए, एक एंटरप्राइज़ स्तरीय डीक्यूआई, जिसे $DQI_{EP \text{ level}}$ स्तर से दिखाया जाता है, उसे इस तरह बताया गया है:

$$DQI_{EP \text{ level}} = \frac{1}{N} \sum_{h=1}^{(l,m)} \delta_h DQI_h(E, R)_{(h)} \quad (8)$$

जहां $N = \sum \delta_h$, यह एंटरप्राइज़ स्तरीय डीक्यूआई तक पहुंचने के लिए संस्था भारिता या रिटर्न भारिता के चुनाव पर निर्भर करता है।

³ जब रिटर्न का एक ही सेट सभी आरई के लिए लागू हो। यदि रिटर्न अलग-अलग संस्थाओं के लिए अलग-अलग हैं, तो रिटर्न भारिता की गणना केवल किसी विशेष इकाई पर लागू रिटर्न के लिए की जा सकती है।

⁴ यह बैंकों के लिए पिछले वित्तीय वर्ष की कुल जमा राशि और कुल ऋण के योग से अनुमानित हो सकता है।

भले ही केंद्रीय बैंक की डेटा संग्रह प्रक्रिया एक केंद्रीकृत वातावरण में काम कर रही है, लेकिन ऐसे कई विभाग/कार्यक्षेत्र/डोमेन हैं जो डेटा के विभिन्न सेटों का ध्यान रखते हैं। उदाहरण के लिए, विदेशी मुद्रा बाजार डेटा विदेशी मुद्रा विभाग द्वारा एकत्र और प्रकाशित किया जाता है, जो विदेशी मुद्रा से संबंधित डेटा और रिटर्न का डोमेन मालिक है। इसी तरह, बैंकिंग डेटा बैंकिंग विभाग या नियामक विभाग द्वारा एकत्र और प्रसारित किया जाता है। विभाग-स्तर या डोमेन-स्तरीय डेटा गुणवत्ता सूचकांकों का अनुमान उन विवरणों को समूहीकृत करके भी लगाया जा सकता है जो किसी विभाग या डोमेन द्वारा अपनाए जाते हैं। तदनुसार, एक नियामक इकाई द्वारा किसी विभाग या वर्टिकल को प्रस्तुत किए गए डेटा पॉइंट्स/सेलों की संख्या को ध्यान में रखते हुए एक भारित औसत डीक्यूआई प्राप्त किया जा सकता है। इस तरह के विभाग-स्तरीय डेटा गुणवत्ता सूचकांक का उपयोग विभिन्न विभागों या डोमेन के बीच तुलना के लिए किया जा सकता है। यह केंद्रीय बैंकों के विभिन्न क्षेत्रों/विभागों द्वारा नियंत्रित अलग-अलग रिटर्न से संबंधित डेटा संग्रह गुणवत्ता की निगरानी को सक्षम करेगा।

इस आलेख में उल्लिखित डीक्यूआई पद्धति का पालन करते हुए, उचित आयामों और भारिता का उपयोग करके प्रत्येक सांख्यिकीय सारणियों (रिटर्न के समान) या प्रकाशन (कई सारणियों सहित) पर विचार करते हुए प्रसार गुणवत्ता सूचकांक भी प्राप्त किए जा सकते हैं। इस आलेख में दिए गए दृष्टिकोणों को अपनाना संगठनों के लिए अपने डेटा गुणवत्ता माप को संस्थागत बनाने और समग्र डेटा गुणवत्ता ढांचे को बढ़ाने और समग्र डेटा अभिशासन को बढ़ाने के लिए उपयोगी हो सकता है।

4.4 डीक्यूआई की व्याख्या

डीक्यूआई प्रत्येक आयाम के महत्व और उस आवृत्ति पर विचार करते हुए, जिसके साथ डेटा का उपयोग किया जाता है, समग्र डेटा गुणवत्ता का एक ही माप प्रदान कर सकता है। डेटा को वर्गीकृत करने के लिए डीक्यूआई के लिए थ्रेसहोल्ड होना वांछनीय है। 100 (≥ 80) के करीब एक डीक्यूआई स्कोर उत्कृष्ट डेटा गुणवत्ता का सुझाव देता है, जबकि कम स्कोर सुधार के क्षेत्रों का संकेत देता है। यदि $70 \leq DQI < 80$ है, तो डेटा गुणवत्ता अच्छी है, जबकि यदि $DQI < 70$ है, तो संगठन को अपनी डेटा

गुणवत्ता में सुधार की आवश्यकता है। समान मानदंड का उपयोग किसी भी आयाम या डीक्यूआई के किसी भी स्तर के लिए भी किया जा सकता है।

5. निष्कर्ष

यह शोधपत्र पूरे साहित्य में विभिन्न डेटा गुणवत्ता आयामों की समीक्षा करता है और प्रासंगिक और सामग्री-निर्भर डेटा गुणवत्ता आयामों और उनके अनुमान के चयन के लिए एक मजबूत और मापने योग्य ढांचा प्रदान करता है। इससे केंद्रीय बैंकों या संगठनों को अपने डेटा की गुणवत्ता की निगरानी और सुधार के लिए विभिन्न स्तरों पर उपयुक्त डेटा गुणवत्ता आयामों और डेटा गुणवत्ता सूचकांक को अपनाने और लागू करने में सुविधा होगी। भले ही, आलेख डेटा संग्रह और प्रसार प्रक्रियाओं के लिए आठ गुणवत्ता आयामों और दो अलग-अलग दृष्टिकोणों का सुझाव देता है, संगठन अपने संचालन के क्षेत्र के आधार पर किसी भी प्रक्रिया को नियोजित कर सकते हैं।

