

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

مدل‌ها و روش‌های مدیریت کیفیت داده: نگرشی کاربردی بر داده‌های علم و فناوری

مؤلفان

محمدجواد ارشادی

محمد مهدی ارشادی

ویراستار

حمید کشاورز



وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران
(ایرانداک)

تهران ۱۳۹۸

سرشناسه	: ارشادی، محمدجواد، ۱۳۶۲ -
عنوان و نام پدیدآور	: مدل‌ها و روش‌های مدیریت کیفیت داده: نگرشی کاربردی بر داده‌های علم و فناوری / مؤلفان محمدجواد ارشادی، محمدمهدی ارشادی؛ ویراستار حمید کشاورز.
مشخصات نشر	: تهران: پژوهشگاه علوم و فناوری اطلاعات ایران: چاپار، ۱۳۹۸.
مشخصات ظاهری	: ۱۸۴ ص.: مصور، جدول.
شابک	: ۲۸۰۰۰ ریال 978-622-95282-2-8
وضعیت فهرست نویسی	: فیبا
یادداشت	: کتابنامه.
عنوان دیگر	: نگرشی کاربردی بر داده‌های علم و فناوری.
موضوع	: داده‌کاوی -- روش‌ها
موضوع	: Data mining -- methods
شناسه افزوده	: ارشادی، محمدمهدی، ۱۳۷۲ -
شناسه افزوده	: کشاورز، حمید، ۱۳۵۸ - ویراستار
شناسه افزوده	: پژوهشگاه علوم و فناوری اطلاعات ایران
رده بندی کنگره	: QA۷۶/۹
رده بندی دیویی	: ۰۰۶/۳۱۲
شماره کتابشناسی ملی	: ۵۶۷۸۸۷۳

این کتاب از سلسله انتشارات پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک) است که با همکاری انتشارات چاپار منتشر شده است.

مدل‌ها و روش‌های مدیریت کیفیت داده: نگرشی کاربردی بر داده‌های علم و فناوری



دانشگاه تهران
پژوهشگاه علوم و فناوری اطلاعات ایران
ایراندک

مؤلفان: محمدجواد ارشادی، محمدمهدی ارشادی

ویراستار: حمید کشاورز

© ۱۳۹۸: پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)

صفحه‌آرا: رحیم کبیرصابر

طراح جلد: سرور محمدی بزنج

نوبت چاپ: اول ۱۳۹۸ | شمارگان: ۵۰۰ نسخه

تهران - خیابان ولی‌عصر (عج) - بالاتر از زرتشت

کوچه نوربخش - پلاک ۴۰ - نشر چاپار

تلفن پخش: ۸۸۸۹۹۶۸۰ - ۰۹۱۲۱۹۶۰۲۱۴

بها: ۲۸۰.۰۰۰ ریال

فهرست مطالب

فصل اول: مفاهیم اولیه کنترل و مدیریت کیفیت	۱۳
۱-۱. مقدمه	۱۳
۲-۱. تفاوت کنترل کیفیت و بازرسی	۱۷
۳-۱. اهمیت روش‌های آماری در کنترل کیفیت	۱۸
۱-۳-۱. طوفان فکری	۱۹
۲-۳-۱. نمودار ایشی کاوا	۲۱
۳-۳-۱. روش چراهای پنج‌گانه	۲۳
۴-۳-۱. روش پارتو	۲۴
۴-۱. چرخه دمینگ و جایگاه آن در مدیریت کیفیت	۲۵
۵-۱. کیفیت داده	۲۶
فصل دوم: مبانی کیفیت داده و اطلاعات	۳۵
۱-۲. مقدمه	۳۵
۲-۲. گام‌های تجزیه و تحلیل روش‌ها	۳۷
۳-۲. مقایسه دیدگاه‌ها در تجزیه و تحلیل روش‌ها	۳۸
۱-۳-۲. مراحل و گام‌های مشترک	۳۹
۲-۳-۲. استراتژی‌ها و روش‌ها	۴۲
۳-۳-۲. معرفی ابعاد کیفیت داده	۴۵
۴-۳-۲. هزینه‌ها	۵۱
۵-۳-۲. انواع داده‌ها	۵۲
۶-۳-۲. ارتباط کیفیت داده با سیستم‌های اطلاعاتی	۵۵
۴-۲. مقایسه روش‌های مختلف کیفیت داده‌ها	۵۷
۱-۴-۲. روش‌ها، گام‌ها و مراحل	۵۹

۶۱.....	۲-۴-۲. چگونگی ارزیابی در هر مدل.....
۶۲.....	۲-۴-۳. گام بهبود.....
۶۳.....	۲-۵. روش‌ها، استراتژی‌ها و شیوه‌ها.....
۶۴.....	۲-۵-۱. شیوه‌های داده محور.....
۶۷.....	۲-۵-۲. شیوه‌های فرایند محور.....
۶۸.....	۲-۶. ارتباط مدل‌های مدیریت کیفیت با ابعاد کیفیت داده.....
۷۷.....	۲-۷. روش‌ها و هزینه‌ها.....
۷۷.....	۲-۷-۱. دسته‌بندی هزینه‌ها.....
۷۹.....	۲-۷-۲. معیارهای کمی‌سازی هزینه‌ها.....
۸۱.....	۲-۸. روش‌ها و انواع داده‌ها.....
۸۳.....	۲-۹. روش‌ها و انواع سیستم‌های اطلاعاتی.....
۸۹.....	۲-۱۰. مقایسه کوتاه روش‌ها.....
۹۸.....	۲-۱۱. از کیفیت داده به سمت کیفیت اطلاعات.....
۹۹.....	۲-۱۲. کیفیت داده‌ها و کیفیت فرایند.....
۱۰۱.....	۲-۱۳. کیفیت داده‌ها و انواع سیستم‌های اطلاعاتی جدید.....
۱۰۵.....	فصل سوم: مدل‌های مدیریت کیفیت داده.....
۱۰۵.....	۳-۱. مقدمه.....
۱۰۵.....	۳-۲. روش TDQM (مدیریت جامع کیفیت داده).....
۱۰۹.....	۳-۳. روش DWQ (کیفیت انبار داده).....
۱۱۲.....	۳-۴. روش TIQM (مدیریت کیفیت جامع اطلاعات).....
۱۱۴.....	۳-۵. روش AIMQ (روشی برای ارزیابی کیفیت اطلاعات).....
۱۱۵.....	۳-۶. روش CIHI (موسسه کانادایی اطلاعات سلامت).....
۱۱۸.....	۳-۷. روش DQA (ارزیابی کیفیت داده‌ها).....
۱۱۹.....	۳-۸. روش IQM (اندازه‌گیری کیفیت اطلاعات).....
۱۲۲.....	۳-۹. روش ISTAT (اداره ملی سرشماری ایتالیا).....
۱۲۵.....	۳-۱۰. روش AMEQ (اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول).....
۱۲۸.....	۳-۱۱. روش COLDQ (تأثیر کیفیت پائین داده بر هزینه‌ها).....
۱۳۰.....	۳-۱۲. روش DAQUINCIS (کیفیت داده در سیستم‌های اطلاعاتی تعاونی).....
۱۳۴.....	۳-۱۳. روش QAFD (ارزیابی کیفیت داده‌های مالی).....
۱۳۶.....	۳-۱۴. روش CDQ (کیفیت داده‌های کامل).....
۱۳۹.....	۳-۱۵. خلاصه‌ای از روش‌های ذکر شده.....

فصل چهارم : کاربرد مفاهیم کیفیت داده در علم و فناوری	۱۴۱
۱-۴. مقدمه	۱۴۱
۲-۴. معرفی فیلدهای سیستم ثبت	۱۴۲
۳-۴. معرفی مدل کیفیت فراداده انتخابی	۱۴۵
۴-۴. نگاشت اولیه ابعاد کیفی فراداده روی اقلام اطلاعاتی سیستم ثبت	۱۴۷
۵-۴. نتیجه‌گیری و جمع‌بندی	۱۵۸
منابع	۱۵۹
منابع فارسی	۱۵۹
منابع لاتین	۱۵۹
نمایه	۱۶۷
واژه‌نامه فارسی به انگلیسی	۱۷۱
واژه‌نامه انگلیسی به فارسی	۱۷۷

فهرست جدول‌ها

جدول ۱-۱ معرفی کلی روش‌های پیشنهادی تحلیل ریشه‌ای مشکلات کیفی	۱۹
جدول ۱-۲ تعاریف مربوط به جامعیت در مراجع مختلف	۴۷
جدول ۲-۲ تعاریف موجود برای ابعاد مرتبط به زمان	۴۹
جدول ۳-۲ مدل‌های کیفیت داده‌ها	۵۸
جدول ۴-۲ مقایسه مدل‌های کیفیت داده‌ها مبتنی بر روش‌های ارزیابی داده	۵۹
جدول ۵-۲ مقایسه مدل‌های کیفیت داده‌ها مبتنی بر گام‌های بهبود	۶۰
جدول ۶-۲ مقایسه مدل‌های کیفیت داده‌ها مبتنی بر گام‌های بهبود	۶۰
جدول ۷-۲ روش‌ها و انواع استراتژی‌ها	۶۱
جدول ۸-۲ روش‌ها و ابعاد کیفی	۷۰
جدول ۹-۲ ابعاد و معیارها	۷۲
جدول ۱۰-۲ روش‌ها و معیارهای کیفی	۷۴
جدول ۱۱-۲ مقایسه از دیدگاه هزینه‌های کیفیت میان دو مدل COLDQ و TIQM	۷۶
جدول ۱۲-۲ روش‌ها و انواع داده‌ها	۸۱
جدول ۱۳-۲ روش‌ها و سیستم‌های اطلاعاتی	۸۳
جدول ۱-۳ جدولی از جمع‌بندی روش‌های مطالعه شده	۱۴۰
جدول ۱-۴ شناسنامه فیلدهای فراداده‌های ثبت پایان‌نامه‌ها/رساله‌ها	۱۴۲
جدول ۲-۴ طرح کیفی فراداده‌های فیلدهای ثبت پایان‌نامه‌ها در سیستم ثبت (کیفیت ذاتی)	۱۴۸
جدول ۳-۴ طرح کیفی فیلدهای فراداده‌های ثبت پایان‌نامه‌ها در سیستم ثبت (کیفیت زمینه‌ای)	۱۵۲
جدول ۴-۴ طرح کیفی فیلدهای فراداده‌های ثبت پایان‌نامه‌ها در سیستم ثبت (کیفیت بازنمایی)	۱۵۵
جدول ۵-۴ طرح کیفی فیلدهای فراداده‌های ثبت پایان‌نامه‌ها در سیستم ثبت (قابلیت دسترسی)	۱۵۷

فهرست شکل‌ها

شکل ۱-۱ چرخه دمینگ	۲۵
شکل ۲-۱ دسته‌بندی ابعاد (استراندگ و دیگران ۱۹۹۷)	۲۹
شکل ۳-۱ چهار مرحله اصلی در مدیریت کیفیت جامع داده	۳۱
شکل ۱-۲ حوزه‌های اصلی کیفیت داده	۳۷
شکل ۲-۲ مقادیر خالی و جامعیت داده‌ها (بتینی و دیگران ۲۰۰۹)	۴۹
شکل ۳-۲ نمایش‌های گوناگون از اطلاعاتی یکسان در دنیای واقعی (بتینی و دیگران ۲۰۰۹)	۵۳
شکل ۴-۲ طرح کلی روش ISTAT (بتینی و دیگران ۲۰۰۹)	۸۵
شکل ۵-۲ معماری چارچوب ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی (بتینی و دیگران ۲۰۰۹)	۸۷
شکل ۶-۲ دسته‌بندی از روش‌ها	۸۹
شکل ۱-۳ گام‌های مدیریت کیفیت داده	۱۰۷
شکل ۲-۳ گام‌های روش‌شناسی کیفیت انبار داده	۱۱۱
شکل ۳-۳ گام‌های مدیریت کیفیت جامع اطلاعات	۱۱۳
شکل ۴-۳ مدل PSP/IQ	۱۱۴
شکل ۵-۳ گام‌های ارزیابی کیفیت اطلاعات	۱۱۵
شکل ۶-۳ گام‌های روش موسسه کانادایی برای اطلاعات سلامت	۱۱۷
شکل ۷-۳ گام‌های ارزیابی کیفیت داده	۱۱۹
شکل ۸-۳ گام‌های اندازه‌گیری کیفیت اطلاعات	۱۲۱
شکل ۹-۳ گام‌های روش ISTAT	۱۲۴
شکل ۱۰-۳ گام‌های اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول	۱۲۷
شکل ۱۱-۳ مراحل ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها	۱۳۰

- شکل ۳-۱۲ گام‌های ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی..... ۱۳۲
- شکل ۳-۱۳ زیرساخت‌های ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی..... ۱۳۳
- شکل ۳-۱۴ گام‌های ارزیابی کیفیت داده‌های مالی..... ۱۳۶
- شکل ۳-۱۵ گام‌های روش جامع برای مدیریت کیفیت داده..... ۱۳۸
- شکل ۴-۱ ابعاد کیفیت فراداده‌ها در سیستم ثبت..... ۱۴۷

پیشگفتار

در پی رشد فناوری‌ها و ابزارهای ارتباطی، شاهد تولید پایگاه داده‌های بزرگ در اکثر سازمان‌ها هستیم. از طرفی برای کسب جایگاه مناسب در دنیای کسب‌وکار کنونی لازم است سازمان‌های مختلف با تغییرات سازگار شوند و به نوسانات مختلف بازار حساسیت داشته باشند. بعلاوه، کلید حل اکثر مشکلات سازمانی در دل داده‌ها و روندهای تولیدشده توسط همان سازمان است و بررسی داده‌های سازمان‌های دیگر تنها می‌تواند راهنمای رسیدن به پاسخ باشد. بنابراین ارزشمندترین و مهم‌ترین موجودی هر سازمان، داده‌های تولید شده توسط همان سازمان است که در زمان‌های قدیم کمتر موردتوجه بودند. افزایش رقابت و تغییرات شدید بازار در سال‌های اخیر، ارزشمندی داده‌های سازمان‌ها را بیش از پیش نمایان ساخته و تمامی سازمان‌ها در تلاشند تا با بررسی داده‌های خود و سایر داده‌های مرتبط با بازار، عملکرد بهتری از خود نشان دهند. متأسفانه به دلیل ناسازگاری ابعاد مختلف داده و ضرورت ساختاریافتگی داده‌های مختلف، بسیاری از داده‌ها قابل بررسی نیستند.

روش‌های کنترل و مدیریت کیفیت می‌توانند با تعریف چارچوب‌های تولید داده و رفع تناقضات موجود در پایگاه‌های داده، روندی منطقی و بهبودیافته را به سازمان‌ها پیشنهاد دهند تا جریان تولید داده و چگونگی استفاده از آن در سازمان‌هایی با شرایط منحصربه‌فرد به‌درستی تعریف شود. هرچند روش‌های کنترل و مدیریت کیفیت در حوزه‌های مختلفی تعریف شده‌اند ولی می‌توان آن‌ها را برای حوزه داده نیز استفاده کرد. برای این منظور بایستی مفاهیم و روش‌های کنترل کیفیت

در چارچوب داده تعریف شوند تا توانایی رفع تناقضات و مشکلات مربوطه را داشته باشند.

کتاب حاضر باتوجه به ضرورت بررسی ابعاد مختلف پایگاه‌های داده و چگونگی روند تولید آن‌ها نگارش شده است. در این کتاب تلاش می‌شود تا با بررسی معیارهای داده و مدل‌های مدیریت کیفیت داده، پاسخی مناسب برای رفع تناقضات و ناسازگاری‌های پایگاه‌های داده بیان گردد. برای این منظور، در ابتدا به بررسی مفاهیم کنترل و مدیریت کیفیت پرداخته می‌شود و چگونگی استفاده از آن‌ها در بررسی کیفیت داده‌ها بیان می‌گردد. سپس مبانی کیفیت داده و اطلاعات مورد تجزیه و تحلیل قرار می‌گیرد تا روش‌ها، مراحل، گام‌ها، استراتژی‌ها و روش‌های مرتبط با ابعاد مختلف کیفیت داده مشخص شود. در ادامه، مدل‌های مختلف مدیریت کیفیت داده بررسی شده و ویژگی‌ها و کاربردهای خاص هریک از مدل‌های توصیف می‌شود. در فصل آخر به کاربرد مفاهیم کیفیت داده در علم و فناوری پرداخته می‌شود. در این فصل پس از بررسی فرایند ثبت پایان‌نامه‌ها در سیستم ثبت، فرایند جدیدی برای آن پیشنهاد می‌گردد تا مشکلات مدل قبلی برطرف گردد به گونه‌ای که بتوان از سایر مزایای بالقوه داده‌های ثبت‌شده در این پایگاه داده به‌طور مناسب استفاده کرد.

دکتر محمدجواد ارشادی پست الکترونیکی: ershadi@irandoc.ac.ir

مهندس محمد مهدی ارشادی پست الکترونیکی: ershadi.mm1372@aut.ac.ir

فصل اول

مفاهیم اولیه کنترل و مدیریت کیفیت

۱-۱. مقدمه

از هنگامی که انسان توانایی ساخت هر محصول را پیدا کرده است، جهت کنترل کیفیت آن نیز به تلاش پرداخته است. ظرافت، دقت و هنرمندی موجود در آثار باستانی و محصولات متعلق به زمان‌های گذشته حکایت از این تلاش دارد. با شروع انقلاب صنعتی در اواسط قرن هیجدهم ماشین‌های تولیدی به تدریج جایگزین ابزارها و مهارت‌های فردی افراد هنرمند و صنعت‌گر گردید. با پیدایش روش‌های جدید و پیچیده تولید، اشتیاق به تولید بیشتر افزایش یافت و بدین ترتیب نیاز به کنترل کیفیت محصول نهایی نیز ابعاد تازه‌ای پیدا نمود.

بین روش‌های اولیه کنترل کیفیت و روش‌های آماری امروزی کنترل کیفیت، شباهت اندکی به چشم می‌خورد. پایه و اساس کنترل کیفیت آماری به مفهوم امروزی آن در سال ۱۹۲۰ بنا نهاده شد. برای فهم بهتر واژه کنترل کیفیت، تعریف جداگانه کلمات کیفیت و کنترل لازم است. کیفیت و کنترل در تحقیقات انجام‌شده به گونه‌های مختلفی تعریف شده‌اند. با جمع‌بندی تعاریف مختلف می‌توان چنین گفت که کیفیت یعنی شایستگی محصول جهت استفاده خاص، به عبارت دیگر کیفیت برابر شاخصی است که محصول توانسته انتظارات مصرف‌کننده خود را برآورده

سازد. کنترل به معنی اعمال ضوابط و راهنمایی در مورد کسی یا چیزی جهت اطمینان از کسب نتایج موردنظر تعریف کرده‌اند (بالو^۱ و دیگران ۱۹۹۸).

معنی کنترل کیفیت مجموع تعاریف دو واژه فوق را در برمی‌گیرد. در صنعت برداشت‌های مختلفی از کنترل کیفیت وجود دارد. بعضی از شرکت‌ها عملیات بازرسی جهت کسب اطمینان از کیفیت محصول را کنترل کیفیت می‌نامند. در مقابل در شرکت‌هایی با تاکید بر کنترل کیفیت، علاوه بر عملیات بازرسی به سایر عملیات‌ها مانند برنامه‌ریزی کیفیت، کنترل کیفی تولید، کنترل موارد ورودی، تجزیه و تحلیل و اقدام چاره‌جویانه در رابطه با نقص‌های تولید و تهیه روش‌های گزارش‌های مربوط به مسائل کیفی نیز کنترل کیفیت گفته می‌شود. کنترل کیفیت نباید به کاربرد محدود واژه آن منحصر گردد بلکه باید تمامی فعالیت‌های لازم برای کسب سطح مطلوبی از درستی و بی‌عیبی محصول را شامل شود. کنترل کیفیت سیستمی است جهت رسیدن به سطح مطلوبی از کیفیت یک محصول یا یک فرایند تولید و نگهداری با برنامه‌ریزی دقیق، استفاده از ماشین‌آلات مناسب، بازرسی مستمر و عمل اصلاح‌کننده هرگاه که لازم باشد (بیاگی^۲ و دیگران ۲۰۱۴).

کیفیت یک محصول معمولاً در رابطه با سه عامل زیر تعیین می‌شود:

- کیفیت طرح
- کیفیت انطباق
- کیفیت عملکرد

عوامل فوق با هزینه و اقتصاد کیفیت ارتباط نزدیکی دارند. در ادامه به معرفی هریک از عوامل فوق‌الذکر خواهیم پرداخت.

- **کیفیت طرح:** دو محصول ممکن است برای کار مشابهی استفاده شوند ولی در طراحی آن‌ها اختلاف زیادی وجود داشته باشد. یک محصول ممکن است از پلاستیک معمولی و از قطعاتی با کیفیت پایین ساخته شده باشد

که مسلماً در این صورت عمر استفاده از آن هم کوتاه است. ممکن است محصول دیگری دارای قطعات سازنده باکیفیت و طراحی مناسبی باشد، اجزای آن با تلورانس‌های دقیق ساخته شوند و قابلیت استفاده طولانی را داشته باشد. کاملاً مشخص است که کیفیت طراحی محصول دوم به مراتب از کیفیت طراحی محصول اول برتر می‌باشد.

- **کیفیت انطباق:** کیفیت انطباق درجه همسویی محصول با مشخصات، استانداردها و معیارهای تعیین شده برای ساخت آن محصول است. بنابراین اگر محصولی طبق مشخصات و حدود کنترل فرایند تولید ساخته شود و ویژگی‌های آن به خوبی بیان‌گر نیازهای مصرف‌کننده باشد، می‌تواند از کیفیت خوبی برخوردار باشد و رضایت مشتری را جلب کند.
- **کیفیت عملکرد:** کیفیت عملکرد یک محصول تابعی از کیفیت طرح و کیفیت انطباق آن محصول است. حفظ سطح بالایی از هر دو کیفیت مذکور، به سطح بالایی از کیفیت عملکرد منجر می‌شود. اگر کیفیت طرح ضعیف باشد یا تطابق با مشخصات مناسب وجود نداشته باشد، کیفیت عملکرد محصول کاهش می‌یابد. مسلماً داشتن بالاترین سطح تطابق با مشخصات محصول چنانچه طراحی محصول ضعیف باشد، فایده‌ای در پی نخواهد داشت.

کنترل کیفیت آماری در دهه ۱۹۲۰ توسط والتر شوهارت^۱ از آزمایشگاه‌های بل تلفن آمریکا پایه‌گذاری گردید. وی در یادداشتی در ۱۶ می ۱۹۲۴ اولین تصویر نمودارهای کنترل را ترسیم کرد و به بررسی بیشتر این روش در مطالعات بعدی خود پرداخت. وی نتیجه پژوهش‌های خود را در کتابی تحت عنوان (کنترل اقتصادی کیفیت محصولات ساخته شده)^۲ در سال ۱۹۳۱ منتشر ساخت. دو همکار دیگر شوهارت به نام‌های داج و رومینگ^۳ نیز کاربرد تئوری آمار را در نمونه‌گیری بررسی کرده و نتیجه

1. Walter A. Shewhart
3. Dodge And Romig

2. Economic Control Of Quality Of Manufactured Product

کار آنان منجر به انتشار جدول‌های معروف بازرسی داج-رومیگ در سال ۱۹۴۴ گردید. مجموعه کارهای شوهارت، داج و رومیگ اساس دانشی را تشکیل می‌دهد که امروزه کنترل کیفیت آماری خوانده می‌شود.

از اوایل سال ۱۹۴۰، انجمن استانداردهای آمریکا به دعوت وزارت جنگ پروژه‌ای را شروع کرد که سرانجام منجر به تدوین استانداردهای نظامی آمریکایی^۱ ۱/۱-۱۹۴۱ و ۱/۲-۱۹۴۱ «راهنمای کنترل کیفیت و روش‌های نمودارهای کنترل برای تجزیه و تحلیل داده‌ها» و ۱/۳-۱۹۴۲ «روش نمودارهای کنترل برای کنترل کیفیت در حین تولید» گردید. مطالب فوق به عنوان مواد درسی دوره‌های آموزشی به کار گرفته شد.

در بافالو^۲ انجمن مهندسين کنترل کیفیت با همکاری دانشگاه بافالو در سال ۱۹۴۴ تأسیس گردید و مجله کنترل کیفیت صنعتی^۳ منتشر شد. مجله جدید به سرعت در سراسر کشور راه یافت. تمام این کوشش‌ها زمینه استفاده هر چه بیشتر از کنترل کیفیت آماری را در سال‌های بعد از جنگ جهانی دوم به وجود آورد.

در فاصله کوتاهی از پایان جنگ جهانی دوم، زمینه لازم برای تأسیس سازمان ملی به نام انجمن کنترل کیفیت آمریکا فراهم گردید. این انجمن انتشار مجله کنترل کیفیت صنعتی را به دست گرفت و به بزرگ‌ترین مروج کاربرد کنترل کیفیت آماری در قاره آمریکا تبدیل شد و شعبه‌ای نیز در ژاپن تشکیل داد. به دنبال سفر شوهارت به لندن در ماه می ۱۹۳۲، ایگان اس.پی یرسون^۴ مقاله‌ای درباره کاربرد صنعتی آمار در انجمن سلطنتی آمار ارائه کرد. در فاصله کمی پس از آن، انجمن مذکور بخشی را برای پژوهش‌های صنعتی و کشاورزی در نظر گرفت و پیوستی درباره کاربرد این رشته جدید به مجله خود افزود.

در سال ۱۹۵۰ از تلفیق روش‌های نمونه‌گیری نظامی و جدول‌های بازرسی وصفی استاندارد MIL-STD-105 تهیه و منتشر گردید. در سال ۱۹۵۷ نیز استاندارد

1. American War Standard Z
3. Industrial Quality Control

2. Buffalo
4. Igon S.P. Ryerson

MIL-STD-414 جدول‌های بازرسی متغیرها منتشر گردید. در انگلستان کنترل کیفیت آماری در اوایل پیشرفتی سریع‌تر از آمریکا داشت. انجمن استانداردهای انگلستان نیز علاقه خود را با انتشار کتابی تحت عنوان (کاربرد روش‌های آماری در استانداردسازی صنایع و کنترل کیفیت)^۱ نشان داد.

عکس‌العمل صنایع انگلستان نسبت به روش‌های جدید آماری سریع و همه‌جانبه بود، بطوریکه تا سال ۱۹۳۷ در مورد محصولاتی از قبیل زغال‌سنگ، الیاف پنبه، منسوجات پنبه‌ای و پشمی، شیشه، لامپ، مصالح ساختمانی و محصولات شیمیایی از روش‌های جدید کنترل کیفیت آماری استفاده شد. روش‌های کنترل کیفیت آماری از آمریکا و انگلستان به سایر کشورها راه یافت. تحت راهنمایی‌های دکتر ادواردز دمنینگ^۲ روش‌های کنترل کیفیت آماری در ژاپن تا سطح بهترین در جهان توسعه یافت. در اروپا انجمن اروپایی کنترل کیفیت به وجود آمد. امروزه تقریباً تمام کشورهای صنعتی جهان از روش‌های کنترل کیفیت آماری استفاده می‌کنند. در ادامه به بررسی و مقایسه مباحث حوزه‌های کنترل کیفیت و بازرسی خواهیم پرداخت.

۲-۱- تفاوت کنترل کیفیت و بازرسی

برای هر واحد صنعتی که با ساخت محصولی سروکار دارد بازرسی و کنترل کیفیت هر دو لازم است اما بین این دو تفاوت آشکاری وجود دارد. در بازرسی به کیفیت محصولی که ساخته شده است اهمیت داده می‌شود و کار بازرسی با تعیین محصولات خوب و بد پایان می‌پذیرد. برای مثال، اگر ساخت ۱۰۰۰ قطعه به‌اندازه $10/000 - 0/250$ و $10/000 + 0/250$ سانتیمتر مورد نظر باشد؛ بازرس قطعات ساخته‌شده را اندازه‌گیری کرده و قطعات خارج از اندازه موردنظر را رد می‌کند و تولید تا تهیه تعداد لازم یعنی ۱۰۰۰ قطعه ادامه می‌یابد. برعکس بازرسی در کنترل کیفیت به فرایند تولید محصول تمرکز دارد و درباره آینده آن تصمیم‌گیری می‌شود.

1. The Application Of Statistical Methods To Industrial Standardization And Quality Control, E.S Pearson
2. W. Edwards Deming

در کنترل کیفیت تلاش می‌شود امکانات تولید به نحوی تنظیم شوند که همیشه محصول با کیفیت تولید شود. کنترل کیفیت از بازرسی به‌عنوان وسیله‌ای جهت رسیدن به این هدف استفاده می‌کند.

در صورتی که برنامه کنترل کیفیت موفقیت‌آمیز باشد، نیاز به بازرسی نسبت به گذشته کمتر خواهد بود و ممکن است نیاز به بازرسی کاملاً از بین برود. روش‌های مختلفی برای کنترل کیفیت وجود دارد که یکی از مهم‌ترین آن‌ها روش‌های آماری کنترل کیفیت است. برای مثال، در حین تولید قطعات مثال بالا می‌توان نمونه‌هایی متناوب از قطعات ساخته شده برداشت و هر قطعه را در نمونه برداشته‌شده بازرسی کرد. اگر کیفیت اقلام یک نمونه بخصوص رضایت‌بخش بود، ادامه تولید مجاز است و در غیر این صورت باید بلافاصله اقدام اصلاحی انجام گیرد. به‌طور خلاصه، اطلاعاتی که از بازرسی یک نمونه از محصول به دست می‌آید در تصمیم‌گیری درباره اینکه آیا ایجاد تغییری در فرایند تولید لازم است یا خیر مورد استفاده قرار می‌گیرد.

۱-۳- اهمیت روش‌های آماری در کنترل کیفیت

مؤثرترین شیوه موجود برای کنترل کیفیت محصولات، روش‌های آماری^۱ است. با روش‌های آماری می‌توان تصویری از وضعیت کل تولید به دست آورد. باتوجه به اینکه تغییرپذیری^۲، پدیده‌ای دائمی و جزء ضروری همه محصولات بوده و مشخصه‌های کیفی هر محصول می‌تواند تغییر کند، روش‌های آماری مؤثرترین ابزار برای بررسی و کنترل این تغییرات است. مادامی که از مواد، افراد، روش‌ها و ماشین‌ها برای تولید استفاده می‌شود، مشکل تغییر کیفیت وجود خواهد داشت و مادامی که این مشکل وجود داشته باشد، روش‌های آماری کنترل کیفیت^۳ نیز ضروری است.

به‌طور کلی، روش‌های آماری کنترل کیفیت بیشتر در کنترل کیفیت محصولاتی که به‌صورت انبوه تولید می‌شوند، مؤثر هستند. هرگاه تولید تکراری باشد، چه به‌صورت جریان پیوسته و ساخت پی‌درپی محصولات و یا به‌صورت بهر به بهر،

1. Statistical Methods

2. Variability

3. Quality Control Statistical Methods

روش‌های آماری کنترل کیفیت بیشترین استفاده را پیدا می‌کنند. همانگونه که در بخش مقدمه اشاره شد روش‌های مختلفی جهت شناسایی ریشه‌ای مشکلات کیفی وجود دارد که هر کدام به فراخور مشکل موجود و هدف کارگروه می‌توانند در جایگاه خود مورد استفاده قرار گیرند. در این بخش به معرفی برخی از این شیوه‌ها خواهیم پرداخت. در ادامه به منظور استفاده کامل‌تر از هریک از روش‌های تحلیلی و پیش از معرفی جداگانه هریک از روش‌ها در قالب جدول زیر به معرفی اجمالی هریک از روش‌ها پرداخته می‌شود.

جدول ۱-۱ معرفی کلی روش‌های پیشنهادی تحلیل ریشه‌ای مشکلات کیفی

نام روش	کاربرد
طوفان فکری	این روش به منظور بررسی و تحلیل مشکلات کیفی با علت نامعلوم مورد استفاده قرار می‌گیرد. جلسات طوفان فکری با حضور طیف متنوعی از متخصصین مختلف برگزار می‌گردد.
نمودار ایشی کاوا	این نمودار در بررسی سیستماتیک و دسته‌بندی شده علل وقوع مشکلات کیفی کاربرد دارد.
چراهای پنج‌گانه	این روش در زمان بروز مشکلات کیفی در یک حوزه تخصصی خاص (مانند نرم‌افزار) به منظور تعیین ریشه‌های این دسته از مشکلات توسط افراد خبره کاربرد دارد.
نمودار پارتو	این نمودار به منظور تعیین میزان و درجه اهمیت ریشه‌های متعدد شناسایی شده برای یک مشکل کیفی مورد استفاده قرار می‌گیرد.

۱-۳-۱- طوفان فکری^۱

طوفان فکری یا طوفان مغزی یکی از ابزارها و شیوه‌هایی است که به صورت معمول جهت ریشه‌یابی و حل مشکلات کیفی سیستم‌ها و محصولات مورد استفاده قرار می‌گیرد. در حقیقت این شیوه به عنوان بستر مناسب به منظور استفاده از سایر شیوه‌ها می‌تواند مورد استفاده قرار گیرد. طوفان فکری یک روش ایجاد خلاقیت در کار گروهی است که برای پیدا کردن یک راه‌حل برای یک مشکل خاص یا جمع‌آوری فهرستی از ایده‌ها مورد استفاده قرار می‌گیرد (لهر^۲ ۲۰۱۳). اصطلاح طوفان فکری برای اولین بار توسط الکس آزبورن^۳ در سال ۱۹۵۳ مورد استفاده قرار

1. Brainstorming

2. Lehrer

3. Alex Faickney Osborn

گرفت. او از سال ۱۹۳۹ به‌عنوان یک مدیر تبلیغات استفاده از روش‌های خلاق برای حل مسائل استفاده کرد و به‌تدریج جلساتی را ترتیب داد که در قالب آن‌ها ایده‌های جدید در راه‌اندازی کمپین‌های تبلیغاتی موردبررسی قرار می‌گرفت. آذربورن در سال ۱۹۵۳ روش خود را در کتاب «قدرت خلاق شما»^۱ مطرح نمود (آذربورن ۱۹۶۳). وی این ادعا را مطرح کرد که دو اصل زیر به ارائه مؤثر ایده‌ها کمک شایانی می‌کند (اسکریبنر^۲ ۱۹۴۸):

- ارائه قضاوت‌های متفاوت از یک مسئله
 - در نظر گرفتن جنبه‌های کمی در ارائه ایده‌ها
- او پس از این دو اصل، چند قانون عمومی در استفاده از روش طوفان فکری ارائه نمود تا بتواند در راستای این قوانین به اهداف زیر دست یابد:
- کاهش موانع اجتماعی در میان اعضای گروه
 - تحریک تولید ایده
 - افزایش خلاقیت در کار گروهی
- قوانین مطرح‌شده توسط وی به شرح زیر هستند:
- **فراوانی زیاد ایده‌ها:** فرض اساسی در نظر گرفته‌شده در این قانون این است که افزایش ایده‌ها شانس دستیابی به یک راه‌حل با بالاترین کیفیت را افزایش می‌دهد. در واقع تولید طیف وسیعی از ایده‌ها، فرایند حل مسئله را تا حد زیادی تسهیل خواهد نمود.
 - **خودداری از نقد:** در طوفان فکری، انتقاد از ایده‌های مطرح‌شده بایستی متوقف گردد و به زمان مناسب دیگری موکول گردد. حاضرین در جلسه بایستی به توسعه ایده‌ها و ارائه ایده‌های نو پردازند و در جلسه دیگری ایده‌ها را نقد نمایند. با تعلیق قضاوت، شرکت‌کنندگان قادر خواهند بود به تولید ایده‌های غیرمعمول پردازند.

- استقبال از ایده‌های وحشی: برای دریافت یک فهرست بلند از ایده‌ها، ایده‌های وحشی بایستی تشویق شوند. طرز تلقی‌های جدید و غیرمعمول از مسائل می‌توانند راه‌حل‌های بهتر و مؤثری را ایجاد نمایند.
- ترکیب و بهبود ایده‌ها: همان‌گونه که شعار « $1+1=3$ » اثرات هم‌افزایی را تحلیل کرده است، تجمیع و ترکیب ایده‌ها راه‌حل‌های مناسب و مؤثرتری را ارائه خواهند نمود.

در جلسات مربوط به بررسی مشکلات کیفی که ممکن است اتفاق نظر جمعی روی راهکارها و ریشه‌ها وجود نداشته باشد، می‌توان از رویکرد طوفان فکری استفاده و نتایج کلیه صاحب‌نظران و کارشناسان را جمع‌آوری نمود. این روش کمک خواهد کرد تنوع بسیار زیادی از ایده‌ها در جلسه مطرح شود و همه ایده‌ها ارائه شوند. کارگروه بهبود می‌تواند در موارد بروز مشکلات حاد کیفی با تشکیل جلسات طوفان فکری ایده‌های مختلف را جمع‌آوری نموده و احتمال دستیابی به راهکارهای عملیاتی مؤثر و کارا را تا حد بالایی افزایش دهد و در نتیجه به حل ریشه‌ای مشکل کمک ویژه‌ای داشته باشد.

۱-۳-۲- نمودار ایشی کاوا^۱

زمانی که عیب، اشکال و یا اشتباهی شناسایی می‌شود باید علل بالقوه آن نیز تعیین گردد. در مواقعی که مجموع علل بروز مشکل واضح نیست یا فقط دو یا چند مورد از آن‌ها مشخص است، نمودار علت و معلول می‌تواند ابزار مفیدی برای شناسایی علل بالقوه باشد. نمودار علت و معلول با نام نمودار «ایشی کاوا» یا «نمودار استخوان ماهی»^۲ نیز شناخته می‌شود؛ زیرا این نمودار توسط دکتر ایشی کاوا در سال ۱۹۴۳ مطرح شد. این نمودار شکلی شبیه یک ماهی دارد که مشخصه کیفی در سر آن و علت‌ها به شکل استخوان ماهی ترسیم می‌شوند. مراحل تهیه نمودار علت و معلول عبارتند از:

1. Ishikawa Chart

2. Fishbone Diagram

۱. مشکل یا معلولی که باید تجزیه و تحلیل شود را تعریف کنید؛
۲. گروهی برای انجام تجزیه و تحلیل‌های مورد نیاز تشکیل دهید. در اغلب موارد تیم بهبود کیفیت، علل بالقوه ایجاد مشکل را از طریق طوفان ذهنی تعیین می‌کند؛
۳. خط مرکز را رسم کرده، مشکل (معلول) را در سمت راست آن (در جلو پیکان) قرار دهید؛
۴. گروه‌های علل بالقوه را تعیین و آن‌ها را به خط مرکزی متصل کنید. معمولاً گروه‌های اصلی به صورت زیر دسته‌بندی می‌شوند و معروف به 5M+1E و یا 6M می‌باشند.

- | | | |
|----------------------------|-----------------------------|------------------------------|
| ۱. مواد اولیه ^۱ | ۳. تجهیزات ^۳ | ۵. نیروی انسانی ^۳ |
| ۲. روش‌ها ^۴ | ۴. اندازه‌گیری ^۵ | ۶. محیط ^۶ |

۵. علل ممکن را شناسایی کرده، آن‌ها در گروه‌های تعیین شده در مرحله ۴ قرار دهید. در صورت نیاز، گروه‌های دیگری تشکیل بدهید. علل باید تا پایین‌ترین سطح فهرست شود؛
۶. علل را رتبه‌بندی کنید تا آن‌هایی که اثر زیادی بر مشکل دارند شناسایی شوند؛

نمودار علت و معلول یکی از ابزارهای قوی برای بررسی مشکلات است. یک نمودار علت و معلول که با جزئیات مناسب تهیه شده باشد را می‌توان به عنوان ابزار گره‌گشای مؤثر استفاده نمود. به علاوه، تهیه یک نمودار علت و معلول کمک خواهد کرد تا با تشکیل گروه، افراد به صورت گروهی مشکل را حل کنند.

به منظور ایجاد اطمینان از بررسی دقیق و همه‌جانبه کلیه عوامل مؤثر در بروز مشکلات کیفی در جلساتی که به بررسی مشکلات کیفی سیستم‌های اطلاعاتی می‌پردازند می‌توان از این روش استفاده نمود. در قسمت تحلیل ریشه‌ای بروز

عدم انطباق فرم‌های اقدام اصلاحی می‌توان از نمودار استخوان ماهی جهت بیان دقیق و دسته‌بندی شده ریشه‌های بروز مشکل بهره برد.

۱-۳-۳- روش چراهای پنج‌گانه^۱

چراهای پنج‌گانه یک روش استقرایی به منظور کشف علت یا علل اصلی وقوع یک مشکل خاص است (بانک توسعه آسیا^۲ ۲۰۰۹). هدف اصلی استفاده از این روش این است که بتوان علت ریشه‌ای وقوع یک نقص یا مشکل را به کمک تکرار کلمه «چرا؟» پیدا کرد. در این روش هر سؤال مبنا و زیربنای سؤال بعدی را تشکیل خواهد داد. متخصصین کیفیت به تجربه و براساس مطالعات میدانی صورت گرفته به عدد ۵ در تکرار سؤال‌ها دست یافته‌اند. این روش به‌طور رسمی توسط ساکیچی توئودا^۳ در شرکت توئوتا^۴ توسعه داده شده است. تایچی اوهنو^۵ در شرکت توئوتا ادامه‌دهنده این روش بود و در حوزه‌های مختلف عملیاتی آن‌ها را به کار برد. در خصوص استفاده از این روش نکات زیر لازم به ذکر است:

همه مشکلات لزوماً یک علت ریشه‌ای واحد ندارند لذا اگر کسی بخواهد به کشف علل ریشه‌ای متعدد بپردازد، بایستی روش چراهای پنج‌گانه را با طرح سؤالات متفاوت مورد استفاده قرار دهد.

روش چراهای پنج‌گانه کاملاً بر دانش و تداوم افراد درگیر در حل مسئله وابسته است؛ لذا لازم است افراد مختلف با طیفی از توانایی‌ها و تخصص در زمان کاربرد این روش حضور داشته باشند.

در جلسات تخصصی تضمین کیفیت سیستم‌های اطلاعاتی در قالب کارگروه بهبود و در زمان بروز مشکلات خاص مانند مشکلات سخت‌افزاری و نرم‌افزاری، با حضور گروه متخصص می‌توان از روش چراهای پنج‌گانه بهره‌برداری کرد.

1. 5whys

2. Asian Development Bank

3. Sakichi Toyoda

4. Toyota Automotive Manufacturer

5. Taiichi Ohno

۱-۳-۴- روش پارتو^۱

آلفرد پارتو^۲ در اواخر قرن ۱۹ و اوایل قرن ۲۰ میلادی ضمن مطالعه‌ای که در مورد توزیع ثروت در اروپا داشت به این نتیجه رسید که جمعیت زیادی با ثروت کم در مقابل جمعیت کمی با ثروت زیاد قرار دارند؛ به بیان دیگر اقلیتی پردرآمد در مقابل اکثریتی کم درآمد در اروپا در حال زندگی هستند. جوران^۳ به عنوان یکی از بزرگان کنترل کیفیت چندسال پس از ارائه این اصل توسط پارتو، اصل پارتو را در مورد کنترل کیفیت و عوامل ایجاد عدم انطباق‌ها جامعیت داد. بنابراین در کیفیت محصول هم می‌توان عوامل کمی را یافت که از تأثیر فراوانی برخوردار هستند و در مقابل، عوامل زیادی نیز ممکن است یافته شوند که مجموع تأثیرشان در کیفیت محصول ناچیز است. براساس اصلی که پارتو در مورد اقتصاد بیان کرد ۸۰ درصد نتایج و مسائل از ۲۰ درصد علل ناشی می‌شوند. به بیان دیگر، اگرچه ممکن است برای مسائل موجود علل فراوانی وجود داشته باشد، ولی اندکی از این علل اهمیت داشته و با رفع آن‌ها می‌توان بخش عمده مسائل را حل کرد. برای نمونه می‌توان گفت:

- ۸۰ درصد خطاها توسط ۲۰ درصد افراد صورت می‌گیرد؛
- ۸۰ درصد ضایعات محصول به علت مشکل در ۲۰ درصد فرایندهای موجود است.

به کمک نمودار پارتو می‌توان علل پدید آمدن نواقص را طبقه‌بندی کرد و به روشنی نشان داد که کدام دسته از علل از اهمیت بیشتری برخوردار است. برای رسم نمودار پارتو میله متناظر هر کدام از علل با همان ترتیب صعودی رسم می‌شود. سپس با محاسبه جمع تجمعی درصد فراوانی‌ها می‌توان عللی را که باعث بروز حدود ۸۰ درصد مشکلات شده‌اند را تعیین کرد. به منظور تعیین میزان تأثیر هریک از علل ریشه‌ای در گام بهبود کیفیت داده که علل ریشه‌ای مشکلات کیفی تا حدی مشخص

1. Pareto Method

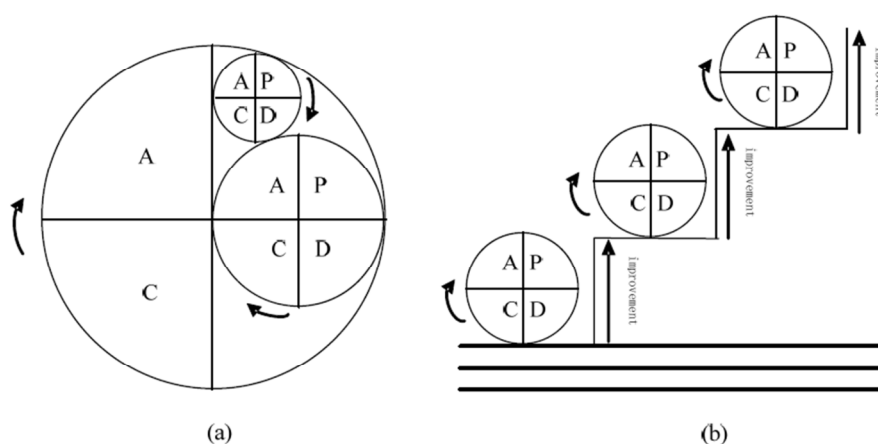
2. Vilfredo Federico Damaso Pareto

3. Joseph M. Juran

شده‌اند، می‌توان از روش پارتو استفاده نمود. این روش در زمانی که تعداد علل ریشه‌ای شناسایی شده زیاد بوده و لازم است از بین موارد متعدد از ریشه‌ها تعداد محدودی را انتخاب نمود، کاربرد فراوانی دارد. در ادامه به تاریخچه و مبانی کیفیت داده خواهیم پرداخت.

۴-۱- چرخه دمینگ و جایگاه آن در مدیریت کیفیت

رویکرد اصلی بهبود مستمر به عنوان یکی از مهم‌ترین اصول مدیریت کیفیت شامل چهار مرحله اصلی چرخه دمینگ^۱ است. در هر چهار مرحله برنامه‌ریزی^۲، اجرا^۳، کنترل^۴ و اقدام^۵، لازم است سازمان برای تضمین کیفیت خدمات خود و کاهش خطا در فرایندهای تولید داده و اطلاعات، اقدامات لازم را انجام دهد. همان‌گونه که در شکل ۱-۱ مشاهده می‌شود با بهره‌مندی دائم از چرخه دمینگ، بهبود مرحله به مرحله در سیستم یا فرایند به کار رفته رخ خواهد داد. همچنین مدل شماتیک چهار مرحله‌ای چرخه دمینگ در قالب این شکل مشاهده می‌شود.



شکل ۱-۱ چرخه دمینگ

برنامه‌ریزی: در این مرحله برنامه‌ریزی‌های مقتضی براساس بهبود مستمر کیفیت انجام می‌گیرد. طرح کیفیت، سندی است که در این رابطه در مرحله برنامه‌ریزی تدوین می‌شود. برای کسب اطمینان از تحت کنترل بودن مستندات و فرایندهای سازمان، رویه‌ای با عنوان «روش اجرایی کنترل مستندات»^۱ در نظر گرفته می‌شود. همچنین، روش اجرایی بازنگری طراحی و توسعه سیستم معرفی می‌شود که در قالب آن اقدام‌ها و ضرورت‌های لازم برای بهبود کیفیت نرم‌افزار مشخص خواهد شد.

اجرا: در این مرحله فرایندها متناسب با چارچوب و ضوابط سازمان اجرا خواهند شد. باتوجه به متناسب بودن فعالیت‌های هر فرایند با الزامات طرح کیفی، کیفیت خروجی فرایندها نیز مطلوب خواهد بود.

کنترل: در این مرحله بررسی‌های لازم به‌منظور اطمینان از عدم بروز خطا و کنترل در سیستم صورت می‌گیرد که براساس اقدام‌های انجام‌گرفته به دو گروه کنترل سیستمی و کنترل کیفی داده‌ها دسته‌بندی می‌شوند.

اقدام: در این مرحله تمام مغایرت‌های مشاهده‌شده در کنترل‌های سیستمی و داده‌ها بررسی و اقدام‌های اصلاحی و پیشگیرانه لازم باتوجه به ملزومات بهبود کیفیت اعمال خواهد شد.

۱-۵- کیفیت داده^۲

داده‌ها دارایی‌های اطلاعاتی بنیادینی هستند که تمرکز اصلی امنیت سیستم‌های اطلاعاتی به آنها وابسته است (دهیلون و بکهوس^۳ ۲۰۰۰). از این‌رو امروزه کیفیت داده‌ها و فراداده‌ها^۴ در فرایند کسب‌وکار تأثیر فراوانی داشته و نقش بسیار مهمی را در بسیاری از سازمان‌ها ایفا می‌کند. داده‌ها در سازمان‌ها نیز منبع حیاتی برای حمایت از فرایند کسب‌وکار و تصمیم‌گیری مدیریتی به شمار می‌روند. (خسروانجوم^۵ و دیگران ۲۰۱۱). رشد انبار داده‌ها و دسترسی مستقیم به اطلاعات از منابع مختلف

1. Documentary Execution Control Method
3. Dhillon & Backhouse

2. Data Quality
4. Metadata

5. Khosroanjom

توسط مدیران و کاربران اطلاعات، نیاز و همچنین آگاهی از کیفیت بالای داده‌ها را در سازمان‌ها افزایش داده است (لی^۱ و دیگران ۲۰۰۲).

همان‌گونه که تولید محصول را می‌توان به‌صورت سیستم پردازشی دانست که بر روی مواد خام فعالیت کرده تا آن‌ها را به محصولات فیزیکی تبدیل کند، تولید اطلاعات را نیز می‌توان به‌صورت یک سیستم پردازشی توصیف کرد که بر روی داده‌های خام فعالیت می‌کند تا آن‌ها را به محصولات اطلاعاتی تبدیل کند (وانگ^۲ ۱۹۹۸).

جدول ۱-۲ مقایسه تولید محصول و اطلاعات (وانگ ۱۹۹۸)

تولید اطلاعات	تولید محصول	
داده خام	مواد خام	ورودی
سیستم اطلاعاتی	خط تولید	پردازش
محصولات اطلاعاتی	محصولات فیزیکی	خروجی

امروزه با پیشرفت فناوری اطلاعات و تغییر سیستم‌های اطلاعاتی از ساختار سلسله‌مراتبی به ساختار شبکه‌ای، با حجم بزرگی از داده‌ها مواجه شده‌ایم. بدین ترتیب با افزایش حجم داده‌ها از نظر اندازه و دامنه در سال‌های اخیر مدیریت آن‌ها پیچیده‌تر شده و خطر کاهش کیفیت داده‌ها رو به افزایش است. داده‌های نامرغوب نیز می‌توانند بر قابلیت سیستم و همچنین تضمین عملکرد عملیاتی زیان وارد کنند (ردمن^۳ ۱۹۹۸). از جهت دیگر داده‌های نامرغوب می‌توانند اثرات اجتماعی و اقتصادی قابل‌توجهی داشته (وانگ و استرانگ^۴ ۱۹۹۶) و بر رضایت مشتریان و روحیه‌ی کارکنان و اطمینان میان سازمان‌ها نیز تأثیرگذار باشند (دهیلون و بکهوس ۲۰۰۰). در سیستم‌های اطلاعات شبکه‌ای، اگر اطلاعات ورودی و کیفیت فرایندها کنترل نشده باشند، می‌توانند بر روی کیفیت کلی داده‌های سیستم اطلاعات تأثیر گذاشته و به‌مرور زمان کیفیت داده‌ها را کاهش دهند. همان‌گونه که گفته شد، عوامل

1. Lee

2. Wang

3. Redman

4. Wang And Strong

بسیاری بر فرایندهای تصمیم‌گیری و نتایج آن‌ها تأثیرگذار است که داده‌های نامرغوب یکی از آن‌ها بوده و تصمیمات ضعیفی را در پی دارند. به دلیل وجود حجم بالای داده‌های در دسترس، کنترل داده یک امر مهم در سازمان‌ها به شمار می‌آید. پژوهش‌ها حاکی از آن است که تصمیم‌گیرندگانی که با کنترل داده آشنایی داشته‌اند از آن بیشتر برای پشتیبانی از تصمیم‌گیری‌ها استفاده کرده و دقت بالاتری در تصمیمات خود داشته‌اند (موگس^۱ و دیگران ۲۰۱۶).