संगठनों/विभाग/अनुभाग स्तरों पर अतिरिक्त जानकारी को उचित भारिता के साथ डेटा गुणवत्ता आयामों में भी शामिल किया जा सकता है। आलेख में दिए गए वेटिंग पैटर्न संगठनों पर कड़ाई से लागू नहीं होते हैं, यह संगठनों पर छोड़ दिया जाता है कि वे अपनी प्रक्रियाओं और प्रणालियों के संबंध में स्वयं निर्णय लें।

यद्यपि डेटा गुणवत्ता प्रबंधन एक सतत प्रक्रिया है, इस आलेख में प्रदान किया गया ढांचा अन्य वित्तीय संस्थानों या डेटा-संचालित नीति निर्माताओं के लिए एक बेंचमार्क के रूप में काम कर सकता है, जो डेटा गुणवत्ता को अपनी डेटा अभिशासन रणनीतियों में एकीकृत करने का लक्ष्य रखते हैं। यह आलेख केंद्रीय बैंकों और अन्य डेटा-संचालित संगठनों के भीतर डेटा गुणवत्ता ढांचे को बढ़ाने पर जारी चर्चा में योगदान देता है।

संदर्भ

Atzeni, P. and De Antonellis, V. (1993), "Relational Database Theory", The Benjamin Publishing Company.
Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. (2009), "Methodologies for Data Quality Assessment

and Improvement", *ACM Computing Surveys*, 41(3):1-41.

Batini, C. and Scannapieca, M. (2006), "Data Quality Concepts, Methodologies and Techniques", Springer-Verlag, Berlin Heidelberg.

Carvalho, A.M., Soares, S., Montenegro, J. and Conceição, L. (2025). "Data Quality: revisiting dimensions towards new framework development", *Procedia Computer Science* 253, 247–256.

Chandola, V., Banerjee, A., and Kumar, V. (2009), "Anomaly Detection: A Survey", *ACM Computing Surveys*, 41(3):1-58.

Csiszar, I. (1975). "I-Divergence Geometry of Probability Distributions and Minimization Problems". *Annals of Probability*. 3 (1): 146–158.

Enrico, G. and Ward, D. (2004). "Quality framework for OECD statistics getting our own house in order", paper presented in the conference on data quality for international organizations, Germany, May 2004.

Fadahunsi, K. P., Akinlua, J. T., O Connor, S., Wark, P. A., and Gallagher, J. (2019), "Protocol for a systematic review and qualitative synthesis of information quality frameworks in eHealth", *BMJ Open*, 9(3).

Garreitt, J., Hoitash, R. and Prawitt, D.F. (2014), "Trust and Financial Reporting Quality" *Journal of Accounting Research*, 52 (5).

Girard, M. (2020), "Helping Organizations Master Data Governance", *Policy Brief No. 163*, Centre for International Governance Innovation.

IMF (2003), "Data quality Assessment Framework and Data Quality Program", International Monetary Fund, Washington.

Khatri, V. and Brown, C. V. (2010), "Designing data governance", *Communications of the ACM*, 53(1):148-152.

Lee, Y.W., Pipino, L., Strong, D.M., and Wang, R.Y. (2004), "Process embedded data integrity", *Journal of Database Management*, 15(1):87-103.

- Lemire, D., MacLellan, C., and Kargupta, H. (2009), "Task-Dependent Data Quality", *IEEE Transactions on Data Engineering*, 31(4):205-221.
- Muller, H. J., Rojas, R. G., and Wilke, G. (2012), "Big data analytics and the role of data quality", *Information Systems and E-Business Management*, 10(1):37-52.
- Nielsen, F. (2021). On a variational definition for the Jensen-Shannon symmetrization of distances based on the information radius. *Entropy*, 23 (4).
- OECD (2003), "Quality Framework and Guidelines for OECD Statistical Activities", OECD, Paris.
- Olson, J. E (2003), "Data Quality: The Accuracy Dimension", *The Morgan Kaufmann Series in Data Management Systems*, 3-23.
- Rahm, E., and Do, H. H. (2000), "Data Cleaning: Problems and Current Approaches", *IEEE Transactions on Knowledge and Data Engineering*, 11(4):147-162.
- RBI (2025). "Supervisory Data Quality Index for Scheduled Commercial Banks", Reserve Bank of India, Press release March 2025.
- Redman, T. C. (2008), "Data Quality: The Field Guide", Digital Press.
- Smallwood, R.F. (2014), "Information Governance: Concepts, Strategies, and Best Practices", John Wiley and Sons.
- Strong, D. M. Yang W. L, and Wang, R.Y. (1997), "Data Quality in Context", *Communications of the ACM*, 40(5).
- Van Gils, B. (2023), "Data in Context-Models as Enablers for Managing and Using Data". The Enterprise Engineering Series. Springer.
- Verma P. and Nandi, D (2017), "Data Quality of Data Warehouse: A Case Study", *International Journal of Advances in Electronics and Computer Science*, 4(9).
- Wang, R.Y (1998), "A Product Perspective on Total Data", *Communications of the ACM*, 41(2).
- Wang, R.Y. and Strong, D.M. (1996), "Beyond accuracy: What data quality means to data consumers", *Journal of Management Information System*, 12(4).
- Zhang, S., Lee, K. P., and Chen, D. (2005), "Measuring Perceived Data Quality", *Data and Knowledge Engineering*, 55(3): 289-319.