کیفیت داده‌ها خود داده‌ای است که برای مصرف‌کنندگان داده مناسب است. علاوه بر این ابعاد کیفیت داده را به‌عنوان مجموعه‌ای از ویژگی‌ها که یک جنبه و یا ساختار کیفیت داده را نشان می‌دهند، در نظر می‌گیرند (وانگ و استرانگ ۱۹۹۶). مطالعات حاکی از آن است که کیفیت داده‌ها یک مفهوم چندبعدی است (پپینو^۲ و دیگران ۲۰۰۲). از این‌رو برای درک بهتر، ابعادی برای اندازه‌گیری و بهبود کیفیت داده مشخص شده است (وزیری و دیگران ۲۰۱۷). در این بخش از کتاب از ابعاد کیفیت داده ارائه شده توسط وانگ و استرانگ در سال ۱۹۹۶ استفاده شده است. این ابعاد شامل ۱۵ بعد هستند که در ۴ دسته زیر طبقه‌بندی می‌شوند:

۱. ذاتی

۲. دسترس‌پذیری

۳. متنی

۴. نمایشگری

ابعاد نمایش سازگاری و کامل بودن، دو بعد مهم در میان ابعاد کیفیت داده‌ها و از شروط اساسی برای زیرساخت فناوری اطلاعات و منبع کسب‌وکار بزرگ به شمار می‌آیند و همچنین دارایی بسیار مهمی برای دستیابی به مزایای رقابتی بلندمدت است (کوان^۳ و دیگران ۲۰۱۴).

برای اندازه‌گیری ابعاد کیفیت داده‌ها، ابزارها و یا معیارهایی مشخص شده است

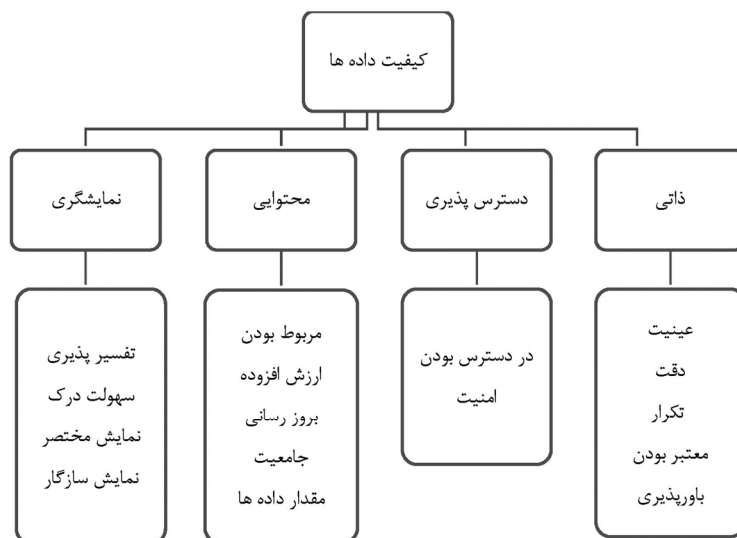
1. Moges

2. Pipino

3. Kwon

که این معیارها یا به صورت ذهنی^۱ و یا به صورت عینی^۲ هستند. معیارهای ذهنی براساس نظرات و تجربیات کاربران داده اعم از مدیران و یا متخصصان داده مشخص می شوند. سنجش ذهنی، معمولاً با استفاده از مصاحبه هایی که مربوط به داده ها و یا پرسشنامه ها است صورت می پذیرد و منعکس کننده نیازها و تجارب ذینفعان از جمله گردآورندگان، نگهبانان و مصرف کنندگان داده می باشد. از سوی دیگر، سنجش عینی بر مبنای فرمول های ریاضی است که برای اندازه گیری یک مجموعه داده استفاده می شود (وزیری و دیگران ۲۰۱۷).

نخستین دسته بندی صورت گرفته در حوزه کیفیت داده ابعادی است که توسط وانگ و استرانگ در سال ۱۹۹۶ ذکر شده است. این ابعاد عبارت اند از: دقت، عینیت، باورپذیری، معتبر بودن، مربوط بودن، ارزش افزوده، بروز بودن، جامعیت، مقدار داده ها، تفسیرپذیری، سهولت درک، نمایش مختصر، نمایش سازگار، در دسترس بودن و امنیت. که این ابعاد به طور مشخص در ۴ دسته با نام های ذاتی، متنی، نمایشگری و دسترس پذیری در قالب شکل زیر قرار می گیرند.



شکل ۱-۲ دسته بندی ابعاد (استرانگ و دیگران ۱۹۹۷)

در حال حاضر با گسترش فن‌آوری اطلاعات و ارتباطات و استفاده از آن در تمامی زمینه‌ها، با انبوه زیاد داده‌ها فرایندهای سازمانی مواجهیم (کارلو^۱ و دیگران ۲۰۱۱). حجم بالای داده‌ها و یا فراداده‌ها باعث می‌شود تا کیفیت داده‌ها به‌عنوان مسئله‌ای دارای اهمیت مطرح شود. از طرفی کیفیت پایین داده‌ها خود نیز می‌تواند باعث بروز مشکلات زیادی شود. کیفیت داده به سازمان‌دهی سیستم اطلاعاتی و چگونگی پردازش داده بستگی زیادی دارد و اندازه‌گیری و بهبود کیفیت داده در سازمان‌ها کاری پیچیده است (خسروانجوم و همکاران ۲۰۱۱). نخستین پژوهش‌ها در حوزه مدیریت کیفیت جامع داده‌ها^۲ به دهه ۱۹۸۰ میلادی بازمی‌گردد. پژوهش‌های اولیه در حوزه کیفیت داده‌ها به‌طور عمده بر توسعه شیوه‌ها برای پرس‌وجو از منابع اطلاعاتی متعدد و ایجاد مخازن اطلاعاتی بزرگ متمرکز بود. تحقیق وانگ و مادنیک^۳ در سال ۱۹۸۹ رویکرد سیستماتیک را برای بررسی کیفیت داده‌های مرتبط بکار می‌گیرد. این کار بیشتر هنگام تفکیک هویت اطلاعات و یکپارچه‌سازی اطلاعات از منابع مختلف و با حضور رکوردهایی که تداخل و همپوشانی دارند شناخته‌شده است. این پژوهشگران برای تعیین هویت رکوردهای جداشده روش‌هایی را پدید آوردند. این مسئله با واژه‌هایی مانند اتصال رکورد، غیرتکراری کردن رکورد و تطبیق رکورد شناخته می‌شود. پس از آن، آن‌ها یک مدل پلی‌ژن^۴ را توسعه دادند تا پردازش برچسب‌های منبع داده را در پردازنده پرس‌وجو زیر نظر بگیرند و به این طریق پاسخی برای پرسش‌های مرتبط با کیفیت داده‌ها بیابند (وانگ و مادنیک ۱۹۹۰). آن‌ها در ادامه پژوهش‌هایشان با تمرکز بر پرسش‌هایی مانند «این داده‌ها از کجا آمده‌اند؟» و «کدام منابع داده‌ای میانجی برای رسیدن به این داده‌ها مورداستفاده قرار گرفته است؟» به توسعه یک روش مدل‌سازی بر اساس مدل، رابطه، هویت و کیفیت پرداختند (ونگ و دیگران ۱۹۹۳) تا بتوانند در مرحله طراحی پایگاه داده مفهومی به جمع‌آوری روش‌مند معیارهای جامع کیفیت داده‌ها به‌عنوان فراداده پردازد (استوری^۵ و یانگ ۱۹۹۸). در این روش پردازنده پرس‌وجو با استفاده از جبر رابطه‌ای بسط یافته، فراداده‌ی کیفیت داده‌های سلسله مراتبی را پردازش می‌کند (ونگ و دیگران ۱۹۹۳).

1. Carlo

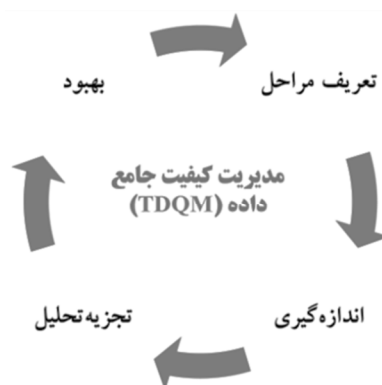
3. Wang, Madnick

2. Total Data Quality Management, TDQM

4. Polygen

5. Storey

این جریان پژوهشی، بر روی تحقیق و طراحی پایگاه داده مدرن، مانند منشأ و تبار داده‌ها که بانمن^۱ و دیگران در سال ۲۰۰۱ ارائه نمودند و همچنین بر روی سایر زمینه‌های مرتبط با جبر رابطه‌ای برای امنیت و مدیریت حفظ حریم خصوصی داده‌ها اثرگذار بوده است. نتایج اولیه تلاش‌های پژوهشی، پژوهشگران را برای آغاز پژوهش‌های روشمند در تمام گستره مسائل کیفیت داده‌ها تشویق کرد که به نوبه خود منجر به افتتاح برنامه مدیریت کیفیت داده جامع در اوایل ۱۹۹۰ در دانشگاه ام‌آی‌تی^۲ شد. تحقیقات اولیه در برنامه توسعه این چارچوب، بهبود کیفیت مستمر داده‌ها را با دنبال کردن مراحل تعریف، اندازه‌گیری، تجزیه و تحلیل و بهبود حمایت می‌کند (ونگ و مندیگ^۳ ۱۹۹۲). این مدل، چارچوب مدیریت کیفیت جامع^۴ را برای بهبود کیفیت در حوزه تولید به حوزه داده توسعه می‌دهد. دمینگ در سال ۱۹۸۲ و جوران و جوفری در سال ۱۹۹۹ از مهم‌ترین پایه‌گذاران و توسعه‌دهندگان این حوزه بودند. بینش کلیدی این بود، اگرچه داده در واقع یک محصول تولیدشده یا محصول جانبی توسط بسیاری از سازمان‌ها است، ولی آن‌طور که باید با آن رفتار نشده و مورد مطالعه قرار نگرفته است. چهار مرحله اصلی رویکرد صنعتی مدیریت کیفیت جامع داده در شکل ۱-۳ قابل مشاهده است. در ادامه به پژوهش‌های صورت گرفته در حوزه نظریه‌ها، روش‌ها، و شیوه‌ها برای چهار مرحله از این چارچوب توسعه خواهیم پرداخت.



شکل ۱-۳ چهار مرحله اصلی در مدیریت کیفیت جامع داده

1. Buneman
2. Madnick

3. Massachusetts Institute Of Technology
4. Total Quality Management, TQM

تعریف مراحل

تعریف کیفیت داده‌ها از نقطه نظر مصرف‌کننده، دستیابی به یک موفقیت بزرگ به شمار می‌رود. مطابق با این تعریف تناسب بین استفاده و شناسایی ابعاد کیفیت داده‌ها از طریق یک مطالعه پیمایشی چندمرحله‌ای روشمند عملی شد (ونگ و استورم ۱۹۹۶). ابعاد کلیدی کیفیت داده‌ها با استفاده از تحلیل عاملی بر روی بیش از ۱۰۰ نشانه مشخص کیفیت داده‌ها از طریق مطالعه پیمایشی کشف شد. این ابعاد به صورت چهار طبقه کیفیت داده یعنی دسترس‌پذیری، محتوایی، نمایشگری و ذاتی سازمان‌دهی شده‌اند. در ادامه و در زیربخش بعدی به مرحله بعدی از مدل پیشگفته یعنی اندازه‌گیری اشاره خواهیم کرد.

اندازه‌گیری

لی و دیگران در سال ۲۰۰۲ دومین مرحله از رویکرد مدیریت کیفیت جامع یعنی ارزیابی و اندازه‌گیری کیفیت داده‌ها را مشابه شرایط عملیاتی به منظور اندازه‌گیری کیفیت داده‌ها در سازمان‌ها توسعه دادند (لی و دیگران ۲۰۰۲). این ابزار هر یک از ابعاد معرفی‌شده در انتهای بخش قبل را به چهار پنج شاخص قابل اندازه‌گیری تقلیل می‌دهد و صورت‌های کاربردی مناسب برای این آیتم‌ها بکار گرفته می‌شوند تا هر یک از ابعاد را ارزیابی کنند. برای مطالعه بیشتر تحقیق صورت گرفته توسط پپینو، یانگ و ونگ^۱ در سال ۲۰۰۲ را ملاحظه کنید. این ابزار همچنین این قابلیت را دارد که با نیازهای خاص هر سازمانی منطبق گردد.

تجزیه و تحلیل

نتایج اندازه‌گیری در این مرحله تفسیر می‌گردند (لی و دیگران ۲۰۰۲). شیوه‌های تجزیه و تحلیل شکاف تفاوت‌های ادراکی بین ابعاد داده‌ها و نقش‌های از قبل تعیین شده در حوزه مربوط به کیفیت داده را نشان می‌دهد (استرانگ، لی و وانگ ۱۹۹۷). لی و استرانگ در سال ۲۰۰۴ سه نقش عمده جمع‌آوری‌کننده‌های اطلاعات،

1 Pipino, Yang And Wang

متولیان داده‌ها، و مصرف‌کنندگان داده را در این حوزه معرفی نمودند (لی و استرانگ ۲۰۰۴). تجزیه و تحلیل‌های صورت گرفته در این مرحله ابعادی که بیشتر نیاز به بهبود دارند را شناسایی کرده و به تحلیل ریشه مشکلات کیفیت داده می‌پردازد. در زیربخش بعدی به معرفی مرحله آخر یعنی بهبود خواهیم پرداخت.

بهبود

در این مرحله، اقدامات صورت گرفته در راستای تغییر ارزش داده‌ها و نیز تغییر فرایندهایی است که داده را تولید می‌کنند. روش دوم همان‌گونه که در مطالعات بالو و دیگران (۱۹۹۸) و وانگ و دیگران (۱۹۹۸) مورد بحث قرار داده‌اند، مؤثرتر است. در این روش مراحل به‌سوی مدیریت اطلاعات به‌عنوان یک محصول فراهم شده است. علاوه بر این، فناوری‌های قبلی مانند پلی‌ژن و مدل رابطه هویت کیفیت، به‌عنوان بخشی از فرایند بهبود مستمر اعمال می‌شود.

علاوه بر توسعه و غنی‌سازی چارچوب مدیریت جامع کنترل داده، برنامه مدیریت جامع کنترل داده موجب یکپارچه‌سازی حوزه کیفیت داده شده است. همچنین باعث گسترش تأثیر پژوهش و ترویج همکاری دولت با دانشگاه و صنعت از طریق نشریات، سمینارها، دوره‌های آموزشی و همایش‌های بین‌المللی سالانه در حوزه کیفیت اطلاعات^۱ از سال ۱۹۹۶ به بعد شده است. دانشگاه آرکانزاس^۲ در شهر لیتل راک^۳ با کمک برنامه مدیریت جامع کنترل داده‌ها برای اولین بار در ایالات متحده، دوره کارشناسی‌ارشد و دکتری کیفیت داده ایجاد کرده است. ایجاد این دوره‌ها به‌منظور پاسخگویی به افزایش تقاضا و نیاز به متخصصان کیفیت داده آموزش‌دیده در سطح بالا و نیز آماده‌سازی دانشجویان برای پژوهش‌های پیشرفته در زمینه کیفیت داده بوده است. در بخش صنعت و در حوزه کیفیت داده‌ها، فروشندگان اصلی نرم‌افزار شروع به پیاده‌سازی فناوری‌های کیفیت داده در ارائه محصول و خدمات خود کرده‌اند. در بخش دولتی، کیفیت داده به بخشی مهم در بسیاری از

1. International Conference On Information Quality

2. University Of Arkansas

3. Little Rock

بخش‌های دولت الکترونیک و طرح‌های معماری سازمانی تبدیل شده است. در این میان، پژوهش‌های کیفیت داده‌ها با چالش‌های جدیدی سر و کار دارند مانند محیط‌های کسب‌وکار همیشه در حال تغییر، الزامات قانونی، انواع افزایش فرم‌های داده‌ها/رسانه‌ها، و فناوری‌های اینترنت و این‌که اساساً چگونه اطلاعات تولید، ذخیره، دستکاری، و مصرف می‌شود. پژوهش‌های کیفیت داده‌ها که از دو دهه پیش شروع شده، وارد عصر جدیدی شده است که در آن تعداد فزاینده‌ای از پژوهشگران به‌طور فعال به‌منظور افزایش درک از مشکلات کیفیت داده و توسعه راه‌حل برای مسائل نوظهور در این حوزه تلاش می‌کنند.

وانگ و دیگران در سال ۱۹۹۵ چارچوب مدونی بر مبنای الزامات استاندارد ایزو ۹۰۰۱ به‌منظور دسته‌بندی پژوهش‌های کیفیت داده ارائه دادند. چارچوب دسته‌بندی پژوهش‌های آن‌ها شامل گروه‌بندی زیر است:

۱. مسئولیت مدیریت

۲. هزینه‌های تضمین و عملیات

۳. تحقیق و توسعه

۴. تولید

۵. توزیع

۶. مدیریت کارکنان

۷. عملکرد قانونی

نلسون^۱ و دیگران در سال ۲۰۰۵ عواملی را که به‌طور هم‌زمان بر کیفیت سیستم و اطلاعات تأثیرگذار هستند را موردبررسی قرار داده و نتایج را در قالب مطالعه موردی بر روی سازمان‌هایی استفاده کردند که انباشت داده دارند. بیاگی و دیگران در مرکز پژوهش‌های کشور ایتالیا کنترل کیفیت در سیستم‌های اطلاعاتی تحقیقاتی و عوامل مؤثر بر آن را موردبررسی قرار داده‌اند (بیاگی و دیگران ۲۰۱۴).

1. Nelson

فصل دوم

مبانی کیفیت داده و اطلاعات

۲-۱. مقدمه

به علت پیشرفت داده‌های الکترونیک، کیفیت داده‌ها نقش مهمی در تمامی کسب‌وکارها و برنامه‌های دولتی ایفا می‌کند. کیفیت داده‌ها به‌عنوان یک بخش عملکردی و ضروری در ارتباط با فرایندهای عملیاتی (موسسه داده انبارداری^۱ ۲۰۰۶)، تصمیم‌گیری‌ها (چنگالور اسمیث^۲ و دیگران ۱۹۹۹) و همکاری‌های درون‌سازمانی (اسکاناپیکو و بتینی^۳ ۲۰۰۲) شناخته شده است. چندین طرح نوآورانه در بخش‌های دولتی و خصوصی راه‌اندازی شده است که در آن کیفیت داده‌ها نقش برجسته‌ای دارد. از جمله این طرح‌ها می‌توان به تصویب قانون کیفیت داده‌ها توسط دولت ایالات متحده در سال ۲۰۰۲ (دفتر مدیریت و بودجه^۴ ۲۰۰۶) و تصویب چارچوب کیفیت داده‌ها توسط دولت ولز^۵ در سال ۲۰۰۴ اشاره کرد که هدف آن‌ها بهبود کیفیت اطلاعات در حوزه پزشکی عمومی بود (مجله کیفیت اطلاعات و داده‌ها^۶ ۲۰۰۶).

با حرکت سیستم‌های اطلاعاتی از ساختار سلسله‌مراتبی و یکپارچه به ساختار مبتنی بر شبکه، مجموعه منابع داده‌ای بالقوه مورد استفاده در سازمان‌ها، از نظر اندازه

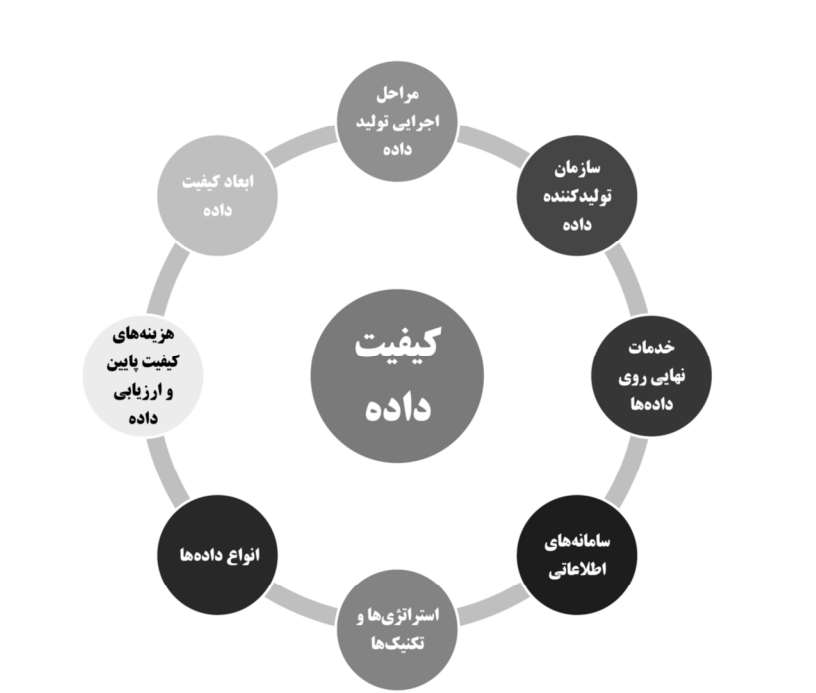
1. Data Warehousing Institute
3. Scannapieco And Batini
5. Wales

2. Chengalur-Smith
4. Office Of Management And Budget
6. Data Quality Information

و دامنه به‌طور چشمگیری افزایش یافت. مسئله کیفیت داده‌ها به دلیل تأثیرپذیری از این تکامل، پیچیده‌تر و بحث‌برانگیزتر شده است. در سیستم‌های اطلاعات شبکه‌ای، فرایندها در تبادل اطلاعات پیچیده‌ای درگیر شده و اغلب براساس ورودی به‌دست‌آمده از منابع خارجی عمل کرده و این ورودی‌ها نیز اغلب قابل مقایسه نیستند. در نتیجه اگر کیفیت هر فرایند و ورودی اطلاعات کنترل شده نباشد، کیفیت کلی داده‌هایی که در سراسر سیستم اطلاعات جریان دارد در طول زمان به‌سرعت کاهش می‌یابد. از سوی دیگر، سیستم‌های اطلاعات شبکه‌ای، فرصت‌های جدیدی را برای مدیریت کیفیت داده‌ها ارائه می‌دهند که می‌توان دسترسی به طیف وسیع‌تری از منابع داده‌ها، توانایی انتخاب و مقایسه اطلاعات از منابع مختلف به‌منظور تشخیص و تصحیح خطاها و بدین ترتیب، بهبود کلی کیفیت داده‌ها را از نتایج آن دانست.

اثر حاضر طیف گسترده‌ای از شیوه‌های ارزیابی و بهبود کیفیت داده‌ها را برای اموری مانند پیوند مدرک و سابقه، قوانین کسب‌وکار و مقیاس‌های هماهنگ ارائه می‌دهد. باگذشت زمان، این شیوه‌ها برای مقابله با افزایش پیچیدگی کیفیت داده‌ها در سیستم‌های اطلاعات شبکه‌ای تکامل یافتند. باتوجه به تنوع و پیچیدگی این شیوه‌ها، پژوهش‌های اخیر بر روی تعریف‌های روش‌شناختی که به انتخاب، سفارشی‌سازی و اعمال شیوه‌های ارزیابی و بهبود کیفیت داده‌ها کمک می‌کنند، متمرکز شده‌اند. کتاب حاضر روش کیفیت داده را به‌عنوان مجموعه‌ای از دستورالعمل‌ها و شیوه‌هایی تعریف می‌کند که با توصیف اطلاعات ورودی در یک زمینه کاربردی شروع شده و فرایند منطقی را برای ارزیابی و بهبود کیفیت داده‌ها بیان می‌کند. هدف این بخش، ارائه یک توصیف منظم و مقایسه‌ای در مورد روش‌های کیفیت داده‌های موجود است.

در متون علمی مختلف ابعاد مختلفی مدیریت داده که به کیفیت داده مرتبط خواهند شد مشاهده می‌گردد. شکل ۱-۲ ابعاد اصلی این حوزه را نمایش می‌دهد که در ادامه این فصل موردتوجه قرار خواهند گرفت.



شکل ۲-۱ حوزه‌های اصلی کیفیت داده

۲-۲. گام‌های تجزیه و تحلیل روش‌ها

این بخش به ترتیب شکل زیر تدوین شده و شامل زیربخش‌های زیر است. در ابتدا در مورد کیفیت داده‌ها توضیحاتی بیان می‌شود. در بخش بعدی موضوعات اصلی مشترک در کیفیت داده برای تمامی روش‌ها بیان می‌شود. دیدگاه‌های مورد استفاده در این بخش برای تحلیل‌های مقایسه‌ای به صورت زیر می‌باشند:

۱. مراحل و گام‌های روش‌شناختی
۲. استراتژی‌ها و شیوه‌ها
۳. ابعاد کیفیت داده
۴. انواع داده و سرانجام
۵. انواع سیستم‌های اطلاعاتی

در فصل دوم، روش‌شناسی‌های موجود در راستای ابعاد معرفی‌شده در فصل سوم مقایسه می‌شوند. مقایسه با ارزیابی‌های کیفی و در قالب جدول‌های مختلفی صورت می‌پذیرد. درنهایت، این بخش با شرح مختصری از هر یک از روش‌ها به پایان می‌رسد که موارد زیر را توصیف می‌کند:

۱. مراحل هر یک از روش‌ها، وابستگی متقابل و تصمیم‌گیری‌های حیاتی هر روش؛

۲. توصیفی کلی از تمرکز هر روش و مشارکت اصلی آن برای فرایند ارزیابی و بهبود کیفیت داده‌ها؛

۳. کاربرد هر یک از این روش‌ها.

در فصل چهارم نیز به بررسی کاربرد مفاهیم کیفیت داده در علم و فناوری پرداخته می‌شود.

۲-۳. مقایسه دیدگاه‌ها در تجزیه و تحلیل روش‌ها

دیدگاه‌های مختلفی برای تجزیه و تحلیل و مقایسه روش‌های کیفیت داده‌ها وجود دارند. برخی از آن‌ها عبارتند از:

۱. مراحل و گام‌هایی که روش را تشکیل می‌دهند؛

۲. استراتژی و شیوه‌هایی که در روش ارزیابی و بهبود سطح کیفیت داده اتخاذ شده‌اند؛

۳. ابعاد و معیارهایی که در این روش‌ها برای ارزیابی سطح کیفیت داده‌ها انتخاب شده‌اند؛

۴. انواع هزینه‌هایی که با مسائل مربوط به کیفیت داده‌ها در ارتباط هستند از جمله:

۱.۴. هزینه‌های مربوط به کیفیت داده‌های ضعیف، که شامل هزینه‌های فرایند ناشی از خطای داده‌ها و هزینه‌های فرصتی است که به علت درآمد از دست‌رفته ایجاد شده است. همچنین این هزینه‌ها به هزینه‌های غیرمستقیم نیز اشاره می‌کنند.

- ۲.۴. هزینه‌های فعالیت‌های ارزیابی و بهبود که به‌عنوان هزینه‌های مستقیم نیز نام‌برده می‌شوند.
۵. انواع داده‌هایی که در روش‌ها در نظر گرفته شده‌اند؛
۶. انواع سیستم‌های اطلاعاتی جهت استفاده، اصلاح و مدیریت داده‌ها که در هر روش در نظر گرفته شده است؛
۷. سازمان‌های درگیر در فرایندها که به ایجاد یا به‌روزرسانی داده‌ها پرداخته و با ساختار و اصول خاص خود در نظر گرفته شده‌اند؛
۸. فرایندهایی که داده‌ها را باهدف تولید خدمات موردنیاز کاربران ایجاد و بروز رسانی می‌کنند در روش‌شناسی در نظر گرفته شده‌اند؛
۹. فرایندهای خدماتی در نظر گرفته شده، در روش‌شناسی نیز ایجاد شده‌اند.
- این روش‌ها از نظر نحوه کاربرد در ابعاد فوق متفاوت هستند. در کتاب حاضر سه دیدگاه پایانی یعنی دیدگاه سازمانی، فرایندی و خدماتی بررسی نخواهند شد زیرا منابع غنی در مورد آن‌ها نیست و به‌ندرت کاربردی از آن‌ها در روش‌های مختلف ذکر شده است.

۲-۳-۱. مراحل و گام‌های مشترک

- در حالتی کلی، مراحل انجام یک روش کیفیت داده از سه‌گام زیر تشکیل شده است:
۱. بازسازی وضعیت موجود (که هدف آن جمع‌آوری اطلاعات براساس فرایندها و خدمات سازمانی است). این گام بررسی مجموعه داده‌ها و روندهای مرتبط با مدیریت، مشکلات کیفی و هزینه‌های مربوطه است. در صورتی که اطلاعات قبلی و گزارش‌هایی از تجزیه و تحلیل‌های قبلی در دسترس باشد می‌توان این گام را نادیده گرفت.
 ۲. ارزیابی/اندازه‌گیری (که مجموعه کیفیت داده‌ها را در ابعاد کیفیت مربوطه اندازه‌گیری می‌کند). اصطلاح اندازه‌گیری برای بررسی ارزش مجموعه‌ای از ابعاد کیفیت داده‌ها استفاده می‌شود. اصطلاح ارزیابی نیز هنگامی استفاده

می‌شود که چنین اندازه‌گیری‌هایی با مقادیر مرجع مقایسه شوند تا قادر به شناخت کیفیت داده‌ها باشیم. در این پژوهش عبارت ارزیابی برای اکثر روش‌هایی اتخاذ شده است که در آن‌ها بر اهمیت علل کیفیت داده‌های ضعیف تأکید شده است.

۳. بهبود فرایند که مربوط به انتخاب مراحل، استراتژی‌ها و شیوه‌ها برای رسیدن به اهداف کیفیت داده‌های جدید است.

اگر گام ارزیابی براساس اسناد و مدارک موجود باشد، انجام مجدد آن اختیاری است. از آنجاکه معمولاً این فرض در مورد داده‌ها وجود دارد، در مورد گام اول یعنی بازسازی وضعیت موجود بحث نمی‌کنیم. معمولاً مراحل ارزیابی و بهبود را در مجموعه‌ای مشترک از گام‌های اساسی سازمان‌دهی می‌کنند. هرچند برای هر مرحله نام‌های مختلفی اتخاذ شده است اما به‌طور کلی گام‌های مرحله ارزیابی عبارت‌اند از:

- تجزیه و تحلیل داده‌ها که به بررسی طرح داده‌ها و انجام مصاحبه برای رسیدن به درک کاملی از داده‌های مرتبط و قوانین موجود در معماری و مدیریت داده‌ها می‌پردازد؛
- تجزیه و تحلیل الزامات کیفیت اطلاعات که به بررسی نظر کاربران داده‌ها و مدیران برای شناسایی مشکلات کیفی و تعیین اهداف جدید می‌پردازد؛
- شناسایی مناطق بحرانی که به انتخاب و ارزیابی کمی مرتبط‌ترین پایگاه داده و جریان داده می‌پردازد؛
- مدل‌سازی فرایند که مدلی از فرایندهای تولید و یا به‌روزرسانی داده‌ها را فراهم می‌کند؛
- اندازه‌گیری کیفیت که براساس مشکلات کیفی مشخص شده و اثرگذار در مرحله تجزیه و تحلیل الزامات کیفیت اطلاعات، ابعاد کیفیتی انتخاب شده و معیارهای مربوط به آن‌ها تعریف می‌شوند. هنگامی که این مسائل مبتنی بر معیارهای کمی باشند، اندازه‌گیری می‌تواند عینی باشد. اما هنگامی که این

مسائل مبتنی بر ارزیابی‌های کیفی توسط مدیران داده‌ها و کاربران باشند، اندازه‌گیری می‌تواند ذهنی باشد.

مراحل گام بهبود عبارت‌اند از:

- بررسی هزینه‌ها که هزینه‌های مستقیم و غیرمستقیم کیفیت داده‌ها را برآورد می‌کند؛
 - تعیین مسئولیت‌های فرایند که به شناسایی صاحبان فرایند و تعریف مسئولیت آن‌ها براساس فعالیت‌های تولید و مدیریت داده‌ها می‌پردازد؛
 - تعیین مسئولیت داده‌ها که به شناسایی صاحبان داده‌ها و تعریف مسئولیت‌های مدیریت داده آن‌ها می‌پردازد؛
 - شناسایی علل خطاها که علت مشکلات کیفیت را شناسایی می‌کند؛
 - انتخاب استراتژی‌ها و شیوه‌ها که تمام استراتژی‌های بهبود داده‌ها و شیوه‌های مربوطه را شناسایی کرده و مطابق با دانش موجود، اهداف کیفیت و محدودیت‌های بودجه به بررسی و انتخاب استراتژی می‌پردازد؛
 - طراحی راه‌حل‌های بهبود داده که مؤثرترین و کارآمدترین استراتژی‌ها، مجموعه شیوه‌ها و ابزارهای مرتبط با بهبود کیفیت داده‌ها را انتخاب می‌کند؛
 - کنترل فرایند که نقاط بررسی را در فرایندهای تولید داده‌ها، برای نظارت بر کیفیت آن‌ها در طول اجرای فرایند تعریف می‌کند؛
 - طراحی مجدد فرایند، تعریف اقدامات بهبود فرایند که می‌تواند پیشرفت‌های مربوط به کیفیت اطلاعات را ارائه دهد؛
 - مدیریت بهبود که قوانین سازمانی جدید را برای کیفیت داده‌ها تعریف می‌کند؛
 - نظارت بر بهبود که به ارائه فعالیت‌های نظارت دوره‌ای می‌پردازد که بازخوردی از نتایج حاصل از فرایند بهبود و تنظیم پویای آن را فراهم کند.
- در بخش‌های بعدی روش‌های پیشگفته باتوجه به توانایی‌های ارزیابی و بهبود آن‌ها

در طول مجموعه‌ای از مراحل و گام‌های معرفی شده مقایسه می‌شوند. شایان ذکر است که معمولاً هر یک از روش‌ها به یک قابلیت ارزیابی و یا بهبود خاص در شرایط مختلف مرتبط هستند.

۲-۳-۲. استراتژی‌ها و روش‌ها

در مراحل بهبود، هر یک از روش‌های معرفی شده می‌توانند دو نوع استراتژی کلی اتخاذ کنند:

۱. استراتژی داده محور^۱

۲. استراتژی فرایند محور^۲

استراتژی‌های داده محور، کیفیت داده‌ها را با اصلاح مستقیم ارزش داده‌ها بهبود می‌بخشد. برای نمونه، کیفیت داده‌های منسوخ شده با به‌روزرسانی آن‌ها از یک پایگاه داده دیگر و جایگزینی آن‌ها با داده‌های به‌روز شده بهبود می‌یابد. استراتژی‌های فرایند محور نیز کیفیت را توسط طراحی مجدد فرایندهای تولید و تغییر داده بهبود می‌بخشند. برای مثال، می‌توان با گنجاندن فعالیت‌هایی که فرمت داده‌ها را قبل از ذخیره‌سازی کنترل می‌کنند، یک فرایند را دوباره طراحی کرد.

استراتژی‌های داده محور و فرایند محور، انواع مختلفی از شیوه‌ها مانند الگوریتم‌ها، فناوری هوشمند و فعالیت‌های مبتنی بر دانش را اعمال می‌کنند که هدف آن‌ها بهبود کیفیت داده است. شیوه‌های بهبود مرتبط با استراتژی‌های مبتنی بر داده عبارتند از:

۱. دستیابی به داده‌های جدید که باعث بهبود داده‌های قدیم و کسب کیفیت

بالاتر از آن‌ها می‌شود. این شیوه بیشتر به‌جای روش‌هایی مورد استفاده قرار

می‌گیرد که موجب مشکلات کیفی در داده‌ها می‌شوند؛

۲. استانداردسازی^۳ (یا نرمالیزه کردن) که مقادیر داده‌های غیراستاندارد را با

مقادیری که با استاندارد مطابقت دارد جایگزین یا کامل می‌کند. برای

1. Data-Driven Strategy

2. Process-Driven Strategy

3. Standardization

نمونه، نام مستعار با نام اصلی جایگزین می‌شود مانند باب با رابرت و عبارات اختصاری متناظر که با نام کامل جایگزین می‌شوند نظیر عبارت Channel Str که بایستی با Channel Street جایگزین شود؛

۳. پیوند سابقه‌ها^۱ که نمایش اطلاعات را در دو یا چند جدول را شناسایی

می‌کند که امکان دارد به موجودیتی یکسان در دنیای واقعی اشاره کنند؛

۴. یکپارچه‌سازی داده‌ها و طرح‌ها که یک دیدگاه یکپارچه را از اطلاعات

ارائه شده توسط منابع داده‌های ناهمگن تعریف می‌کند. هدف اصلی ادغام

و یکپارچه‌سازی، دسترسی یک کاربر به داده‌های ذخیره شده توسط منابع

داده ناهمگن و از طریق دیدگاه یکپارچه این داده‌ها است.

سیستم‌های اطلاعاتی توزیعی، تعاونی، دوطرفه (بخش ۲-۲-۶) و منابع داده در ارتباط

با انواع حالت‌ها و ناهمگنی‌ها را می‌توان به‌طور کلی به موارد زیر دسته‌بندی کرد:

۱.۴. ناهمگنی در فناوری

ناهمگنی فناوری در حالت استفاده از محصولات داده‌ای توسط

تأمین‌کنندگان مختلف، در لایه‌های گوناگونی از زیرساخت‌های اطلاعاتی

و ارتباطاتی ایجاد می‌شوند.

۲.۴. ناهمگنی در طرح^۲

ناهمگنی در طرح به دو صورت ایجاد می‌شوند:

۱.۲.۴. در درجه اول به‌خاطر استفاده از مدل‌های داده‌های مختلف

ایجاد می‌شود مانند منبعی که مدل داده رابطه‌ای را اتخاذ کرده و

منبع‌های مختلفی که مدل داده‌های XML را پذیرفته‌اند.

۲.۲.۴. نمایش‌های مختلف برای یک شی یکسان مانند دو منبع رابطه‌ای

که یک شی را به‌عنوان یک جدول و یک ویژگی نشان می‌دهند.

۳.۴. ناهمگنی در سطح نمونه

1. Record Linkage

2. Shema Heterogeneities

ناهمگنی در سطح نمونه توسط مقادیر داده مختلف و متضاد ایجاد می‌شود که توسط منابع مجزا برای موجودیت‌های مشابه ارائه می‌شود. برای نمونه، این نوع عدم تجانس یا ناهمگنی را می‌توان در فرایندهای مستقل و ناهماهنگی مشاهده کرد که منابع داده مختلف را تغذیه می‌کنند. یکپارچه‌سازی داده‌ها باید با تمام این ناهمگنی‌های ذکر شده روبرو شود.

۵. اعتماد به منبع که منابع داده را براساس کیفیت داده‌های هر منبع انتخاب می‌کند.

۶. شناسایی و تصحیح خطا که داده‌های کثیف^۱ را با شناسایی رکوردهایی که الزامات کیفی را برآورده نکرده‌اند کشف و از بین می‌برد. این شیوه‌ها عمدتاً در حوزه‌های آماری مورد مطالعه قرار می‌گیرند. در مقایسه با داده‌های ابتدایی، داده‌های آماری مانند میانگین، مجموع، حداکثر و مانند آن حساسیت کمتری نسبت به شناسایی احتمالی نادرست و اصلاح مقادیر دارند. شیوه‌هایی برای شناسایی و تصحیح خطا در جهت رفع تناقضات، اطلاعات ناقص و پرت ارائه شده است (داسو جانسون^۲ ۲۰۰۳) (باتینی اسپاناپیکو^۳ ۲۰۰۶)

۷. بهینه‌سازی هزینه‌ها، اقدامات بهبود کیفیت را همراه با مجموعه‌ای از ابعاد برای به حداقل رساندن هزینه‌ها تعریف می‌کند.

دو شیوه اصلی برای مشخص کردن استراتژی‌های فرایند محور عبارت‌اند از:

- بررسی و کنترل روندهای کنترل فرایند تولید داده‌ها هنگامی که:

۱. داده‌های جدید ایجاد می‌شوند؛

۲. مجموعه داده‌ها به‌روز می‌شوند؛

۳. مجموعه داده‌های جدید با روندی مشخص در دسترس قرار می‌گیرند.

در این روش، از استراتژی واکنشی برای اصلاح داده‌ها، اجتناب از تخریب داده‌ها و انتشار خطا استفاده می‌شود.

1. Dirty Data

2. Dasu Johnson

3. Batini Scannapieco

- فرایند طراحی مجدد، به طراحی دوباره فرایندها، حذف علل پایین بودن کیفیت و معرفی فعالیت‌های جدید می‌پردازد که نهایتاً منجر به تولید داده‌هایی باکیفیت بالاتر می‌شود. اگر فرایند طراحی مجدد اساس کار باشد، این شیوه به مهندسی مجدد فرایند کسب‌وکار اشاره می‌کند (هامر چمپی^۱ ۲۰۰۱) (استویکا و همکاران^۲ ۲۰۰۳).

در مطالعه ردمن (۱۹۹۶) چند شیوه معمولی از استراتژی‌های داده محور و فرایند محور با بحث در مورد بهبود ابعاد کیفیتی مختلف و هزینه اجرای هر روش مقایسه شده‌اند. این مقایسه در دیدگاه‌های کوتاه‌مدت و بلندمدت انجام شده است. این مقایسه‌ها بر موارد زیر تمرکز دارند:

۱. جمع‌آوری داده‌های جدید؛
۲. ارتباط رکوردها؛
۳. شناسایی و تصحیح خطاها؛
۴. کنترل فرایند؛
۵. شیوه‌های طراحی مجدد فرایند.

به‌طورکلی، شیوه‌های فرایند محور عملکرد بهتری نسبت به شیوه‌های داده محور در درازمدت دارند زیرا علل ریشه‌ای مشکلات کیفیت را از بین می‌برند. اما از دیدگاه کوتاه‌مدت، طراحی مجدد فرایند می‌تواند بسیار گران باشد (ردمن ۱۹۹۶؛ انگلیش^۳ ۱۹۹۹). در مقابل، استراتژی‌های داده محور در کوتاه‌مدت به‌صرفه هستند اما در بلندمدت گران خواهند شد. آن‌ها برای کاربردهای تک زمانه مناسب‌اند، به همین دلیل برای داده‌های ایستا و مقطعی پیشنهاد می‌شوند.

۲-۳-۳- معرفی ابعاد کیفیت داده

در تمام روش‌های موجود در حوزه کیفیت داده، تعریف ابعاد کیفیت و معیارهای ارزیابی داده‌ها، فعالیتی اساسی است. به‌طورکلی، معیارهای مختلفی را

می‌توان با هر یک از ابعاد کیفیت مرتبط کرد. در برخی موارد، معیار منحصربه‌فرد است و لازم است معیار و بعد مربوط به آن هم به صورت عملیاتی و هم نظری مورد توجه قرار گیرد. به همین دلیل، میان تعاریف نظری و عملیاتی ابعاد تنها زمانی که معیارهای متعددی فراهم باشند، تفاوت گذاشته می‌شود.

ابعاد کیفیت داده‌ها در این تحقیق به مفاهیم مرتبط با ذات و ارزش داده‌ها در سیستم می‌پردازد. اگرچه کیفیت داده‌ها موارد شماتیکی و جنبه‌های ظاهری داده را نیز می‌تواند در برگیرد، اما پرداختن به این جنبه از داده‌ها رویکردهای دیگری از پژوهش را می‌طلبد که از حوزه این اثر خارج است.

پژوهش‌های مربوط به کیفیت داده، طبقه‌بندی کاملی از ابعاد کیفیت داده‌ها را فراهم می‌کنند. با اینهمه، برخی اختلافات در تعریف مهم‌ترین ابعاد باتوجه به ماهیت تعریف‌شده از کیفیت وجود دارد. شش طبقه مهم از ابعاد کیفیت توسط وند^۱ و ونگ (۱۹۹۶)؛ ونگ و استرانگ (۱۹۹۶)؛ ردمن (۱۹۹۶)؛ جاک^۲ و دیگران (۱۹۹۵)؛ بوی^۳ و دیگران (۲۰۰۱) و ناومن^۴ (۲۰۰۲) ارائه شده است. با تجزیه و تحلیل این طبقه‌بندی، تعریف مجموعه اولیه‌ای از ابعاد کیفیت داده‌ها امکان‌پذیر است مانند دقت، جامعیت، ثبات، و به هنگام بودن که مورد تمرکز بیشتر نویسندگان می‌باشد. با اینهمه، هیچ توافق کلی بر روی تعریف مجموعه‌ای از ابعاد کیفیت داده‌ها و یا معنای دقیق هر یک از ابعاد وجود ندارد. در ادامه برخی از مهم‌ترین ابعاد کیفیت داده ارائه می‌شود.

صحت. تعاریف متعددی برای عبارت ارائه شده است. وانگ و استرانگ (۱۹۹۶) صحت را به صورت «میزان و درجه‌ای از بدون غلط بودن، قابلیت اطمینان و تائید داده‌ها» تعریف کرده‌اند. بالو و پازر^۵ داده‌ها را زمانی صحیح می‌دانند که مقادیر داده ذخیره‌شده در پایگاه داده با مقادیر دنیای واقعی مطابقت داشته باشد. (بالو و پازر^۶ ۱۹۸۵). در مطالعه ردمن (۱۹۹۶)، صحت میزان نزدیکی یک مقدار داده^۷، به برخی مقادیر دیگر^۷ است که درست در نظر گرفته شده باشد.

1. Wand

2. Jarke

3. Bovee

4. Naumann

5. Pazer

6. Ballou And Pazer

به طور کلی دو نوع دقت یعنی دقت نحوی و دقت معنایی را می توان از هم متمایز کرد. روش کیفیت داده تنها دقت نحوی را در نظر گرفته و به عنوان نزدیکی مقدار v ، به عناصر دامنه متناظر، D تعریف می کند. در دقت نحوی به دنبال مقایسه مقدار v با مقدار v' دنیای واقعی آن نیستیم، بلکه علاقه مندیم بررسی کنیم که آیا v یکی از مقادیر D است یا خیر و یا اینکه چقدر به مقادیر D نزدیک است. برای نمونه، $v = \text{'John'}$ از لحاظ دستوری دقیق در نظر گرفته می شود، حتی اگر $v = \text{'John'}$ باشد.

جامعیت. جامعیت یعنی تا چه اندازه مجموعه داده های معلوم شامل توصیف داده مربوط به مجموعه ای از دنیای واقعی هستند.

جدول ۱-۲ پژوهش هایی را بیان می کند که به ارائه تعریف جامعیت پرداخته اند. با مقایسه این تعاریف، می توان توافق قابل توجهی در تعریف انتزاعی جامعیت مشاهده کرد. لازم به ذکر است تعاریف در حوزه هایی که به آن اشاره کرده اند متفاوتند. برای نمونه، حوزه سیستم اطلاعات در پژوهش وند و ونگ (۱۹۹۶)، انبار داده ها در جارك و دیگران (۱۹۹۵)، نهاد در بوی و دیگران (۲۰۰۱) موردنظر هستند.

جدول ۱-۲ تعاریف مربوط به جامعیت در مراجع مختلف

منبع	تعریف
وند و ونگ ۱۹۹۶	توانایی یک سیستم اطلاعاتی برای نشان دادن حالتی معنی دار از یک سیستم دنیای واقعی
ونگ و وند ۱۹۹۶	میزانی که داده ها باتوجه به طول، عرض و دامنه برای حوزه موردنظر کافی باشند
ردمن ۱۹۹۶	درجه ای که در آن مقادیر اقلام اطلاعاتی در یک مجموعه داده های موردنظر قرار می گیرند
جارك و دیگران ۱۹۹۵	درصدی از اطلاعات دنیای واقعی که در منابع داده ها و/یا انبار داده ها وارد شده است
بوی و دیگران ۲۰۰۱	میزانی که اطلاعات اجزاء ضروری ^۱ از یک نهاد ^۲ را شامل می شوند
لیو و چی ۲۰۰۲	نسبتی از مقادیر جمع آوری شده در منابع اطلاعاتی به کل مقادیر دنیای واقعی

معمولاً در حوزه پژوهش مربوط به پایگاه داده رابطه ای^۳، جامعیت به مقادیر خالی^۴ (خالی، صفر) به شدت مرتبط می باشد. واژه خالی معادل واژه داده از دست رفته^۵ است. داده ای که در دنیای واقعی وجود دارد اما در یک مجموعه داده قابل شناسایی و

دسترسی نیست. به‌منظور تعیین بعد جامعیت، درک اینکه چرا این داده ازدست‌رفته است دارای اهمیت است. یک مقدار به چند دلیل از دست می‌رود مانند اینکه وجود دارد اما شناخته‌شده نیست یا به دلیل آن‌که اصلاً وجود ندارد و مشخص نیست که آیا وجود دارد یا خیر (آنتونلیز و آتzeni^۱ ۱۹۹۳).

فرض کنید در شکل ۲-۲، یک فرد با ویژگی‌های نام، نام خانوادگی، تاریخ تولد و ایمیل در نظر گرفته‌شده است. فرض کنید فرد شماره ۲ ایمیل نداشته باشد پس اطلاعات ردیف ۲ کامل محسوب می‌شود (از بعد جامعیت). اما اگر فرد شماره ۳ ایمیل داشته باشد اما در این سطر شناسایی یا درج نشده باشد، اطلاعات این فرد ناقص بوده و از بعد جامعیت دارای نقص است.

ثبات. بعد ثبات به نقض قوانین معنایی^۲ تعریف‌شده بر روی مجموعه‌ای از اطلاعات اشاره دارد. با در نظر گرفتن نظریه رابطه‌ای، محدودیت یکپارچگی با سایر اقلام اطلاعاتی دسته‌ای از قوانین معنایی است. در نظریه رابطه‌ای، دو دسته اساسی مربوط به محدودیت یکپارچگی را می‌توان متمایز کرد. این دو دسته عبارت‌اند از: محدودیت‌های درون رابطه‌ای^۳ و محدودیت‌های بین رابطه‌ای^۴.

محدودیت درون رابطه‌ای طیف وسیعی از مقادیر مجاز را برای دامنه یک ویژگی تعریف می‌کند. در ادامه، به چند مثال اشاره می‌شود. «سن باید بین محدوده ۰ و ۱۲۰ باشد»، و یا «اگر سابقه کاری کمتر از ۳ سال است، پس حقوق می‌تواند بالاتر از ۲۵.۰۰۰ یورو در سال باشد».

محدودیت یکپارچگی بین رابطه‌ای می‌تواند شامل روابط متفاوت باشد. ارتباط میان دانشگاه، رشته و مانند آن از محدودیت‌های میان رابطه‌ای است. برای نمونه، در انتخاب رشته کنکور سراسری در صورتی که فرد حوزه فنی مهندسی را انتخاب کند فقط جدول‌های مربوط به رشته‌های فنی مهندسی برای وی باز شود.

1. Antonellis And Atzeni
3. Intra-Relation Constraints

2. Semantic
4. Inter-Relation Constraints

در نظریه رابطه‌ای، دودسته اساسی مربوط به محدودیت یکپارچگی را می‌توان متمایز کرد، این دودسته عبارت‌اند از:

- محدودیت‌های درون رابطه‌ای
- محدودیت‌های بین رابطه‌ای

ID	Name	Surname	BrithDate	Email	
1	John	Smith	1974/17/3	smith@abc.it	موجود نیست
2	Edward	Mornoe	1967/3/2	NULL	موجود اما غیر قابل شناسایی
3	Anthony	White	1936/1/1	NULL	
4	Marianne	Collines	1955/20/11	NULL	

شکل ۲-۲-۲ مقادیر خالی و جامعیت داده‌ها (بتینی و دیگران ۲۰۰۹)

جدول ۲-۲ تعاریف موجود برای ابعاد مرتبط به زمان

منبع	تعریف
وند و ونگ ۱۹۹۶	به‌هنگام بودن تنها به تأخیر بین تغییر یک حالت دنیای واقعی و اصلاح ناشی از حالت سیستم اطلاعات اشاره دارد.
وند و ونگ ۱۹۹۶	به‌هنگام بودن یعنی تا چه حد سن داده‌ها برای کار در دست مناسب هستند.
ردمن ۱۹۹۶	رواج یعنی اینکه داده چقدر به‌روز است. باوجود اختلاف ناشی از تغییرات مربوط به زمان تا مقدار صحیح، ارزش داده به‌روز است.
جارك و دیگران ۱۹۹۵	رواج داشتن زمانی است که اطلاعات به‌موقع در منابع و/یا انبار داده‌ها وارد شوند. فراریت، دوره زمانی را توصیف می‌کند که اطلاعات آن در دنیای واقعی معتبر باشد.
بوی و دیگران ۲۰۰۱	به‌هنگام بودن دو مؤلفه دارد: سن و فراریت. سن و یا ارز معیار چندسالگی اطلاعات، براساس مدت سنی است که ثبت شده است. فراریت معیار بی‌ثباتی اطلاعات، فرکانس تغییر مقدار برای یک ویژگی نهادی است.
ناثومن ۲۰۰۲	به‌هنگام بودن، متوسط سن داده‌ها در یک منبع است.
لیو و چی ۲۰۰۲	به‌هنگام بودن، میزان به‌روز بودن کافی داده‌ها برای یک کار است.

محدودیت یکپارچگی بین رابطه‌ای می‌تواند شامل ویژگی‌هایی^۱ با روابط متفاوت باشد. برای نمونه، فرض کنید دو جدول با ساختار رابطه‌ای داریم. در جدول I، سه ستون با عنوان نام فیلم، نام کارگردان و سال تولید و در جدول II که به جایزه و اسکار مربوط است. دو ستون با عنوان نام فیلم و سال برنده شدن فیلم را در نظر

1 Attributes

بگیرید. در ساختار رابطه‌ای و در محدودیت بین رابطه‌ای، علی‌رغم اینکه دو ستون از دو جدول مختلف، یکی با عنوان «سال تولید» و دیگری با عنوان «سال برنده شدن» داریم، اما از بعد سازگاری لازم است مقادیر این دو ستون یکسان باشند.

متون گسترده‌ای در مورد پایگاه داده‌های سازگار وجود دارد، مثلاً ارناس^۱ و دیگران (۱۹۹۹) مسائل مربوط به سازگاری در پایگاه‌های داده‌ای رابطه‌ای را موردتوجه قرار می‌دهند و مشکلاتی را که مانع ایجاد یکپارچگی در این پایگاه‌های داده می‌شوند را از بعد سازگاری در کیفیت داده موردبررسی قرار می‌دهند. آن‌ها روشی را برای محاسبه پاسخ‌های سازگار، با حفظ صحت و جامعیت داده‌ها ارائه می‌دهند و محدودیت یکپارچگی را به‌عنوان یک توانمندساز در یکپارچه‌سازی داده‌ها نیز ذکر می‌کنند (کال^۲ و همکاران ۲۰۰۴).

برای نمونه، بعد سازگاری در تحلیل مطالعات میدانی به کمک پرسشنامه از جنبه‌های آماری نیز می‌تواند کاربرد داشته باشد. داده‌های به‌دست‌آمده از پرسشنامه‌های سرشماری دارای ساختار پرسشنامه‌ای هستند. قوانین معنایی با عنوان ویرایش^۳، می‌توانند در طرح پرسشنامه‌ای برای مشخص کردن صحت پاسخ تعریف شوند. چنین قوانینی معمولاً بر پیشگیری از خطا دلالت دارند. برای مثال، یک ویرایش می‌تواند به صورت زیر باشد:

- اگر وضعیت تأهل «ازدواج کرده» باشد، سن نباید کمتر از ۱۴ باشد.
- پس از تشخیص گزارش‌های خطا، عمل تصحیح مقادیر شروع می‌شود (هولت و فلگی^۴ ۱۹۷۶).

ابعاد مربوط به زمان: این ابعاد عبارت‌اند از: رواج^۵، فراریت^۶، و به‌هنگام بودن^۷. یک جنبه مهم از داده‌ها، به‌روزرسانی آن‌ها در طول زمان است. ابعاد مربوط به زمان اصلی پیشنهادی در متون، رواج، فراریت و به‌هنگام بودن می‌باشد. جدول ۲-۳ تعاریف ارائه‌شده در متون را برای این سه بعد زمان مقایسه می‌کند. وند و ونگ

1. Arenas

5. Time Related Dimensions

2. Cal

6. Currency

3. Edit

7. Volatility

4. Holt And Fellegi

8. Timeliness

(۱۹۹۶) و ردمن (۱۹۹۶) تعاریف بسیار مشابهی را برای به هنگام بودن و رواج ارائه می‌دهد. ونگ و استرانگ (۱۹۹۶) و لیو و چی (۲۰۰۲) همان معنی را برای به هنگام بودن در نظر گرفتند، درحالی‌که بوی و دیگران (۲۰۰۱) تعریفی را برای به هنگام بودن از نظر رواج و فراریت ارائه می‌دهند. تعریف ارز توسط بوی و دیگران (۲۰۰۱) بیان‌شده اما متناظر با تعریف به‌هنگام بودن توسط ونگ و استرانگ (۱۹۹۶) و لیو و چی (۲۰۰۲) می‌باشد. این مقایسه نشان می‌دهد که در تعریف انتزاعی ابعاد مربوط به زمان توافق وجود ندارد و معمولاً رواج و به‌هنگام بودن برای اشاره به یک مفهوم استفاده می‌شوند.

۲-۳-۴. هزینه‌ها

هزینه‌ها برابر با مصرف منابعی هستند که در هر روش باتوجه به اثرات داده‌های با کیفیت پایین بر روی فعالیت‌ها در نظر گرفته می‌شوند. بنابراین، هزینه کیفیت داده برابر با مجموع هزینه‌های فعالیت‌های ارزیابی و بهبود کیفیت داده‌ها است که به هزینه برنامه‌های کیفیت داده و هزینه کیفیت داده‌های ضعیف مربوط می‌شود. با تخصیص هزینه به داده‌هایی باکیفیت پایین، می‌توان پس از اجرای یک برنامه آن‌ها را به داده‌هایی با کیفیت مؤثرتر ارتقا داد که معمولاً بین روش‌ها گران‌تر هستند. بنابراین با افزایش هزینه برنامه‌های کیفیت داده، هزینه‌های تخصیص‌یافته برای بهبود کیفیت داده‌های ضعیف کاهش می‌یابد. این کاهش می‌تواند به نفع برنامه کیفیت داده‌ها باشد. هزینه تخصیص‌یافته برای برنامه بهبود کیفیت داده‌ها را می‌توان به‌عنوان یک هزینه پیشگیرانه در نظر گرفت که توسط سازمان‌ها برای کاهش خطای داده‌ها پرداخت می‌شود. این دسته هزینه شامل هزینه تمام گام‌ها و مراحل است که فرایند ارزیابی و بهبود کیفیت داده‌ها را تشکیل می‌دهد.

هزینه‌های بهبود کیفیت پایین داده‌ها را می‌توان به شرح زیر طبقه‌بندی کرد (انگلیش ۱۹۹۹):

۱. هزینه‌های فرایند مانند هزینه‌های مربوط به اجرای دوباره کل فرایند به دلیل اشتباه در داده‌ها؛

۲. هزینه‌های فرصت باتوجه به زیان و درآمد ازدست‌رفته.

هزینه اصلاح کیفیت داده‌های ضعیف برعکس هزینه‌های برنامه کیفیت داده‌ها به شدت وابسته به حوزه و نحوه کاربرد داده می‌باشد. این باعث می‌شود ارزیابی آن دشوار باشد چراکه ارزش داده‌های مشابه و سطح کیفیت مربوطه اثرات مختلفی بسته به کاربر آن‌ها دارد. مثلاً با دریافت اطلاعات منسوخ در سهام توسط یک معامله‌گر فعال ممکن است خسارات اقتصادی قابل توجهی، باتوجه به تصمیمات سرمایه‌گذاری ایجاد شود. در مقابل، یک روزنامه همان اطلاعات منسوخ را دریافت کرده و گزارش‌های تجاری ماهانه‌ای منتشر می‌کند به گونه‌ای که شاید از این طریق کسی زیان اقتصادی نبیند.

۲-۳-۵. انواع داده‌ها

هدف نهایی یک برنامه کیفیت داده، تجزیه و تحلیل داده‌هاست. که در حالت کلی، چنین برنامه‌ای عناصر دنیای واقعی را در شکلی توصیف می‌کند که بتواند با استفاده از روش‌های نرم‌افزاری و طراحی یک شبکه، ذخیره، بازیابی و پردازش شود. در زمینه کیفیت داده بسیاری از نویسندگان به طور ضمنی یا صریح سه نوع داده را شناسایی کرده‌اند:

۱. داده‌های ساختاریافته^۱: مجموعه‌ای ساختاریافته از داده‌ها که توسط ویژگی‌های از پیش تعیین شده در یک دامنه^۲ و چارچوب مشخص توسعه داده شده‌اند. دامنه نشان‌دهنده سطح قابل قبولی از مقادیری است که می‌توان به ویژگی‌ها اختصاص داد و معمولاً با انواع اطلاعات اولیه زبان‌های برنامه‌نویسی، مانند مقادیر عددی یا رشته متن مطابقت دارند. جدول‌های رابطه‌ای و داده‌های آماری رایج‌ترین نوع داده‌های ساختاریافته هستند؛

1. Structured Data

2. Domain

۲. داده‌های ساختار نیافته^۱: مجموعه‌ای از علائم، کلمات و حروف می‌باشند، که معمولاً در زبان طبیعی کدگذاری شده‌اند. نمونه‌های معمولی از داده‌های بدون ساختار، پرسشنامه‌های حاوی سؤالات باز با پاسخ‌های آزاد و یا متن پیام یک ایمیل است؛

۳. داده نیمه ساختاریافته^۲: داده‌هایی که ساختار آن‌ها تا حدودی انعطاف‌پذیر است. داده‌های نیمه ساختاریافته، داده‌های بدون طرح^۳ و یا خود توصیفی^۴ نیز نامیده می‌شوند (ابیتبول^۵ و دیگران ۲۰۰۰) (بونمن^۶ ۱۹۹۷) (کالوانس^۷ ۱۹۹۹).

Patrick	Metzisi	Masai Mara	KE	(۱) داده‌های ساختاریافته
Mr Patrick Metzisi, Born in the Masai Mara region that is part of Kenya				(۲) داده‌های ساختار نیافته
<PersonalData> <name> Patrick Metzisi </name> <BornIn> Masai Mara, Kenya </BornIn> </PersonalData>				(۳) داده‌های نیمه ساختاریافته

شکل ۲-۳ نمایش‌های گوناگون از اطلاعاتی یکسان در دنیای واقعی (بتینی و دیگران ۲۰۰۹)

XML زبان نشانه‌گذاری مشترک استفاده‌شده برای نشان دادن داده‌های نیمه ساختاریافته است. برخی از ویژگی‌های مشترک این دسته از داده‌ها عبارت‌اند از:

۱. داده‌ها می‌توانند شامل زمینه‌های ناشناخته‌ای در زمان طراحی باشند. مثلاً یک فایل XML دارای فایل طرح XML مربوطه نیست؛
۲. همان نوع داده‌ها ممکن است از راه‌های متعددی نشان داده شوند؛ مثلاً یک داده ممکن است توسط یک زمینه یا با زمینه‌های متعدد، حتی در درون یک مجموعه داده واحد نشان داده شوند؛

۳. در میان زمینه‌های شناخته‌شده در زمان طراحی، بسیاری از زمینه‌ها بدون ارزش می‌باشند.

روش‌های کیفیت داده در مقابل داده‌های ساختار نیافته همواره پیچیده‌تر هستند. مثلاً فرض کنید یک سیستم ثبت‌کننده اطلاعات شخصی مانند نام، نام خانوادگی، منطقه، و مکان تولید را در نظر بگیریم. شکل ۲-۳ مثال آقای پاتریک متسیزی متولد منطقه ماسایی مارا در کنیا را با استفاده از انواع حالات ساختاریافته (شکل ۲-۳ (۱))، ساختار نیافته (شکل ۲-۳ (۲)) و نیمه ساختاریافته (شکل ۲-۳ (۳)) نشان می‌دهد. دقت نحوی آن‌گونه که در بخش ۲-۲-۳ در مورد داده‌های ساختاریافته توضیح داده شد، اندازه‌گیری می‌شود. برای محاسبه تابع فاصله داده‌های نیمه ساختاریافته، باید فاصله کلی مربوط به شکل درخت XML را علاوه بر فاصله محلی زمینه‌ها در نظر گرفت. اکثر کارهای پژوهشی در متون مربوط به کیفیت داده‌ها بر روی داده‌های ساختاریافته و نیمه ساختاریافته متمرکز شده‌اند. به همین دلیل، اگرچه ما به ارتباط داده‌های ساختار نیافته اشاراتی می‌کنیم اما این اثر بر داده‌های ساختاریافته و نیمه ساختاریافته متمرکز شده است.

طبقه‌بندی مبتنی بر مشاهده داده‌ها در متون کیفیت داده‌ها، همچون تولید محصولات می‌باشد (شانکارانارایان^۱ ۲۰۰۰). از این منظر، سه نوع داده مشخص شده‌اند:

- داده‌های خام: این داده‌ها به‌عنوان داده‌هایی تعریف می‌شوند که دست‌خوش هیچ پردازشی نشده‌اند و با ایجاد و ذخیره‌سازی اولیه می‌توان آن‌ها را برای مدت طولانی نگهداری کرد؛
- محصولات داده‌ای: که نتیجه یک فعالیت تولیدی انجام‌شده بر روی داده‌ها می‌باشند؛
- اقلام داده‌ای تکمیل‌کننده: که در زمان ایجاد اطلاعات محصول مربوطه موردنیاز می‌باشند و موقتاً ذخیره می‌شوند تا محصول نهایی را بسازند.

1. Shankaranarayan

همان گونه که در ادامه گفته خواهد شد، این نوع طبقه‌بندی موجب می‌شود تا امکان کاربرد شیوه‌های سنتی مدیریت کیفیت که سال‌ها در تولید محصولات استفاده شده‌اند در حوزه کیفیت داده نیز ممکن و میسر شود.

۲-۳-۶. ارتباط کیفیت داده با سیستم‌های اطلاعاتی

روش‌های کیفیت داده توسط نوع سیستم‌های اطلاعاتی مورد استفاده برای ارزیابی و بهینه‌سازی نیز تحت تأثیر قرار می‌گیرند. پژوهش‌های گذشته معماری سیستم‌های اطلاعاتی^۱ را به عنوان یک عامل هماهنگ‌کننده کل سیستم اطلاعاتی شرکت پشتیبانی معرفی می‌کند (زاگمن^۲ ۲۰۰۶). معماری سیستم‌های اطلاعاتی براساس نوع داده، سطح پردازش و چگونگی پشتیبانی توسط یک سیستم فنی متمایز می‌گردند. هرچه سطح، پردازش و مدیریت داده کاهش می‌یابد، ارزیابی کیفیت داده و شیوه‌های بهینه‌سازی پیچیده‌تر شده و ارزیابی و بهبود کیفیت داده‌ها چالش برانگیزتر می‌شود. انواع سیستم‌های اطلاعاتی زیر را می‌توان براساس سطح و نحوه پردازش آن‌ها شناسایی کرد:

- در یک سیستم اطلاعات یکپارچه نرم‌افزارها تک لایه‌ای بوده و دسترسی به داده‌ها را ارائه نمی‌دهند. اگرچه معمولاً داده‌ها در پایگاه داده‌ای قابل جستجو مرتب می‌شوند، اما نرم‌افزارهای جداگانه‌ای را به اشتراک نمی‌گذارند. این امر می‌تواند موجب تکرار داده‌ها شود، که احتمالاً بر تمام ابعاد کیفیت نیز مؤثر می‌باشد؛
- یک انبار داده، مجموعه‌ای متمرکز از اطلاعات بازیابی شده از پایگاه داده‌های متعدد است. انبار داده به صورت دوره‌ای با داده‌های به روز از پایگاه داده اصلی به‌هنگام سازی می‌شود. داده‌ها در تطابق با وضع فیزیکی موجود نیز به‌روزرسانی می‌شوند؛
- سیستم اطلاعات توزیع شده^۳، مجموعه‌ای از ماژول‌های نرم‌افزاری است که

1. Information System, IS

2. Zachman

3. Distributed IS

توسط یک گردش کار با یکدیگر هماهنگ شده‌اند. جنبه‌های کاربردی نرم‌افزارها در سطوحی مانند نمایش، منطق کاربرد، مدیریت داده‌ها و نوع دسترسی به داده‌های خروجی تقسیم می‌شوند. در این نوع از سیستم اطلاعات می‌توان پایگاه داده‌های مختلفی ذخیره کرد، اما امکان یکپارچه کردن آن‌ها تضمین می‌گردد.

- سیستم اطلاعات تعاونی^۱ را می‌توان به‌عنوان یک سیستم اطلاعاتی در مقیاس بزرگ نیز تعریف کرد که سیستم‌های مختلف در سازمان‌های مختلف و مستقل را برای به اشتراک‌گذاری اهداف مشترک به‌صورت داخلی به هم وصل می‌کند (دی میشلیس^۲ و دیگران ۱۹۹۷). هماهنگی با دیگر سیستم‌های اطلاعاتی نیاز به توانایی تبادل اطلاعات دارد. در سیستم‌های اطلاعات تعاونی، داده‌های منطقی یکپارچه نیستند زیرا در پایگاه داده‌های جداگانه‌ای با توجه به طرح‌های مختلف، ذخیره می‌شوند. با این وجود، نرم‌افزارها روند انتقال و تبادل داده‌ها را ترکیب می‌کنند، تا بتوانند موجب همکاری میان فرایندهای موجود شوند. به عبارت دیگر، ادغام در سطح فرایند تحقق می‌یابد.

- در متون علمی برای نشان دادن هر نوع اطلاعات اتخاذ شده از فناوری‌های وب، از اصطلاح سیستم اطلاعاتی وب^۳ (ایساکوویتز^۴ و دیگران ۱۹۹۸) استفاده می‌شود. از دیدگاه فنی سیستم اطلاعاتی وب یک نرم‌افزار سرویس گیرنده/سرویس دهنده است. چنین سیستمی معمولاً از داده‌های ساختاریافته، نیمه ساختاریافته و ساختار نیافته استفاده کرده و توسط ابزارهای توسعه و مدیریت براساس شیوه‌های خاص برای هر نوع داده پشتیبانی می‌شود.

1. Cooperation Information System, CIS
3. Web Information System, WIS

2. De Michelis
4. Isakowitz
5. Peer To Peer, P2P

- در یک سیستم اطلاعاتی دوطرفه^۵، تمایزی بین کاربر و سرویس‌دهنده وجود ندارد. این سیستم توسط مجموعه‌ای از گره‌های یکسان تشکیل شده است که داده‌ها و خدمات نرم‌افزاری را به‌منظور برآوردن نیازهای کاربران، به اشتراک می‌گذارد. سیستم‌های دوطرفه توسط چند خاصیت مشخص می‌شود:

- نداشتن هماهنگی مرکزی؛
- نداشتن پایگاه داده مرکزی؛
- همتا نداشتن در دیدگاه جامع سیستم.

همتایان خودمختارند و می‌توانند به‌صورت پویا به سیستم وصل شده یا از سیستم قطع شوند. با اینهمه، همتایان معمولاً روش‌های اجرایی و گردش کارها را به اشتراک می‌گذارند.

۴-۲. مقایسه روش‌های مختلف کیفیت داده‌ها

در این بخش، روش‌ها براساس معیارهای پیشگفته مقایسه می‌شوند. جدول ۲-۳ فهرست روش‌های مدنظر توسط کلمات اختصاری همراه با نام توسعه‌یافته روش و مرجع اصلی را نشان می‌دهد و جدول ۴-۲ به مقایسه آن‌ها می‌پردازد. همچنین از مخفف‌هایی برای اشاره به هر روش در ادامه این پژوهش استفاده خواهد شد.

جدول ۲-۳ مدل‌های کیفیت داده‌ها

مخفف روش	نام اصلی	مراجع
TDQM	مدیریت کیفیت جامع داده ^۱	ونگ ۱۹۹۸
DWQ	روش‌شناسی کیفیت انبار داده ^۲	جنوسفلد ^۳ و دیگران ۱۹۹۸
TIQM	مدیریت کیفیت جامع اطلاعات ^۴	انگلیش ۱۹۹۹
AIMQ	روشی برای ارزیابی کیفیت اطلاعات ^۵	لیو دیگران ۲۰۰۲
CIHI	روش موسسه کانادایی برای اطلاعات سلامت ^۶	لانگ و سکو ^۷ ۲۰۰۵
DQA	ارزیابی کیفیت داده ^۸	پیپینو و دیگران ۲۰۰۲
IQM	اندازه‌گیری کیفیت اطلاعات ^۹	اپلر و مونزنمایر ^{۱۰} ۲۰۰۲
ISTAT	روش ^{۱۱} ISTAT	فالرسی ^{۱۲} و دیگران ۲۰۰۳
AMEQ	روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول ^{۱۳}	سو و جین ^{۱۴} ۲۰۰۴
COLDQ	تأثیر کیفیت پائین داده بر هزینه‌ها ^{۱۵}	لوشین ^{۱۶} ۲۰۰۴
DaQuinCIS Data	کیفیت در سیستم‌های اطلاعات تعاونی ^{۱۷}	اسکنپیکو ^{۱۸} ۲۰۰۴
QAFD	روشی برای ارزیابی کیفیت داده‌های مالی ^{۱۹}	دامیک و بتینی ^{۲۰} ۲۰۰۴
CDQ	روشی جامع برای مدیریت کیفیت داده ^{۲۱}	بتینی و اسکنپیکو ۲۰۰۴

- 1 Total Data Quality Management
- 2 The Data warehouse Quality Methodology
- 3 Jeusfeld
- 4 Total Information Quality Management
- 5 A Methodology For Information Quality Assessment
- 6 Canadian Institute For Health Information Methodology
- 7 Long And Seko
- 8 Data Quality Assessment
- 9 Information Quality Measurement
- 10 Eppler And Munzenmaier
- 11 Istat Methodology (Italian National Bureau Of Census)
- 12 Falorsi
- 13 Activity-Based Measuring And Evaluating For Product Information Quality Methodology
- 14 Su And Jin
- 15 Loshin Methodology Or Cost-Effect Of Low Data Quality
- 16 Loshin
- 17 Data Quality In Cooperative Information Systems
- 18 Scannapieco
- 19 Methodology For The Quality Assessment Of Financial Data
- 20 De Amicis And Batini
- 21 Comprehensive Methodology For Data Quality Management

جدول ۲-۴ مقایسه مدل‌های کیفیت داده‌ها مبتنی بر روش‌های ارزیابی داده

قابلیت توسعه به دیگر ابعاد و معیارها	اندازه‌گیری کیفیت	مدل‌سازی فرایند	شناسایی حوزه‌های کلیدی	تحلیل الزامات کیفیت داده	تحلیل داده‌ها	مخفف روش / مرحله
ثابت	+	+	+		+	TDQM
باز	+		+	+	+	DWQ
ثابت	+	+	+	+	+	TIQM
ثابت	+		+		+	AIMQ
ثابت			+		+	CIHI
باز	+		+		+	DQA
باز	+				+	IQM
ثابت	+				+	ISTAT
باز	+	+	+		+	AMEQ
ثابت	+	+	+	+	+	COLDQ
باز	+	+	+		+	DaQuinCIS Data
ثابت	+		+	+	+	QAFD
باز	+	+	+	+	+	CDQ

هزینه‌ها، ابعاد و کیفیت گام‌های اجرا به‌عنوان متمایزترین معیار به‌منظور ایجاد تمایز در روش‌شناسی‌های مختلف مطرح بوده و در ادامه مورد بحث و مقایسه قرار خواهند گرفت.

۲-۴-۱. روش‌ها، گام‌ها و مراحل

جدول‌های ۲-۵ و ۲-۶ مراحل و گام‌های هر یک از روش‌ها را به‌منظور ایجاد بهبود در کیفیت نشان می‌دهد. اگر در مدل مربوطه به گام یا روش ایجاد بهبود اشاره شده باشد از علامت «+» و در غیر این صورت آن سلول جدول خالی خواهد بود. برای نمونه، روش‌شناسی کیفیت انبار داده عموماً به مرحله مدل‌سازی فرایند اشاره دارد اما جزئیات اجرا را ارائه نمی‌دهد، لذا در جدول ۲-۵ این گام برای مدل روش‌شناسی کیفیت انبار داده در نظر گرفته نشده است. به همین ترتیب، مرحله

اندازه‌گیری کیفیت موسسه کانادایی برای روش اطلاعات سلامت در جدول ۵-۲ در نظر گرفته نشده است. ارزیابی و بهبود به‌طور جداگانه در دو بخش بعدی بحث شده‌اند.

جدول ۵-۲ مقایسه مدل‌های کیفیت داده‌ها مبتنی بر گام‌های بهبود

شناسایی علت ریشه‌ای بروز خطاها	انتخاب استراتژی	تخصیص مالک به داده‌ها	تخصیص مالک به معرفی فرایند	ارزیابی هزینه‌ها	مخفف روش / مرحله
+	+	+	+	+	TDQM
+	+	+		+	DWQ
+	+	+	+	+	TIQM
+					DQA
+	+				ISTAT
+					AMEQ
+	+			+	COLDQ
+	+				DaQuinCIS Data
+	+	+	+	+	CDQ

جدول ۶-۲ مقایسه مدل‌های کیفیت داده‌ها مبتنی بر گام‌های بهبود

پایش مستمر	مدیریت بهبود	طراحی مجدد فرایند	طراحی راه‌کارهای بهبود داده	کنترل فرایند	نام روش / مرحله
+	+	+			TDQM
	+		+		DWQ
+		+	+		TIQM
					DQA
		+	+		ISTAT
+					AMEQ
+		+	+	+	COLDQ
					DaQuinCIS Data
		+	+	+	CDQ

۲-۴-۲. چگونگی ارزیابی در هر مدل

در جدول ۷-۲ چگونگی ارزیابی داده‌ها در هر مدل ارائه و مورد مقایسه قرار گرفته است. به‌طور کلی، روش‌های ارائه‌شده در این بخش مشابه روش‌های معرفی‌شده در بخش ۲-۲ است. ممکن است تجزیه و تحلیل داده‌ها و اندازه‌گیری کیفیت از مهم‌ترین اقدامات ارزیابی داده باشند که در اکثر مدل‌ها مشترک هستند. با اینهمه، شاید به روش‌های مختلف انجام شوند. مثلاً مرحله اندازه‌گیری کیفیت، به‌صورت پرسشنامه در یک روش ارزیابی کیفیت اطلاعات با ترکیبی از معیارهای ذهنی و عینی و یا با تجزیه و تحلیل در یک روش برای ارزیابی کیفیت داده‌های مالی انجام می‌شود. روش‌های اندازه‌گیری، باید متناسب با کسب و کار سازمانی، فرایندها، کاربران و یا خدمات مختلف بومی‌سازی شوند.

جدول ۷-۲ روش‌ها و انواع استراتژی‌ها

نام روش/استراتژی	داده محور	فرایند محور
TDQM		طراحی مجدد فرایند
DWQ		
TIQM	پاک کردن داده نرمالیزه کردن	طراحی مجدد فرایند
ISTAT	نرمالیزه کردن پیوند میان رکوردها	
COLDQ	بهینه‌سازی هزینه	کنترل فرایند پیوند میان رکوردها
DaQuinCIS	اعتماد به منبع پیوند میان رکوردها	
CDQ	نرمالیزه کردن پیوند میان رکوردها ادغام داده‌ها و طرح محلی سازی و تصحیح خطا	کنترل فرایند طراحی مجدد فرایند

تنها چند روش است که مرحله تجزیه و تحلیل الزامات کیفیت داده، شناسایی مسائل کیفیت داده و جمع‌آوری سطح کیفیت هدف جدید از کاربران را انجام می‌دهند. این

مرحله به‌طور ویژه به ارزیابی و حل اختلافات در سطوح هدف کیفیت داده از دیدگاه ذینفعان مختلف مربوط می‌باشد. مثلاً یک روش ارزیابی کیفیت داده‌های مالی مجموعه‌ای از سطوح کیفیت هدف را با انواع مختلف برای کارشناسان مانند کارشناسان کسب‌وکار و کارکنان واحد مالی را پیشنهاد می‌کند اما به اصلاح ناسازگاری داده کمکی نمی‌کند. چند روش از مدل‌سازی فرایند پشتیبانی می‌کنند. توجه داشته باشید که به‌غیر از روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول، روش‌ها از مدل‌سازی فرایند پشتیبانی کرده و همچنین یک استراتژی فرایند محور را برای مرحله بهبود اتخاذ می‌کنند.

۲-۴-۳. گام بهبود

جدول‌های ۲-۶ و ۲-۷ مراحل بهبود روش‌های مختلف را مقایسه می‌کند. معمولاً در گام بهبود به شناسایی علل ریشه‌ای خطاها پرداخته می‌شود. روش ارزیابی کیفیت داده بر اهمیت گام شناسایی علل خطاها تأکید می‌کند، اما در مورد اجرای آن‌ها بحث نمی‌کند. همچنین روش‌شناسی کیفیت انبار داده به یک مدل ریاضی براساس مفهوم وابستگی علت و معلول در گام شناسایی علل خطاها اشاره می‌کند، اما مدل ریاضی ارائه‌شده در این روش به‌طور کامل و همراه با اجزاء آن ارائه نشده است. فعالیت‌های بهبود در روش‌های مختلف به‌غیر از روش کیفیت انبار داده، عمدتاً براساس طراحی مجدد فرایند هستند که آن معیار از قبل تعیین‌شده، سؤال متریک هدف^۱ را فراهم می‌کنند (باسیلی^۲ و دیگران ۱۹۹۴). رویکرد مذکور ابتدا در زمینه مهندسی نرم‌افزار ارائه‌شده بود. مرحله ارزیابی هزینه، معمولاً در همه روش‌های مدیریت کیفیت داده الزامی است. این مرحله برای اندازه‌گیری مزیت اقتصادی راه‌حل‌های بهبود و انتخاب کارآمدترین شیوه‌های بهبود در نظر گرفته می‌شود. از سوی دیگر، در مدل مدیریت کیفیت جامع داده مرحله‌ای با عنوان مدیریت راهکارهای بهبود دیده‌شده تا فرایند بهبود مستمر کیفیت داده بهتر صورت گیرد.

1. Goal Question Metric

2. Basili

روش‌های دیگر به طیف گسترده‌ای از شیوه‌ها و بهترین شیوه‌های مدیریت تغییر اشاره دارند (کیتینگر و گراور^۱ ۱۹۹۵)). علاوه بر این، ممکن است مرحله ارزیابی روش به‌منظور ارزیابی نتایج حاصل از گام بهبود تکرار شود. مثلاً روش ارزیابی کیفیت داده گام‌های روش قبلی را به‌منظور بررسی اثربخشی بهبود توصیه می‌کند.

درنهایت، رابطه بین کیفیت داده، فرایند و سازمان توسط روش‌های مدیریت کیفیت جامع اطلاعات، مدیریت کیفیت داده و روشی جامع برای مدیریت کیفیت داده در نظر گرفته می‌شود. این روش‌ها به موضوع تخصیص مسئولیت در فرایندها و داده‌ها می‌پردازند. در گام تخصیص مسئولیت، اطلاعات شناسایی وضع موجود به‌عنوان ورودی تکمیلی بایستی در نظر گرفته شود. این روش برای مدیریت کیفیت جامع داده، مجموعه‌ای از ماتریس‌ها را به‌منظور نشان دادن رابطه بین فرایندها، واحدهای سازمانی و پایگاه داده معرفی می‌کند، که در طول مرحله بهبود و پس‌از آن در مرحله تعیین مسئولیت استفاده می‌شود.

۲-۵. روش‌ها، استراتژی‌ها و شیوه‌ها

جدول ۲-۸، استراتژی‌ها و شیوه‌های به خدمت گرفته‌شده توسط روش‌های متفاوت را نشان می‌دهد. یک روش با یک استراتژی در ارتباط بوده و رهنمودهایی برای انتخاب و طراحی شیوه‌های مربوط به آن استراتژی، فراهم کند. باید توجه داشت که ستون‌های جدول ۲-۸ با برچسب فرایند محور همان اطلاعات ستون‌های کنترل فرایند و طراحی مجدد فرایند در جدول ۲-۷ را نشان می‌دهند و ستون‌هایی با برچسب داده محور صراحتاً به شیوه‌های داده محور ضمنی در جدول‌های ۲-۶ و ۲-۷ اشاره دارند.

جدول ۲-۸، روش‌های مختلف کیفیت داده برای استراتژی‌های ترکیبی را به نمایش می‌دهد، و ترکیب‌های متنوعی از روش‌های داده محور و شیوه‌های فرایند محور را ارائه می‌دهد. روش مدیریت کیفیت جامع اطلاعات برای طیف وسیع‌تری از

شیوه‌های داده محور و فرایند محور استفاده می‌شود. در مقابل آن، روش مدیریت کیفیت داده دستورالعمل‌هایی برای اعمال استراتژی‌های فرایند محور با استفاده از اطلاعات ماتریس تحلیل تولید^۱ (بالو و دیگران ۱۹۹۸) فراهم می‌کند که زمان و چگونگی بهبود داده‌ها را نشان می‌دهد.

شایان ذکر است روشی که به‌طور انحصاری استراتژی داده محور (مثلاً روش‌شناسی کیفیت انبار داده برای ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی) و یا استراتژی فرایند محور را اتخاذ کند، شاید برای سازمان‌هایی با تجربه‌های قبلی در پروژه‌های کیفیت داده انعطاف‌پذیر نباشد. تنها روشی که صراحتاً به این مسئله می‌پردازد روشی جامع برای مدیریت کیفیت داده است که به‌طور مشترک شیوه داده محور و فرایند محور را انتخاب می‌کند. انتخاب مناسب‌ترین استراتژی و شیوه کاملاً وابسته به حوزه کاربرد داده‌ها می‌باشد (باتینی^۲ و دیگران ۲۰۰۸). نرمالیزه کردن، پیوند میان رکوردها و یکپارچه‌سازی داده و طرح نشان‌دهنده شیوه‌های داده محوری است که به‌طور گسترده در روش کیفیت داده اتخاذ شده‌اند. طراحی مجدد فرایند، به روش فرایند محور مربوط می‌شود. اکنون به نوآوری‌های خاص مربوط به شیوه‌های داده محور و فرایند محور می‌پردازیم.

۲-۵-۱. شیوه‌های داده محور

شیوه‌های نرمالیزه کردن در حوزه‌های مختلفی مانند سرشماری و داده‌های منطقه‌ای مطرح شده است. روش ISTAT و روش جامع برای مدیریت کیفیت داده، شیوه‌های نرمالیزه سازی را برای بهبود کیفیت داده از طریق مقایسه داده‌ها با جستجو کردن^۳ در جدول‌ها و تعریف یک طرح جامع ارائه می‌دهند. مثلاً روش ISTAT از جدول‌های موجود در سیستم ثبت ملی خیابان به‌عنوان یک جدول پایه استفاده می‌کند و سپس اطلاعات و داده‌های استانی و منطقه‌ای را به کمک روش جستجو در جدول‌های سیستم ثبت ملی خیابان‌ها بهبود می‌بخشد.

1. Manufacturing Analysis Matrix

2. Batini

3. Look-Up

پیوند رکوردها به عنوان یک روش بهبود کیفیت داده در پژوهش‌های صورت گرفته در حوزه پایگاه داده از دهه پنجاه تاکنون مورد استفاده قرار گرفته است و در بسیاری از زمینه‌ها مانند برنامه‌های مراقبت بهداشتی، اداری و سرشماری استفاده شده است. در چنین شرایطی، ایجاد روش‌های کارآمد منطبق با رایانه می‌تواند استفاده از منابع دفتری را کاهش دهد و خطاها را به حداقل برساند. روش جامع برای مدیریت کیفیت داده در مورد سه نوع از شیوه‌های ثبت ارتباطات بحث می‌کند:

۱. شیوه‌های احتمالی: براساس مجموعه گسترده‌ای از روش‌ها که در طول دو قرن گذشته، در آمار و نظریه احتمالات توسعه یافته‌اند، مثل شبکه‌های بیزین در داده‌کاوی؛

۲. شیوه‌های تجربی: شیوه‌های الگوریتمی مانند طبقه‌بندی، تجزیه و تحلیل درخت، مقایسه همسایه و هرس؛

۳. روش‌های مبتنی بر دانش: براساس استخراج دانش از زمینه‌ها و استفاده از استراتژی‌های استدلال بر مبنای آن‌ها؛

معیارهای انتخاب بین این سه نوع شیوه در روش جامع برای مدیریت کیفیت داده مورد بحث قرار گرفته است. پروژه ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی، شیوه ثبت ارتباطات خاص را توسعه داده است (برتولازی و همکاران^۱ ۲۰۰۳). در پروژه ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی پیوند رکوردها در دو گام انجام شد:

۱. نسخه‌های مختلف از نهادهای مشابه در منابع داده مختلف همسان‌سازی می‌شود؛

۲. با گزارش‌گیری‌های مشابه از منابع مختلف داده‌ای و شناسایی اختلاف میان گزارش‌ها، مشکلات واحد شناسایی و رفع می‌شوند.

روش پیوند رکوردها مبتنی بر روش همسایگی طبقه‌بندی شده^۲، توسعه یافته است (هرناندز و استولفو^۳ ۱۹۹۸). اما برخی از ویژگی‌های جدید زیر به آن اضافه شده است.

1. Bertolazzi

2. Sorted Neighborhood Method

3. Hernandez And Stolfo

- عناصر و ستون‌های کلیدی که به‌منظور بررسی انطباق میان داده‌ها از منابع مختلف استفاده می‌شوند به‌صورت خودکار و به کمک الگوریتم‌های هوشمند انتخاب می‌شوند؛
 - الگوریتم‌های تطبیق مبتنی بر تابعی عمل می‌کنند که فاصله کلاسیک را برحسب طول رشته‌ها، محاسبه و نرمالیزه می‌کند.
- یکپارچه‌سازی داده‌ها و طرح‌ها^۱ (لنزرینی^۲ ۲۰۰۲) حوزه گسترده‌ای از پژوهش‌ها می‌باشد که تا حدودی باکیفیت داده‌ها دارای همپوشانی هستند. اغلب شیوه‌های بهبود داده محور مورد استفاده در روش‌ها، مبتنی بر استفاده از داده‌های جدید هستند، که به‌منظور بهبود کیفیت جمع‌آوری داده‌ها ارائه شده‌اند. در نتیجه، شیوه‌های بهبود کیفیت داده در درجه اول بر روی ناهمگنی نمونه‌های موجود به‌منظور شناسایی موارد مشابه، رفع مغایرت‌ها و در نهایت انتخاب یک نمونه نهایی منحصربه‌فرد صحیح متمرکز هستند.
- روش‌های روش‌شناسی کیفیت انبار داده، ISTAT، ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی و روش جامع برای مدیریت کیفیت داده به ناهمگنی داده‌ها می‌پردازند. در روش کیفیت انبار داده، منابع اطلاعاتی ناهمگن در دسترس، ابتدا به شیوه‌ای یکنواخت از طریق مکانیسم‌های استخراج به نام الگوریتم‌های پوششی شناسایی می‌شوند. سپس به عملیات یکپارچه‌سازی اطلاعات و حل تناقضات می‌پردازند. در نتیجه داده‌های استاندارد و یکپارچه‌شده به‌صورت موردی در انبار داده‌ها ذخیره می‌شوند.
- روش ISTAT نشان می‌دهد که چگونه ناهمگنی در میان داده‌های مدیریت‌شده با مدل مشترک برای نشان دادن فرمت داده‌های تبدیل‌شده براساس زبان نشانه‌گذاری XML، توسط سازمان‌های دولتی مختلف حل می‌شود. در این روش درک ناهمگنی میان سازمان‌ها آسان‌تر شده است، درحالی‌که برای حل این ناهمگنی به توافق دو یا

چندجانبه میان سازمان‌ها نیاز است. در روش ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی به بررسی ناهمگنی سطح نمونه در میان منابع داده مختلف توسط کارگزار کیفیت داده پرداخته می‌شود. نسخه‌های مختلفی از داده‌های مشابه دریافت شده به‌عنوان پاسخ به درخواست، توسط کارگزار کیفیت داده تنظیم می‌شوند و بهترین مقدار کیفیتی انتخاب می‌شود.

روش جامع برای مدیریت کیفیت داده مشابه روش ISTAT می‌باشد، چراکه تأکید بیشتری بر استقلال سازمان در یک سیستم همکاری دارد. در این روش حل ناهمگنی مطالعات موردی که به‌عنوان بهترین شیوه پیشنهاد شده است. از طریق پیوند میان رکوردها در یک‌لایه بسیار نازک از داده‌ها، یعنی شناسه انجام شده است. سایر داده‌ها تنها در صورت تصمیم‌گیری مستقل سازمان‌های درگیر استفاده می‌شوند.

۲-۵-۲. شیوه‌های فرایند محور

روش‌هایی که به گام طراحی مجدد فرایند می‌پردازند، تمایل به استفاده از شیوه‌های مربوط به فرایند مهندسی مجدد کسب‌وکار^۱ دارند (موتو^۲ و دیگران ۱۹۹۹؛ هامر^۳ ۱۹۹۰). روش مدیریت کیفیت داده در این رابطه استثناست به‌طوری‌که رویکردی برای کنترل فرایند طراحی مجدد را پیشنهاد می‌دهد و به یک «سیستم تولید محصول اطلاعاتی»^۴ اشاره می‌کند (بالو و دیگران ۱۹۹۸). این روش مدل نقشه تولید اطلاعات^۵ را پیشنهاد می‌دهد (شانکاراناریان و دیگران ۲۰۰۰) که در مدل‌های محصولات اطلاعاتی مدیریت شده توسط فرایندهای تولید استفاده می‌شود. نقشه تولید اطلاعات یک مدل گرافیکی طراحی شده برای کمک به تجسم تحلیل‌گران فرایند تولید اطلاعات، شناسایی مالکیت مراحل فرایند، درک اطلاعات و مرزهای سازمانی، برآورد زمان و کیفیت معیارهای مرتبط با فرایند تولید فعلی است. توصیف فرایندها، یک فعالیت اجباری و هم‌راستا با جهت‌گیری کلی استراتژی‌های فرایند محور است. پس از

1. Business Process Reengineering, BPR
3. Hammer
5. Information Production Map, IP-Map

2. Muthu
4. Information Production System

مدل‌سازی و ارزیابی فرایند تولید اطلاعات، فعالیت‌های کنترل فرایند جدید مشخص می‌شوند و یا در مورد طراحی مجدد فرایند تصمیم گرفته می‌شود.

راه‌حل‌های پیچیده مانند نقشه تولید اطلاعات به دلیل هزینه‌های بالای پیاده‌سازی همیشه قابل استفاده نبوده و در برخی موارد، کاربرد عملی از طریق اجرای تنهای مدل‌سازی فرایند امکان‌پذیر نیست. به همین دلیل روش‌های دیگر و راه‌حل‌های کمتر رسمی ولی امکان‌پذیر انتخاب می‌شوند. مثلاً روش جامع برای مدیریت کیفیت داده مبتنی بر مجموعه‌ای از ماتریس‌هاست که روابط اصلی را در میان داده‌ها، جریان اطلاعات، فرایندها و واحدهای سازمانی توصیف می‌کنند. رابطه بین واحدهای سازمانی و فرایندها نیز در مدل‌های توسعه‌یافته نقشه تولید اطلاعات گنجانده شده‌اند (اسکاناپیکو و دیگران ۲۰۰۲).

۲-۶. ارتباط مدل‌های مدیریت کیفیت با ابعاد کیفیت داده

ابعاد کیفی در نظر گرفته شده در هر مدل در جدول ۲-۸ نشان داده شده‌اند. در این جدول هر بعد در هر سطر به این شرط آورده شده است که آن روش تعریف متناظری در مورد آن بعد ارائه داده باشد. در جدول ۲-۴ نیز متناظر با هر بعد و روش توضیحات کافی داده شده است (جدول ۲-۴). توجه شود که ابعاد گوناگونی در هر مدل تعریف شده که نشان‌دهنده پیچیدگی مفهوم کیفیت داده است. این امر غافلگیرکننده نیست چراکه امروزه تعداد زیادی از پدیده‌ها می‌توانند برحسب داده‌ها تعریف شوند. روش مدیریت کیفیت جامع اطلاعات، ابعاد را از دو منظر ذاتی^۱ و کاربردی^۲ دسته‌بندی می‌کند. روش ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها (COLDDQ) میان طرح، داده و سیاست‌گذاری تمایز ایجاد می‌کند و هریک را در یک بعد دسته‌بندی می‌کند. روش موسسه کانادایی برای روش اطلاعات سلامت (CIHI)، یک دسته‌بندی دو مرحله‌ای برحسب ابعاد و خصوصیات مرتبط ارائه می‌کند. روش جامع برای مدیریت کیفیت داده ابعاد کیفیت را در دو سطح دسته‌بندی می‌کند. سطح

1. Inherent

2. Pragmatic

اول مانند بقیه مدل‌ها ابعاد کیفیت داده را تعیین می‌کند و در سطح دوم مشخصه‌های^۱ دو بعد معرفی می‌شوند.

در جدول ۲-۱۰، معیارهای فراهم‌شده برای اندازه‌گیری ابعاد کیفی در مدل‌های گوناگون آورده شده است. در این جدول برای بعد صحت معنایی، تعریف و معیاری ارائه نشده است زیرا دو روش ارزیابی کیفیت داده‌های مالی و روش جامع برای مدیریت کیفیت داده روش‌های اندازه‌گیری مشخصی را ارائه نکرده‌اند. با توجه به این ویژگی‌ها، چند معیار یا شاخص اندازه‌گیری برای هر بعد تعریف شده و هر بعد دارای چند ورودی^۲ در جدول است. توجه کنید معیارهای ذهنی نظیر پیمایش‌های^۳ کاربر (از طریق پرسشنامه و مانند آن) تقریباً برای تمام ابعاد کیفی قابل تعریف است. معیارهای متفاوت با علائم اختصاری شناسایی می‌شوند و در جدول ۲-۱۱ آمده‌اند تا بتوان آن‌ها را با مدل‌هایی که مورداستفاده قرار می‌دهند در قالب یک ماتریس مرتبط نمود. آخرین ستون جدول ۲-۱۰ برای هر بعد و هر سنجه مرتبط با آن، دو مورد را مشخص می‌کند:

- تعداد روش‌هایی که از آن سنجه استفاده می‌کنند و تعریف دقیقی ارائه

می‌کنند؛

- تعداد کل روش‌هایی که به بعد متناظر با آن سنجه اشاره می‌کنند.

نسبت میان این دو عدد (حاصل تقسیم این دو عدد) در سنجه‌های مختلف به‌عنوان درجه عمومیت یافتن^۴ آن سنجه در نظر گرفته می‌شود. چنین معیاری برای ابعادی مثل دقت، جامعیت^۵ و ثبات^۶ معمولاً بالا است، درحالی‌که این مقدار به شکل چشمگیری برای دو بعد مرتبط به زمان یعنی به‌موقع بودن و رواج پایین است. در روش اندازه‌گیری کیفیت اطلاعات که به کیفیت اطلاعات وب می‌پردازد، تقریباً برای هر بعد یک سنجه وجود دارد. چنین معیارهایی با در نظر گرفتن ابزارهای اندازه‌گیری در زمینه وب تعریف می‌شود. مثلاً امکان دارد با استفاده از یک تحلیلگر سایت، ابعادی نظیر قابلیت دسترسی، پایداری، بجا بودن، اختصار^۷ و قابلیت نگهداری^۸

1. Characteristics
5. Completeness

2. Entry
6. Consistency

3. Survey
7. Conciseness

4. Consensus
8. Maintainability

را ارزیابی کرد. تحلیلگران می‌توانند برای ارزیابی عملی بودن و سهولت اجرا از روش ارزیابی حجم محاسبات ترافیکی استفاده کنند همچنین ابزار اسکنرهای درگاه^۱ برای ارزیابی امنیت مفید هستند. تعداد بالای ابزارهای اندازه‌گیری در حوزه وب، منجر به وجود تعداد بالای معیارها در متدولوژی اندازه‌گیری کیفیت اطلاعات می‌شود. روش ارزیابی کیفیت اطلاعات دارای معیارها و ابعاد مشخص گوناگونی است. این امر به علت رویکرد بالا به پایین به خدمت گرفته‌شده در این روش برای تعریف ابعاد و معیارهاست که از دو دسته‌بندی متفاوت استفاده می‌کند:

۱. کیفیت محصول در مقابل کیفیت خدمات؛
۲. مطابقت با مشخصات در مقابل برآورده سازی نیازهای مشتری یا عدم تأمین آن‌ها.

جدول ۲-۸ روش‌ها و ابعاد کیفی

مخفف	ابعاد کیفی داده‌ها
TDQM	قابلیت دسترسی، مناسب بودن، باورپذیری، جامعیت، ارائه همسان، سهولت تغییر، ارزش‌افزوده داشتن، بدون خطا بودن، تفسیرپذیری، عینی بودن، مربوط بودن، شهرت، امنیت، به‌جا بودن، درک پذیری.
DWQ	صحت، تمامیت، به‌اندازه بودن، قابلیت ردگیری، قابلیت تفسیر، تکامل‌پذیری فراداده‌ها، قابلیت دسترسی (سیستم، تراکنش، امنیت)، مفید بودن (قابلیت تفسیر)، به‌هنگام بودن (رایج بودن، فراربت)، پاسخ‌دهی، جامعیت، اعتبار، صحت، همسانی، قابلیت تفسیر.
TIQM	ابعاد ذاتی: همسانی، جامعیت، تبعیت از قوانین تجاری، صحت انطباق با منابع داده، صحت انطباق با واقعیت، دقت، بدون تکرار، هم ارزی داده‌های زائد، هم‌زمانی داده‌های زائد. ابعاد عملی: قابلیت دسترسی، به‌هنگام بودن، وضوح متنی، یکپارچگی، قابلیت استفاده، جامعیت، هزینه.
AIMQ	قابلیت دسترسی، مناسب بودن، باورپذیری، جامعیت، ارائه موجز/همسان، سهولت اجرا، بدون خطا، قابلیت تفسیر، عینی بودن، مربوط بودن، شهرت، امنیت، به‌هنگام بودن، درک پذیری.
CIHI	ابعاد: صحت، قابلیت مقایسه، به‌هنگام بودن، قابلیت استفاده، مربوط بودن. مشخصه‌ها: سطح پوشش، قابلیت اطمینان، واحد/آیتم بدون پاسخ، ویرایش پذیری و سندیت داشتن، پردازش، تخمین، به‌هنگام بودن، جامعیت، یکپارچگی، استانداردسازی، هم ارزی، توانایی پیوند، قابلیت مقایسه محصول/تاریخی، قابلیت دسترسی، مستندسازی، قابلیت تفسیر، انطباق‌پذیری، ارزش.

مخفف	ابعاد کیفی داده‌ها
DQA	قابلیت دسترسی، حجم مناسبی از داده، باورپذیری، جامعیت، بدون خطا، همسانی، ارائه موجز/همسان، مربوط بودن، سهولت دست‌کاری، قابلیت تفسیر، عینی بودن، شهرت، امنیت، به‌هنگام بودن، درک پذیری، ارزش‌افزوده.
IQM	قابلیت دسترسی، همسانی، به‌هنگام بودن، موجز بودن، نگهداشت پذیری، رایج بودن، عملی بودن، سهولت، سرعت، جامعیت، وضوح، صحت، قابلیت ردگیری، امنیت، دقت
ISTAT	صحت، تمامیت، همسانی.
AMEQ	ارائه همسان، قابلیت تفسیر، درک پذیری، ارائه موجز، به‌هنگام بودن، جامعیت ارزش‌افزوده، مناسب بودن، معناداری، قابل‌خواندن، معقولیت، دقت، قابلیت اطمینان، کارایی، هزینه، عینی بودن، باورپذیری، شهرت، قابلیت دسترسی، صحت، عدم ابهام، همسانی، نقص طراحی و عملیات.
COLDQ	طرح: روشنی تعریف، جامعیت، انعطاف‌پذیری، نیرومندی، ضرورت، ویژگی دانه‌ای، دقت حوزه‌ها، همگنی، قابلیت شناسایی، قابلیت اکتساب، مربوط بودن، سادگی/پیچیدگی، همسانی معنایی، همسانی نحوی. داده: دقت، مقادیر تهی، تمامیت، همسانی، رایج بودن، بجا بودن، توافق مصرف، هماهنگی، پیوستگی. ارائه: مناسب بودن، صحیح قابلیت تفسیر، انعطاف‌پذیری، دقت قالب، قابلیت حمل، همسانی، استفاده از انبار. کاربرد اطلاعاتی: قابلیت دسترسی، فراداده، حریم خصوصی، امنیت، حسو، هزینه.
DaQuinCIS	صحت، تمامیت، همسانی، رایج بودن، قابلیت اعتماد.
QAFD	صحت معنایی/نحوی، همسانی داخلی/خارجی، تمامیت، رایج بودن، متحصربه‌فرد بودن
CDQ	طرح: درستی برحسب مدل، درستی برحسب الزامات، تمامیت، اقتضا، قابلیت خواندن، نرمال‌سازی، داده: صحت معنایی/نحوی، دقت معنایی، تمامیت، همسانی، رایج بودن، بجا بودن، فراریت، قابلیت تکمیل، شهرت، قابلیت دسترسی، هزینه.

توجه کنید که روش ارزیابی کیفیت اطلاعات فقط از معیارهای ذهنی برای ارزیابی ابعاد کیفی استفاده می‌کند. در این روش کیفیت داده‌ها عمدتاً برحسب پرسشنامه‌هایی با ۴ تا ۶ گویه مستقل برای هر بعد کیفی ارزیابی می‌شود. گویه‌ها دارای ساختار کلی به شرح زیر می‌باشند. «این اطلاعات (وصفی^۱ یا توضیحی^۲) هستند.» مثلاً تمامیت مرتبط با شش فیلد است که برخی از آن‌ها عبارت‌اند از:

۱. این اطلاعات تمام مقادیر ضروری را فراهم می‌کند؛
۲. این اطلاعات برای نیازهای ما کافی است؛
۳. این اطلاعات، نیازهای فعالیت‌های ما را تأمین می‌کند.

جدول ۲-۹ ابعاد و معیارها

ابعاد	نام	تعریف معیارها
صحت	Acc1	دقت نحوی: این به‌صورت فاصله میان مقدار ذخیره‌شده در پایگاه داده و درستی، اندازه‌گیری می‌شود
	Acc2	دقت معنایی = تعداد مقادیر صحیح/تعداد کل مقادیر
	Acc3	تعداد عناصر صحیح تحویل داده‌شده
جامعیت	Compl1	پیمایش کاربر - پرسشنامه
	Compl2	تمامیت = تعداد مقادیر غیر تهی/تعداد کل مقادیر
	Compl3	تمامیت = تعداد عناصر تحویل داده‌شده/تعداد مورد انتظار
	Compl4	تمامیت داده‌های وب = $(T_{\max} - T_{\min}) * (Completeness_{\max} - Completeness_{\min}) / 2$
سازگاری	Cons1	پیمایش کاربر به کمک پرسشنامه
	Cons2	همسانی = تعداد مقادیر همسان/تعداد کل مقادیر
	Cons3	تعداد عناصر نقض‌کننده محدودیت‌ها، تعداد تفاوت‌های برنامه‌نویسی
	Cons4	تعداد صفحات با انحراف از راهنمای سبک
به‌هنگام بودن	Time1	پیمایش کاربر - پرسشنامه
	Time2	به‌هنگام بودن = $(\text{فراریت/رواج} - ۰) * \text{Max}$
	Time3	درصد اجرای فرایندها قابل انجام در چارچوب زمانی موردنظر
رایج بودن	Curr1	پیمایش کاربر - پرسشنامه
	Curr2	رواج = زمانی که در آن داده‌ها در سیستم ذخیره می‌شود - زمانی که در آن داده‌ها در دنیای واقعی به‌روزرسانی می‌شود
	Curr3	زمان آخرین بروز رسانی
	Curr4	رواج = زمان درخواست - آخرین به‌روزرسانی
	Curr5	رواج = سن + (زمان درخواست - زمان ورود)
فراریت	Vol1	پیمایش کاربر - پرسشنامه
	Uni1	طول زمان برای آنکه داده‌ها معتبر باقی بمانند
منحصربه‌فردی		تعداد تکثیرها
حجم مناسب داده	Appr1	حجم مناسبی از داده‌ها = حداقل ((تعداد واحدهای داده‌ای فراهم‌شده/تعداد واحدهای داده‌ای موردنیاز؛ (تعداد واحدهای داده‌ای موردنیاز/تعداد واحدهای داده‌ای فراهم‌شده))
	Appr2	پیمایش کاربر - پرسشنامه

ابعاد	نام	تعریف معیارها
قابلیت دسترسی	Access1	قابلیت دسترسی = ماکزیمم (۰؛ ۱) - (زمان تحویل - زمان درخواست) / (ضرب الاجل - زمان درخواست)
	Access2	تعداد پیوندهای نامعتبر - تعداد مهارهای نامعتبر
	Access3	پیمایش کاربر - پرسشنامه
اعتبار	Cred1	تعداد عناصر با مقادیر پیش فرض
	Cred2	پیمایش کاربر - پرسشنامه
قابلیت تفسیر	Inter1	تعداد عناصر با داده‌های قابل تفسیر، مستندسازی مقادیر کلیدی
	Inter2	پیمایش کاربر - پرسشنامه
قابلیت استفاده	Usa1	پیمایش کاربر - پرسشنامه
یکپارچگی	Integr1	درصد محاسبات صحیح داده‌های مشتق شده بنا به تعریف محاسبه یا فرمول مشتق
موجز بودن	Conc1	تعداد صفحات عمیق (به شدت وابسته به سلسله مراتب)
	Conc2	پیمایش کاربر - پرسشنامه
قابلیت نگهداشت	Main1	تعداد صفحات با فرا اطلاعات از دست رفته
	App1	تعداد صفحات جدا افتاده
عملی بودن	App2	پیمایش کاربر - پرسشنامه
سهولت	Conv1	مسیرهای ناوبری دشوار: تعداد مسیرهای ناوبری گم شده/معیوب
سرعت	Speed1	زمان پاسخ سرور و شبکه
جامعیت	Comp1	پیمایش کاربر - پرسشنامه
شفافیت	Clar1	پیمایش کاربر - پرسشنامه
قابلیت ردگیری	Trac1	تعداد صفحات بدون مؤلف یا منبع
امنیت	Sec1	تعداد ورودهای ضعیف
	Sec2	پیمایش کاربر - پرسشنامه
دقت	Corr1	پیمایش کاربر - پرسشنامه
عینی بودن	Obj1	پیمایش کاربر - پرسشنامه
مربوط بودن	Rel1	پیمایش کاربر - پرسشنامه
شهرت	Rep1	پیمایش کاربر - پرسشنامه
سهولت اجرا	Ease1	پیمایش کاربر - پرسشنامه
تعامل پذیری	Interact1	تعداد فرم‌ها - تعداد صفحات شخصی سازی شده

روش معیار	TDQM	DWQ	TIQM	AIMQ	CIHI	DQA	IQM	ISTAT	AMEQ	COLDQ	DaQuinCIS	QAFD	CDQ	مقدار / بعد
Inter1		+												۱/۲
Inter2				+										۱/۲
Usa1			+											۱/۱
Integr 1			+											۱/۱
Conc1														۱/۲
Conc2							+							۱/۲
Main1				+										۱/۱
App1														۱/۱
App2							+							۱/۱
Conv1							+							۱/۱
Speed1							+							۱/۱
Comp1				+			+			+				۳/۳
Clar1			+				+			+				۳/۳
Trac1							+							۱/۱
Sec1							+							۱/۱
Sec2				+										۱/۱
Corr1							+							۱/۱
Obj1				+										۱/۱
Rel1				+										۱/۱
Rep1				+										۱/۱
Ease1				+										۱/۱
Interact1							+							۱/۱

معیارهای گردآوری‌شده با روش کیفیت انبار داده، دارای رابطه تنگاتنگی با فرایندهای استفاده‌کننده از داده‌ها هستند. تمام ابعاد براساس میزان انتظارات کاربران فرایندی خاص ارزیابی شده‌اند. مثلاً جامعیت به‌صورت نسبت مقادیر غیر تهی به کل تعداد مقادیر موردنیاز یک فرایند خاص تعریف می‌شود و نه با کل تعداد مقادیر ذخیره‌شده در پایگاه داده (بخش ۳-۲ را ببینید).

جدول ۲-۱۱ مقایسه از دیدگاه هزینه‌های کیفیت میان دو مدل TIQM و COLDQ

روش‌ها/دسته‌بندی هزینه‌ها	TIQM	COLDQ
هزینه ارزیابی فعالیت‌های کیفیت داده	هزینه ارزیابی یا بازرسی: • هزینه نرم‌افزار تحلیل کیفیت اطلاعات • زمان افراد در فرایند ارزیابی	هزینه‌های کشف مشکلات کیفی داده
هزینه بهبود فعالیت‌های کیفیت داده	هزینه بهبود فرایند و جلوگیری از عیب	هزینه‌های اصلاح داده‌ها هزینه‌های نگهداشت داده‌ها: • هزینه‌های اکتساب بالاسری • هزینه‌های واپاشی داده • هزینه‌های زیرساخت هزینه‌های بهبود فرایند
هزینه‌های فرایندی ناشی از کیفیت پایین داده‌ها	هزینه‌های خرابی فرایند: • هزینه‌های غیرقابل‌بازایی • هزینه‌های دیون و افشا • هزینه‌های بازایی مشتریان ناراضی ضایعات اطلاعاتی و استفاده مجدد: • هزینه‌های سامان دادن داده‌های زائد و پشتیبانی • هزینه‌های شکار یا تعقیب اطلاعات ازدست‌رفته • هزینه‌های کار مجدد تجاری • هزینه‌های رفع مشکلات و کاهش بهره‌وری • هزینه‌های تأیید داده‌ها • هزینه‌های بازنویسی نرم‌افزار • هزینه‌های حذف و اصلاح داده‌ها • هزینه‌های نرم‌افزار حذف داده‌ها	تأثیرات اجرایی: • هزینه‌های عقب‌گرد • هزینه‌های کار مجدد • هزینه‌های جلوگیری • هزینه‌های وارانته اثرات تاکتیکی و استراتژیک: • هزینه‌های تأخیر • هزینه‌های حق تقدم • هزینه‌های بی‌باری • هزینه‌های افزایش دشواری • هزینه‌های از بین رفتن دشواری • هزینه‌های بدگمانی سازمانی • هزینه‌های ناهمترازی
هزینه‌های فرصت کیفیت پایین داده‌ها	هزینه‌های فرصت ازدست‌رفته یا نادیده انگاشته شده: • هزینه‌های فرصت ازدست‌رفته • هزینه‌های فرصت نادیده انگاشته شده • ارزش ازدست‌رفته سهامدار	هزینه‌های منفعت ازدست‌رفته: • هزینه‌های تنیدگی • هزینه‌های کاهش • هزینه‌های استهلاک • هزینه‌های توقیف

۷-۲. روش‌ها و هزینه‌ها

بعد هزینه فقط در روش‌های مدیریت کیفیت جامع اطلاعات، ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها و روش جامع برای مدیریت کیفیت داده در نظر گرفته شده است. در این بخش هزینه‌ها را از دو نقطه نظر متفاوت تحلیل می‌کنیم:

۱. دسته‌بندی‌های هزینه

۲. معیارهای کمی کردن^۱ هزینه

۷-۲-۱. دسته‌بندی هزینه‌ها

روش مدیریت کیفیت جامع اطلاعات (انگلش ۱۹۹۹) و روش ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها، (لوشین ۲۰۰۴) جزئیاتی را برای دسته‌بندی هزینه‌ها ارائه کردند. دسته‌بندی سوم توسط اپلر و هلفرت^۲ در سال ۲۰۰۴ ارائه شده است.

جدول ۲-۱۲، دسته‌بندی روش‌های مدیریت کیفیت جامع اطلاعات و ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها را مقایسه می‌کند. در روش مدیریت کیفیت جامع اطلاعات به علت کیفیت پایین داده‌ها، هزینه کیفیت داده‌ها متناظر با هزینه فرایندهای تجاری و مدیریت داده است. هزینه‌های ارزیابی کیفیت اطلاعات، ابعاد کیفیت داده‌ها را برای تأیید اینکه فرایندها در حال عملکرد مطلوب هستند یا خیر، اندازه‌گیری می‌کند. سرانجام، هزینه بهبود فرایند یا جلوگیری از خرابی مطرح می‌شود که شامل فعالیت‌هایی برای بهبود کیفیت داده‌ها، با هدف حذف یا کاهش هزینه‌های کیفیت پایین داده‌ها است. در روش مدیریت کیفیت جامع اطلاعات هزینه‌ها، به علت کیفیت پایین داده‌ها تحلیل شده و به سه زیرگروه تقسیم می‌شوند:

۱. *هزینه‌های خرابی فرایند*. زمانی تحمیل می‌شوند، که داده‌های باکیفیت پایین

موجب فرایندی نادرست شوند. مثلاً نشانی پستی غیردقیق، منجر به

تحویل اشتباه محموله‌ها می‌شود؛

۲. **ضایعات اطلاعاتی و کار مجدد.** زمانی که داده‌ها دارای کیفیت پایینی باشند، نیازمند انواع مختلف فعالیت‌های مدیریت خرابی، مانند کار مجدد، حذف یا بازگردانی می‌شوند؛

۳. **هزینه‌های فرصت ازدست‌رفته.** این هزینه‌های متناظر با عواید و سودهای ازدست‌رفته، ناشی از کیفیت پایین داده‌ها است. مثلاً به علت دقت پایین نشانی ایمیل مشتری، کمپین‌های تبلیغات دوره‌ای نمی‌توانند به درصدی از مشتریان دست پیدا کنند، که این امر منجر به کاهش عواید می‌شود. روش ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها، هزینه‌های کیفیت پایین داده‌ها را تحلیل کرده و آن‌ها را باتوجه به اثر حوزه‌ای دسته‌بندی می‌کند:

- **حوزه عملیاتی،** که شامل اجرای سیستم پردازش اطلاعاتی و هزینه‌های عملیاتی سیستم است؛
- **حوزه تاکتیکی،** که تلاش می‌کند مسائل را پیش از قوت گرفتن آن‌ها شناسایی و حل کند؛
- **حوزه استراتژیک،** که تصمیمات بلندمدت را حمایت می‌کند. برای اثرات عملیاتی و اثرات تاکتیکی/استراتژیک، دسته‌بندی هزینه‌ها ارائه شده است. در اینجا هزینه‌های عملیاتی را تشریح می‌کنیم:
- **هزینه‌های عقب‌گرد،** زمانی تحمیل می‌شود، که کاری که انجام شده است باید به حالت قبلی بازگردانده شود؛
- **هزینه‌های کار مجدد،** زمانی تحمیل می‌شود که یک مرحله پردازش باید تکرار شود؛
- **هزینه‌های جلوگیری،** زمانی به وجود می‌آید که فعالیت جدیدی پیاده‌سازی می‌شود تا فعالیت ضروری برای جلوگیری از خرابی عملیاتی به علت مسئله کیفیت داده‌ها را انجام دهد؛
- **هزینه‌های وارانتی،** مربوط به تضمین بازگشت خسارات ناشی از ضرر و زیان است.

روش جامع برای مدیریت کیفیت داده دسته‌بندی را ارائه می‌کند که ناهمگنی میان روش‌های مدیریت کیفیت جامع اطلاعات، ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها و دیدگاه اپلر و هلفرت (۲۰۰۴) را برطرف می‌کند. برای جزئیات بیشتر، به مقاله باتینی و اسکناپیو (۲۰۰۶) مراجعه شود.

۲-۷-۲. معیارهای کمی‌سازی هزینه‌ها

ارزیابی مجموع هزینه کیفیت داده‌ها باتوجه به انواع فعالیت‌های کیفیت داده و اولویت‌های آن‌ها صورت می‌گیرد (به بخش ۲-۲ مراجعه شود). روش‌های مدیریت کیفیت جامع اطلاعات، ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها و روش جامع برای مدیریت کیفیت داده تنها روش‌هایی هستند که معیارها را برای این نوع فعالیت‌ها فراهم می‌کنند.

انتخاب و اولویت‌بندی در روش مدیریت کیفیت جامع اطلاعات با طی مراحل زیر حاصل می‌شود:

- کاربران فعلی که از داده‌ها استفاده می‌کنند را شناسایی می‌کند؛
 - خطاهایی را که اثر منفی بر کیفیت داده‌ها می‌گذارد را فهرست می‌کند؛
 - واحدهای تجاری که اغلب اوقات تحت تأثیر داده‌هایی باکیفیت پایین قرار می‌گیرند شناسایی می‌کند؛
 - هزینه مستقیم برنامه‌های فعلی کیفیت داده‌ها را تخمین می‌زند؛
 - هزینه خطاهای داده‌ها را برای تمام کاربران، تخمین زده و از داده‌ها استفاده می‌کند؛
 - از هزینه‌ها برای توجیه و اولویت‌بندی شروع برنامه‌های کیفیت داده استفاده می‌کند، که شامل ساختاربندی یک برنامه مستمر بهبود کیفیت است.
- هر نوع خطا در یک تناوب زمانی رخ می‌دهد و شامل هزینه‌ای می‌باشد. باید به یاد داشت که هزینه دسته‌های مختلف خطا یک مقدار مشروط است و با تغییر فرایندی که از داده‌ها استفاده می‌کند، تغییر می‌کند. مدل‌های ارائه فرایند، اجازه شناسایی

فعالیت‌های تحت تأثیر خطاهای داده‌ای را می‌دهند. از آنجایی که فعالیت‌ها با مجموع هزینه‌های سازمانی در ارتباط هستند، هزینه کار مجدد می‌تواند تخمین‌های کمی و عینی ایجاد کند.

روش ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها با حمایت از سرمایه‌گذاری در برنامه مدیریت دانش، بر ارزیابی هزینه‌های مرتبط با کیفیت پایین داده تمرکز دارد. این ارزیابی با طی مراحل زیر حاصل می‌شود:

- نقشه زنجیره اطلاعاتی را برای درک چگونگی جریان یافتن اطلاعات در سازمان، ترسیم می‌کند؛
- کاربران فعلی که از داده‌ها استفاده می‌کنند را شناسایی می‌کند؛
- اثر کیفیت داده‌ها را بر مشتریان شناسایی می‌کند؛
- داده‌های معیوب را با تخصیص به نواحی حساس ایزوله می‌کند؛
- حوزه‌های متأثر و مرتبط با هر نمونه از کیفیت پایین داده را شناسایی می‌کند؛
- هزینه‌های اقتصادی را براساس اثرات نهایی داده‌های بد مشخص می‌کند؛
- همه‌چیز را برای ارزیابی اثر کلی اقتصادی گردآوری می‌کند؛
- فرصت‌های بهبود را شناسایی می‌کند.

نتیجه این کار با عنوان کارت امتیاز^۱ کیفیت داده‌ها یاد می‌شود. این نتیجه‌گیری هزینه‌های مرتبط با کیفیت پایین داده را خلاصه می‌کند و می‌تواند به عنوان ابزاری برای یافتن بهترین راه‌حل بهبود استفاده شود.

در روش جامع برای مدیریت کیفیت داده، معیار اصلی برای انتخاب میان فرایندهای بهبود، حداقل‌سازی هزینه برنامه کیفیت داده است. ابتدا فرایندهای بهبود متفاوت در چارچوب روش‌های داده محور و فرایند محور اعمال شده و سپس براساس داده‌ها، جریان داده‌ها و اساس روش بهبود، روش متناسب شناسایی می‌شود.

پس از آن هزینه فرایندهای متفاوت ارزیابی و مقایسه شده و فرایندی با حداقل هزینه انتخاب می‌شود.

جدول ۲-۱۲ روش‌ها و انواع داده‌ها

نیمه ساختاریافته	ساختاریافته	مخفف نوع داده/روش
x	X	TDQM
	X	DWQ
در نظر گرفته شده به‌طور ضمنی	X	TIQM
در نظر گرفته شده به‌طور ضمنی	X	AIMQ
x	X	CIHI
	X	DQA
x	X	IQM
x	X	ISTAT
در نظر گرفته شده به‌طور ضمنی	X	AMEQ
در نظر گرفته شده به‌طور ضمنی	X	COLDQ
x	X	DaQuinCIS
	X	QAFD
X	X	CDQ

۲-۸. روش‌ها و انواع داده‌ها

در بخش ۵-۲ مشاهده کردیم که نوع داده‌ها بر ابعاد کیفیت داده و شیوه‌های ارزیابی و بهبود اثر می‌گذارند. در جدول ۲-۱۳، سه نوع داده دسته‌بندی شده در بخش ۵-۲ با روش‌های کیفیت داده مرتبط شده‌اند. اکثر روش‌ها، داده‌های ساختاریافته را آدرس‌دهی می‌کنند درحالی‌که فقط تعداد کمی از روش‌ها داده‌های نیمه ساختاریافته را آدرس‌دهی می‌کنند. در جدول ۲-۱۳ مشاهده می‌شود که هنگامی که روش دقیقاً به نوع داده اشاره نکرده باشد از عبارت «در نظر گرفته شده به‌طور ضمنی» استفاده می‌شود، اما مراحل و قدم‌ها می‌توانند بر آن اعمال شوند. مثلاً روش ارزیابی کیفیت اطلاعات از عبارات عمومی استفاده می‌کند و ارزیابی کیفی را از طریق پرسش‌هایی که بر داده‌های ساختاریافته اعمال می‌شود، انجام می‌دهد. اما ممکن است به هر نوع داده شامل داده‌های ساختار نیافته نیز اشاره داشته باشد.

در روش ISTAT داده‌های نیمه ساختاریافته، استانداردسازی قالب داده‌ها و بیان آن‌ها را در طرح رایج XML در نظر گرفته می‌شود. این طرح برای حداقل کردن اختلافات میان سازمان‌ها و نحوه همکاری آن‌ها با قسمت‌های دیگر پیاده‌سازی شده است. در روش کیفیت در سیستم‌های اطلاعات تعاونی، مدلی برای ایجاد ارتباط میان مقادیر کیفی و اسناد XML ارائه شده است (اسکاناپیکو ۲۰۰۴). سازمان‌های مختلف در یک سیستم اطلاعاتی تعاونی از مدل داده و کیفیت داده در حوزه جریان‌های داده‌ای تبادل شده استفاده می‌کنند. کیفیت جریان داده برای جلوگیری از انتشار خطا در تبادل داده‌ها حیاتی می‌شود. روش، می‌تواند به منظور تصدیق دقت، جامعیت، رایج بودن و همسانی داده‌ها استفاده شود. مدل‌های نیمه ساختاریافته به هر سازمان اجازه می‌دهند، کیفیت داده‌های خود را با درجه مشخصی از انعطاف‌پذیری، تنظیم کنند. کیفیت می‌تواند با عناصر متنوعی از مدل داده‌ها در ارتباط باشد که بازه آن از مقادیر داده‌های منفرد تا کل منبع داده‌ها متغیر است.

روش جامع برای مدیریت کیفیت داده قدم‌ها و شیوه‌هایی را برای داده‌های ساختاریافته، توسعه داده و برای داده‌های نیمه ساختاریافته نیز بسط داده است. مثلاً هر نوع داده اعم از ساختاریافته و نیمه ساختاریافته در گام بازسازی وضعیت، پیمایش شده است. یک شیوه مشخص بهبود داده محور نیز برای داده‌های غیر ساختاریافته ارائه شده است، که نام آن ثبت پروفایل داده‌ها^۱ است. این شیوه برای مرتبط کردن یک فایل متنی با طرح یک پایگاه داده به وسیله کشف الگوهای تکرار شونده در داخل متن، استفاده می‌شود (آیکن^۲ ۱۹۹۶).

جدول ۲-۱۳ روش‌ها و سیستم‌های اطلاعاتی

وب	تعاونی	انبار داده	توزیع شده	یکپارچه	مخفف نوع سیستم اطلاعاتی/روش
			در نظر گرفته شده به طور ضمنی	متمرکز	TDQM
		به شدت متمرکز			DWQ
			متمرکز	متمرکز	TIQM
			در نظر گرفته شده به طور ضمنی	متمرکز	AIMQ
			متمرکز	متمرکز	CIHI
			در نظر گرفته شده به طور ضمنی	متمرکز	DQA
به شدت متمرکز					IQM
	به شدت متمرکز		متمرکز	متمرکز	ISTAT
				متمرکز	AMEQ
				متمرکز	COLDQ
	به شدت متمرکز		متمرکز	متمرکز	DaQuinCIS
				متمرکز	QAFD
	به شدت متمرکز		متمرکز	متمرکز	CDQ

۲-۹. روش‌ها و انواع سیستم‌های اطلاعاتی

جدول ۲-۱۳ نشان می‌دهد که روش‌های مختلف تا چه حدی با انواع سیستم‌های اطلاعاتی معرفی شده در بخش ۲-۶، سروکار دارند. جدول از یک مقیاس چهار متغیره استفاده می‌کند که:

۱. کاملاً متمرکز بدان معناست که کل سازمان درک شده و روش مطابق با آن نوع سیستم اطلاعاتی طراحی شده است، درحالی‌که راهنماهای کلی برای سایر سیستم‌های اطلاعاتی، نیز فراهم است؛
۲. متمرکز بدان معناست که روش، راهنماها و شیوه‌های دقیقی را برای انواع سیستم اطلاعاتی فراهم می‌کند؛

۳. در نظر گرفته شده به طور ضمنی دارای همان معنای جدول ۲-۱۲ است؛

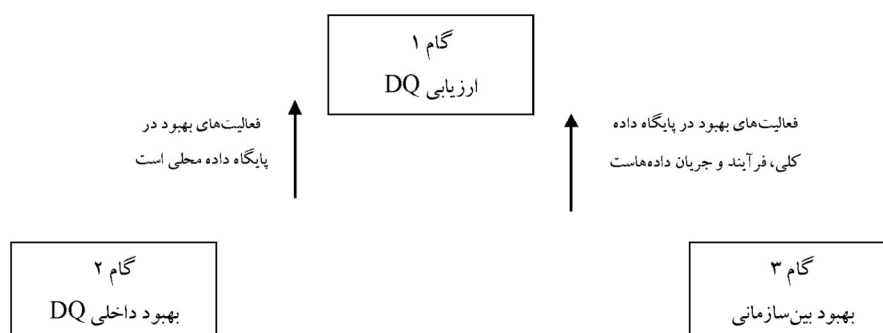
۴. مقدار خالی در جدول نشان می‌دهد که روش، رهنمودهای را کلی فراهم کرده اما به طور دقیق نوع سیستم اطلاعاتی متناظر را آدرس‌دهی نمی‌کند. از آنجایی که هر روشی بر روی ابعادی متمرکز است، پس سیستم‌های اطلاعاتی وب نیز در جدول گنجانده شده و دیگر مسائل مرتبط با چنین سیستم‌هایی در فصل ۳ مورد بحث قرار خواهد گرفت؛ ولی هیچ نوع روشی به سیستم‌های دوطرفه اشاره نکرده است، بنابراین در بخش نتیجه‌گیری و مسائل آتی به آن می‌پردازیم. می‌توان مشاهده کرد که روش‌های اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول، ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها و ارزیابی کیفیت داده‌های مالی بر سیستم‌های اطلاعاتی یکپارچه^۱ تمرکز دارند. این روش‌ها، داده‌های ساختاریافته را در یک سیستم منفرد در نظر گرفته و مسائل کیفیت داده ناشی از تبادل داده میان سازمان‌های مجزا را نادیده می‌گیرند.

همان‌گونه که در جدول ۲-۱۳ مشخص است، برخی روش‌ها مانند روش DWQ و IQM در یک نوع سیستم اطلاعاتی متمرکز است. برای نمونه مدل DWQ فقط بر انبار داده‌ها متمرکز بوده و در این دسته از سیستم‌های اطلاعاتی کاربرد دارد. از سوی دیگر برخی از روش‌ها مانند TDQM، ISTAT و CDQ در انواع مختلفی از سیستم‌های اطلاعاتی متمرکز هستند. برای نمونه روش TDQM در سیستم اطلاعاتی یکپارچه و توزیع شده کاربرد دارد و این نشان‌دهنده جامعیت و پوشش بالای این روش است.

روش مدیریت کیفیت جامع اطلاعات می‌تواند بر سیستم‌های یکپارچه و سیستم‌های توزیعی اعمال شود زیرا هنگامی که ساختار داده‌ها تحلیل شود، سیستم مرکزی و توزیعی فراهم می‌شود. روش موسسه کانادایی برای روش اطلاعات سلامت بر سیستم‌های توزیعی تمرکز دارد و استفاده از آن در پایگاه‌های داده ملی است.

سایر روش‌ها یعنی ISTAT، ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی و روش جامع برای مدیریت کیفیت داده می‌توانند بر چند نوع سیستم اعمال شوند. هر روش راهنماهای دقیقی را برای پیچیده‌ترین نوع سیستم‌ها یعنی سیستم اطلاعات تعاونی فراهم می‌کند. بنابراین می‌تواند بر سیستم‌های یکپارچه و توزیعی نیز اعمال شود. این روش‌ها از رویکردهای متفاوتی برای سامان دادن چند جریان داده و پایگاه‌های داده ناهمگن استفاده می‌کنند.

روش ISTAT دارای رویکرد قوی بین سازمانی بوده و برای مدیریت دولتی ایتالیا استفاده شده است. این روش دارای ساختار توزیعی با مدیریت کاملاً خودمختار برای شناسایی و انجام روابط بوده و در بسیاری از کشورها وجود دارد. سه گام اصلی روش در شکل ۲-۴ آمده است.



شکل ۲-۴ طرح کلی روش ISTAT (بتینی و دیگران ۲۰۰۹)

گام ارزیابی (گام ۱) مرتبط‌ترین فعالیت‌های گام بهبود (گام ۲ و ۳) را به شرح زیر شناسایی می‌کند:

۱. گام ۲ بر روی پایگاه‌های داده محلی که برای مدیریت‌های متفاوت طراحی شده عمل می‌کند. در این گام ابزارهایی طراحی شده تا فعالیت‌های مرتبط با کیفیت داده به‌طور خودمختار انجام و روندهایی نیز برای بهبود مهارت‌های محلی کیفیت ارائه شوند؛

۲. گام ۳ سیستم کلی اطلاعاتی تعاونی را برحسب جریان داده تبادل شده در نظر گرفته و پایگاه داده مرکزی را برای اهداف تعاونی تنظیم می‌کند. این فعالیت‌ها به‌طور متمرکز برنامه‌ریزی شده و به‌صورت هماهنگ انجام می‌گیرد. روش ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی چارچوبی را فراهم می‌کند که در آن خدمات و ابزارهای متنوعی برای سیستم‌های اطلاعاتی تعاونی ارائه می‌شوند (شکل ۲-۵). کارگزار^۱ کیفیت داده‌ها که در بخش ۳-۲ توصیف شده است هسته معماری و مسئول اصلاح پاسخ‌های ناهمگن به تفسیرهاست. خدمات آگاه‌سازی کیفیت یک موتور انتشاری/اشتراکی است که به‌عنوان پیام‌رسان میان اجزای معماری سازمان‌های همکار استفاده می‌شود (اسکاناپیکو و دیگران ۲۰۰۴). در سازمان‌ها به‌خاطر تغییرات کیفیت داده‌ها از این موتور استفاده می‌شود که اشتراکی مبتنی بر کیفیت را ممکن می‌سازد. سپس بهترین مقدار کیفیت انتخاب شده توسط کارگزار در سازمان‌ها انتخاب می‌شود تا سازمان داده‌های خود را با داده‌های باکیفیت بالاتر جایگزین کند (تابع بهبود کیفیت). چارچوب کارخانه کیفیت (شکل ۲-۵) مسئول ارزیابی کیفیت داده‌های داخلی هر سازمان است (برای مطالعه بیشتر به مقاله کاپیلو^۲ و دیگران ۲۰۰۳ مراجعه شود). در این چارچوب درخواست از کاربران بیرونی یا سیستم‌های اطلاعاتی در کارخانه کیفیت به‌وسیله تحلیلگر کیفیت پردازش می‌شوند و یک تحلیل ایستا از ابعاد کیفیت داده‌های مرتبط با داده‌های درخواستی ارائه می‌شود. سپس آن‌ها را با شاخص‌های کیفیت محک^۳ موجود در انبار کیفیت مقایسه می‌کنند. خدمات رتبه‌بندی نهایی در فصل بعد بررسی خواهند شد.

از پژوهش‌های دیگری که از ابزارهای پردازش تفسیری مبتنی بر کیفیت و حل تعارض به کمک نمونه‌برداری برای سیستم‌های اطلاعاتی تعاونی ارائه کرده‌اند، فیوژن پلکس (موترو و آنوخین^۴ ۲۰۰۵)، آی‌فویس (راهم^۵ و دیگران ۲۰۰۵) و هام‌مر (بیلک^۶ و دیگران ۲۰۰۵) را می‌توان نام برد.

1. Broker

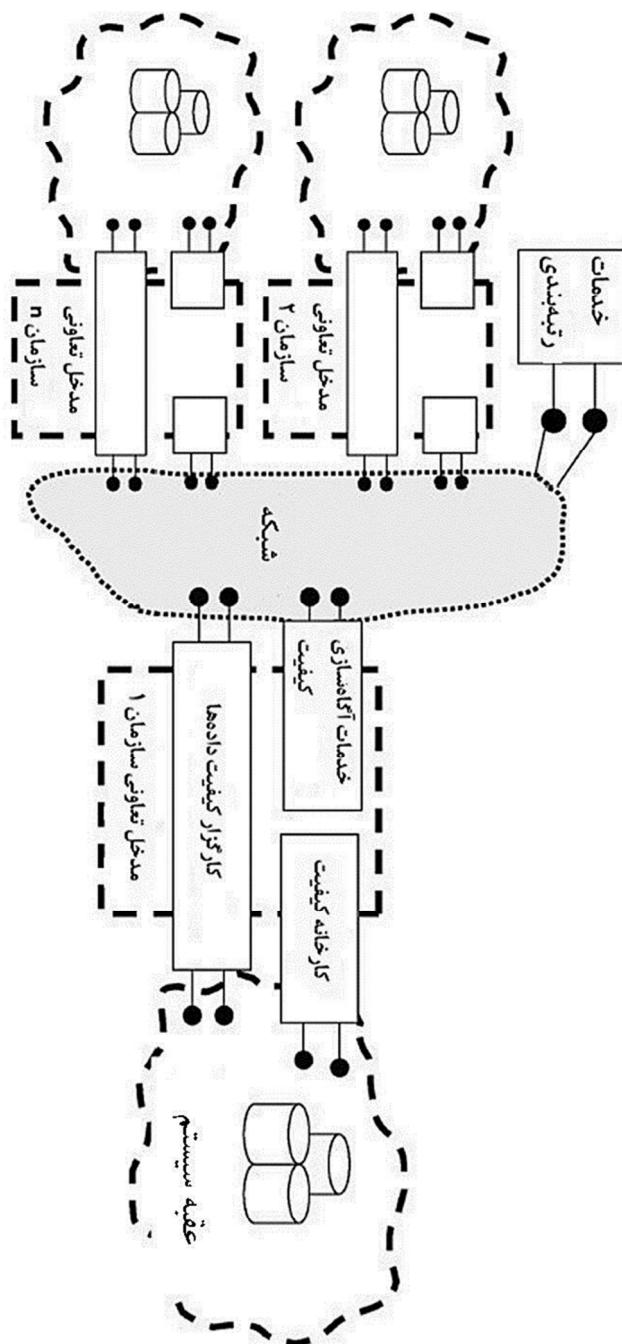
4. Motro And Anokhin

2. Cappiello

5. Rahm

3. Benchmark

6. Bilke



شکل ۲-۵ معماری چارچوب ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی (نبتی و دیگران ۲۰۰۹)

در این قسمت به روش‌هایی که بر انواع سیستم‌های اطلاعاتی خاص یعنی روش‌شناسی کیفیت انبار داده و اندازه‌گیری کیفیت اطلاعات کاملاً متمرکز هستند پرداخته می‌شود؛ بنابراین روش‌شناسی کیفیت استفاده شده مخصوص انبارهای داده است. در پژوهش جارک و دیگران (۱۹۹۵) مشاهده می‌شود که اکثر پژوهشگران در بررسی‌های خود بر روی روش کیفیت انبار داده، به نقش میانجی‌های میان سیستم‌های پردازش تحلیلی آنلاین^۱ و پشتیبانی تصمیم تمرکز کرده‌اند. این مسئله نقش سازمانی انبارسازی داده‌ها را برای کنترل مرکزی جریان‌های اطلاعاتی نادیده می‌گیرد. به این ترتیب، بسیاری از مسائل کیفیت مربوط به روش کیفیت انبار داده نمی‌توانند با فرامدل‌های سنتی روش کیفیت انبار داده بیان شوند. روش‌شناسی کیفیت انبار داده دو تغییر در نحوه حل این نوع مسائل انجام می‌دهد. ابتدا فراداده‌ها در معماری روش کیفیت انبار داده با مدل‌های صریح بنگاهی، غنی می‌شوند. سپس شیوه‌های ریاضی برای اندازه‌گیری یا بهینه‌سازی جنبه‌های متنوع کیفیت ارائه می‌شود. ابزار آرتکوس^۲ نیز می‌تواند برای آدرس‌دهی مسائل کیفیت داده مشارکت کند (واسیلیادیس^۳ و دیگران ۲۰۰۱). آرتکوس، فرامدلی متشکل از چند عنصر بوده که عبارتند از:

۱. فعالیت‌ها، واحدهای مجزای فعالیت پردازش داده‌ها؛
۲. سناریوها، مجموعه از فعالیت‌های که باهم اجرا می‌شوند؛
۳. کارخانه کیفیت، که برای هر فعالیت متناسب با ابعاد و سنجه‌های تعریف‌شده در بخش ۲-۳ تعریف‌شده است.

روش اندازه‌گیری کیفیت اطلاعات بر سیستم‌های اطلاعاتی وب کاملاً متمرکز است زیرا مجموعه وسیعی از ابزارهای موجود را برای ارزیابی کیفیت اطلاعات در حوزه وب در نظر می‌گیرد. ازجمله این ابزارها می‌توان به تحلیلگران وب، تحلیلگران ترافیک، اسکنرهای درگاه، سیستم‌های رصد عملکرد، ابزارهای وب‌کاوی و ابزارهای پیمایش برای تولید بازخورد مبتنی بر عقاید کاربر اشاره کرد. معیارهای کیفیت اطلاعاتی^۴ متنوعی می‌توانند با کمک این ابزارها اندازه‌گیری شوند. روش اندازه‌گیری

1. Online Transaction Processing, OLTP
3. Vassiliadis

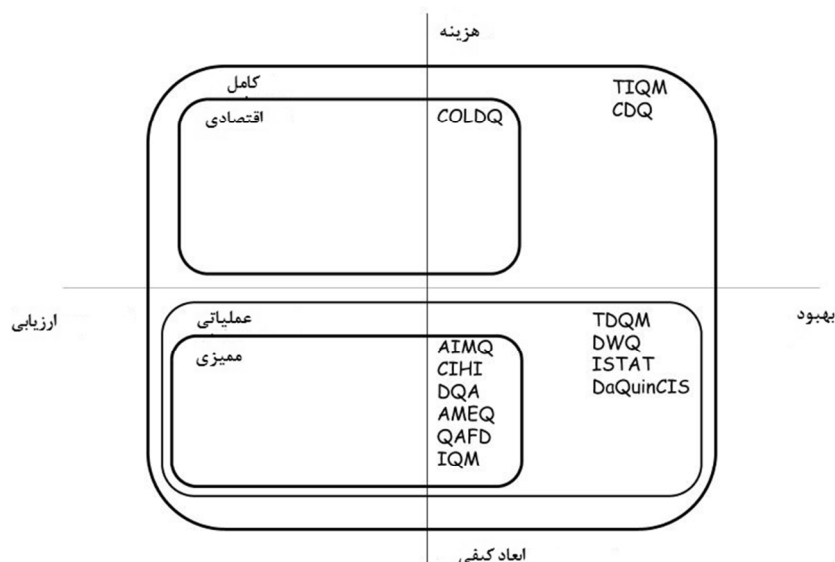
2. Artkos
4. Information Quality Criteria, IQC

کیفیت اطلاعات قدم‌های پی‌درپی و منظمی را برای تطبیق معیارهای کیفیت اطلاعات با ابزارهای اندازه‌گیری فراهم می‌کند.

۲-۱۰. مقایسه کوتاه روش‌ها

مقایسه کوتاه روش‌ها در بخش‌های قبلی به‌روشنی نشان می‌دهد که این روش‌ها بر زیرمجموعه‌ای از مسائل کیفیت داده تمرکز دارند. تفاوت‌های وسیع در نقاط مورد تمرکز روش‌ها را می‌توان از دیدگاهی در چهار گروه به شکل زیر تقسیم‌بندی کرد:

- **روش‌های کامل** که گام ارزیابی و گام بهبود را پشتیبانی می‌کنند، همچنین مسائل فنی و اقتصادی را هم پوشش می‌دهند؛
- **روش‌های ممیزی** که بر گام ارزیابی تمرکز داشته و پشتیبانی محدودی از گام بهبود ارائه می‌دهند؛
- **روش‌های عملیاتی** که بر مسائل فنی گام‌های ارزیابی و بهبود متمرکز بوده اما مسائل اقتصادی را پوشش نمی‌دهند؛
- **روش‌های اقتصادی** که بر هزینه‌های ارزیابی تمرکز دارند.



شکل ۲-۶ دسته‌بندی از روش‌ها

از نظر تاریخی، بین ابعاد کیفی و تکامل فناوری‌های ارتباطات و اطلاعات همبستگی وجود دارد. نسل اول سیستم‌های اطلاعاتی (دهه‌های ۵۰ و ۶۰) دارای معماری تکنولوژیک ویژه و شامل یک پردازنده مرکزی و یک پایگاه داده یکپارچه بودند. جریان‌های اطلاعاتی در آن‌ها ساده و تکراری بود و اکثر خطاها به‌وسیله داده‌های نادرست به وجود می‌آمد. از جمله کاربردهای مهم و در مقیاس بزرگ آن‌ها، مدیریت سرشماری و تحلیل داده‌های درمانی بود و کیفیت داده‌ها بر صحت، همسانی، جامعیت و ابعاد مرتبط با زمان تمرکز داشت (کاتارسی و اسکاناپیکو^۱ ۲۰۰۲). حیاتی‌ترین مسائل در مدیریت کیفیت این داده‌ها، شناسایی و اصلاح خطا در منابع داده و ثبت پیوند میان منابع داده جدید و منابع موجود بود.

تکامل سیستم‌های اطلاعاتی از حالت یکپارچه به حالت مبتنی بر شبکه، سبب رشد تعداد منابع داده هم در اندازه و هم در دامنه شد و به دنبال آن پیچیدگی مدیریت کیفیت داده‌ها به شدت افزایش یافت. روش‌های کیفیت داده بر ابعاد جدید کیفیت نظیر جامعیت منبع داده، رواج داده‌ها و همسانی منابع جدید داده با پایگاه داده بنگاهی تمرکز دارند. با ظهور وب و سیستم‌های اطلاعاتی نظیر آن ارزیابی و کنترل منابع داده دشوار شده است زیرا جستجو از طریق وب نامحدود است. به‌خاطر چنین تکامل سریعی، روش‌ها ابعاد کیفی جدیدی مانند قابلیت دسترسی و شهرت را پوشش دادند. قابلیت دسترسی، توانایی کاربران را برای دسترسی به داده‌ها براساس فرهنگ و وضعیت فیزیکی آن‌ها و فناوری‌های موجود اندازه می‌گیرد که در سیستم‌های اطلاعاتی تعاونی و مبتنی بر شبکه مهم است. شهرت یا قابلیت اعتماد معیاری در منابع داده است که توانایی آن‌ها را برای فراهم کردن اطلاعات صحیح را اندازه می‌گیرد و مشخصاً مرتبط با سیستم‌های مبتنی بر وب و دوطرفه است که در فصل بعد بیشتر توضیح داده خواهد شد.

همان‌گونه که در شکل ۲-۶ نشان داده شده است تکامل فناوری اطلاعات و ارتباطات و رشد پیچیدگی‌های کیفیت داده، دلایل اصلی در تمرکز بر روش‌های

1. Catarci And Scannapieco

مسائل کیفیت داده است. اکثر روش‌ها متعلق به دسته ممیزی هستند و بر مسائل فنی گام ارزیابی تمرکز دارند. این امر نیاز به ارزیابی کیفیت داده‌ها را ضروری می‌سازد زیرا به عنوان بخشی از فعالیت‌های بهبود در راستای ارزیابی اثربخشی شیوه‌های بهبود است. بحث ارزیابی کیفیت جزو مسائل پیچیده و جدید حوزه کیفیت داده است که مشخصاً از دید علمی جالب توجه هستند. برای مثال تعریف معیارهایی نظیر جامعیت و صحت، به شدت وابسته به تئوری پایگاه داده است و دامنه مسائل طراحی پایگاه داده سنتی نظیر دید یکپارچه و نرمال‌سازی طرح را وسیع می‌کند. در مقابل، مسائل اقتصادی ارزیابی و بهبود نیازمند یک رویکرد تجربی است. آزمایش شیوه‌ها در جهان واقعی می‌تواند چالش برانگیز باشد و فقط تعداد کمی از روش‌ها بر شواهد تجربی و اقتصادی کیفیت داده تمرکز دارند.

شایان ذکر است روش‌های ممیزی در گام ارزیابی، دقیق‌تر از روش‌های کامل و عملیاتی هستند زیرا با فراهم کردن مثال‌های بیشتر و دانش مرتبط، اطلاعات دقیق‌تری درباره شیوه انتخاب روش‌های مناسب ارزیابی ارائه می‌دهند. سپس هرگونه مسئله را صرف‌نظر از شیوه‌های بهبود شناسایی می‌کند؛ مثلاً روش‌های ارزیابی کیفیت اطلاعات و ارزیابی کیفیت داده‌های مالی به‌طور دقیق توضیح داده‌اند که ارزیابی عینی و غیر عینی چگونه قابل انجام بوده و راهنمایی را برای تفسیر نتایج ارائه می‌دهد. روش ارزیابی کیفیت داده درباره تعاریف عملیاتی بحث می‌کند که می‌تواند برای اندازه‌گیری ابعاد متفاوت کیفیت داده برای ارزیابی معیارهای تجمعی در پایگاه‌ها و منابع داده استفاده شود.

روش‌های عملیاتی بر ارزیابی کیفیت داده و شناسایی مسائلی تمرکز می‌کنند که رویکرد بهبود آن‌ها بهترین عملکرد را دارد. یکی از وظایف اصلی، شناسایی مجموعه‌ای از ابعاد مرتبط با بهبود و توصیف روش‌هایی برای ارزیابی آن‌ها است. مثلاً روش مدیریت کیفیت داده یک روش باهدف عمومی است و مجموعه کاملی از ابعاد مرتبط و روش‌های بهبودی را در حوزه‌های مختلف ارائه می‌کنند. وقتی که روش‌ها بر

حوزه‌های مشخصی تمرکز می‌کنند، جامعیت کاهش می‌یابد. مثلاً روش‌شناسی کیفیت انبار داده حوزه انبار داده را تحلیل کرده و ابعاد کیفی جدیدی را مطابق با معماری انبار داده تعریف می‌کند. تهیه فهرست ابعاد مرتبط، نقطه شروع مهمی برای فرایند بهبود است چراکه از شرکت‌ها در شناسایی مسائل کیفیت داده موثرتر بر انبار داده آن‌ها پشتیبانی می‌کند. باید توجه داشت که فرایند ارزیابی به‌طور دقیق در روش‌های عملیاتی که بر حوزه مشخصی تمرکز دارند توصیف شده‌اند. بنابراین تخصصی کردن روش‌های عملیاتی، جامعیت و عملیاتی بودن آن‌ها را در مقایسه با روش‌های کامل کاهش داده اما کارایی شیوه‌های ارائه‌شده را افزایش می‌دهد.

همچنین روش‌های عملیاتی زمانی که بر یک مسئله کیفیت داده خاص تمرکز می‌کنند، مؤثرتر هستند (خصوصاً برای فرایندهای داده محور). مثلاً روش ISTAT^۱ بر بومی‌سازی^۱ داده‌ها تمرکز دارد که آدرس‌های موجود در یک قلمروی مدیریتی را با داده‌های شخصی افراد مقیم، مرتبط می‌کند. چنین داده‌هایی به‌طور متوالی در میان آژانس‌ها تبادل می‌شوند. برای این کار روش ISTAT قالب‌های ناهمگن متفاوتی را به خدمت می‌گیرد تا بتواند برای مکان‌یابی داده‌ها و صاحبان داده‌های حوزه‌های متفاوت، اطلاعاتی مانند کد پستی را مصاحبه کرده و یک قالب متحد و رویکردی جدید را برای تبادل داده‌هایی براساس زبان XML ارائه کند.

برای نمونه، اگر مسائل کیفیت داده به دقیق بودن و جامعیت داده‌های شخصی مرتبط باشد، روش‌های بهبود می‌توانند در هدف‌گیری شیوه‌های ثبت، دقیق‌تر عمل کنند زیرا براساس آن روش‌های ISTAT و ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی به فراهم کردن الگوریتم‌هایی برای کشف موارد مشابه در پیوند ثبت و یکپارچه کردن منابع مختلف می‌پردازند. مثلاً حذف نام تکراری خیابان‌ها در نواحی دوزبانی نظیر ناحیه آلتو آدیج^۲ در ایتالیا با به خدمت گرفتن توابع مشابهت و خطاهای پارادایمی نظیر استفاده از u به جای $\bar{u}$ ، انجام شده است.

1. Localization

2. Alto Adige

مسئله دیگری که در روش‌های عملیاتی پوشش داده بررسی می‌شود نظارت بر فرایندهای کیفیت داده، امکانسنجی و مدیریت خطرهای مرتبط است. تفاوت اصلی روش ISTAT نسبت به سایر روش‌ها، تمرکز بر روی مسائلی با مقیاس بزرگ در تمام حوزه‌های مدیریت‌های مرکزی یک کشور است. مثلاً واحدهای سازمانی پراکنده و توزیع شده در کل قلمرو که وابسته به آژانس‌های مرکزی هستند. معمولاً چنین گروهی از مدیریت‌ها دارای ویژگی‌های کلیدی زیر هستند:

۱. پیچیدگی بالا برحسب رابطه متقابل، فرایندها و خدماتی که آژانس‌های متفاوت به علت جدا بودن دارند؛
۲. سطح بالایی از خودمختاری که اجرای قوانین را دشوار می‌کند؛
۳. ناهمگنی زیاد معنا و ارائه داده و جریان داده؛
۴. همپوشانی زیاد میان ثبت‌ها و موارد ناهمگن.

بهبود کیفیت داده در چنین ساختار پیچیده‌ای معمولاً یک فعالیت بسیار پرخطر و هزینه‌بر است. این مطلب خود دلیلی است برای آنکه چرا هدف اصلی روش ISTAT ممکن بودن راه حل است. بنابراین توجه اصلی، تمرکز بر متداول‌ترین نوع داده‌های تبادل شده میان آژانس‌ها است. در اولین مرحله بر روی مرتبط‌ترین داده‌ها و جریان داده‌ها برای کشف مهم‌ترین نواحی ارزیابی صورت می‌گیرد تا پیچیدگی‌های مسئله کاهش یابند. در مرحله بعد ارزیابی‌های عمیق‌تری بر پایگاه‌های داده محلی مدیریت شده توسط مدیریت‌های متفاوت تحت مسئولیت آن‌ها انجام می‌شود و این کار با استفاده از شیوه‌ها و ابزارهای متداول انجام می‌شود. در نهایت، تبادل داده میان آژانس‌هایی با مرکزیت مشخص، طرح‌ریزی و هماهنگی می‌شود.

روش‌های کامل در فراهم‌سازی یک چارچوب جامع برای هدایت برنامه‌های بزرگ کیفیت داده در سازمان‌هایی پردازش‌کننده داده‌های کلیدی مؤثر هستند. از سوی دیگر، آن‌ها تعامل سستی میان قابلیت اجرایی روش و عدم هماهنگی آن در حوزه‌های مشخص یا زمینه‌های فنی را نشان می‌دهند. روش‌های کامل فقط به‌طور

حاشیه‌ای تحت تأثیر نوع، منابع و جریان داده قرار می‌گیرند که بخشی از سیستم‌های مدرن هستند. مثلاً مدل مدیریت کیفیت داده «نقشه پردازش اطلاعات» به‌منظور مدیریت رشد پیچیدگی سیستم‌ها برحسب فرایندها و عملگرها تا نقشه مدل اطلاعاتی یکپارچه^۱ رشد یافته است. هرچند نقش و استفاده آن در چارچوب کلی روش مدیریت کیفیت داده به شکل زیادی تغییر پیدا نکرده است.

شیوه‌های اقتصادی سایر روش‌ها را تکمیل کرده و می‌توانند به‌آسانی در چارچوب کلی ارائه‌شده توسط هر روش کاملی جای گیرد. اکثر روش‌های ممیزی و بهبود دارای هزینه ارزیابی هستند (جدول ۲-۱۱). اگرچه آن‌ها عمدتاً بر هزینه آغازین کیفیت داده تمرکز می‌کنند اما تحلیل کامل هزینه-فایده باید هزینه انجام ندادن کار را نیز در نظر بگیرد. هزینه کیفیت پایین داده‌ها معمولاً به‌خاطر طبیعت سازمانی است. روش‌های اقتصادی بر هر دو جنبه تمرکز دارند. مثلاً روش ارزیابی هزینه مؤثر کیفیت پایین داده‌ها بر ارزیابی هزینه‌های مرتبط با کیفیت پایین داده‌ها تمرکز دارد و اثر اقتصادی مبتنی بر اثرات نهایی داده‌های بد را ساختاربندی می‌کند. نتیجه آن با نام کارت امتیاز کیفیت داده شناخته می‌شود و می‌تواند به‌عنوان ابزاری برای یافتن بهترین راه‌حل برای بهبود استفاده شود. در روش جامع برای مدیریت کیفیت داده، ارزیابی کلی هزینه‌های کیفیت پایین توسعه بیشتری یافته است. در این روش به این واقعیت توجه نشده است که همان بهبود کیفیت می‌تواند با اولویت‌ها و مسیرهای متفاوت نیز به دست آید و حداقل‌سازی هزینه برنامه کیفیت داده، معیار اصلی برای انتخاب فرایندهای بهبود جایگزین است. در این روش می‌توان از فرایندهای بهبود مختلفی برای اعمال شیوه‌های داده محور و فرایند محور بر پایه داده‌ها و جریان داده‌های مستند و سپس شناسایی روند بهبود استفاده کرد.

در ادامه، صحت روش‌ها در حوزه اجرا و مهم‌ترین تجربیات گزارش‌شده در پژوهش‌های پیشین را شرح می‌دهیم. به علت پراکندگی تجربیات گزارش‌شده، این قسمت به‌طور مورد به مورد تشریح می‌شود. تأیید تجربی روش‌ها غالباً مورد غفلت

1. Information Product Maps Unified Modeling Language, IP-UML

واقع شده، یا براساس مطالعه موردی کوچکی است و به ندرت با مقیاس بزرگ آزمایش‌های علمی سنجیده شده است. این پیاده‌سازی‌ها معمولاً در حوزه صنایع اتفاق می‌افتند و بخشی از اهداف آن برای مشاوره گرفتن در آن حوزه است و نحوه آزمون فرایند ناگفته باقی می‌ماند.

تجربه استفاده از نسخه‌های اولیه روش مدیریت کیفیت داده در اسناد وزارت دفاع ایالات متحده^۱ در سال ۱۹۹۴ گزارش شده است. هم‌اکنون استفاده از ابزارهای توسعه‌یافته کیفیت داده در بستر رویکردهای برنامه‌نویسی و اس.کیو.ال برای بررسی کیفیت داده‌ها مورد توجه است. از این روش‌های پایه‌ای در حوزه‌های قانونی نیز استفاده شده است (سشنز^۲ ۲۰۰۷). سایر کاربردهای روش مدیریت کیفیت داده در تسهیلات درمانی ارتش و حوزه درمان وزارت دفاع ایالات متحده نیز گزارش شده است (وانگ ۱۹۹۸) (کوری^۳ و دیگران ۱۹۹۶). یک مزیت روش مدیریت کیفیت داده این است که می‌تواند براساس مسائل حیاتی کیفیت داده و داده‌های متناظر با آن‌ها به صورت کارآمد، نمایش‌گری و جامعیت معیارهای کیفیت داده و اینکه آیا مجموعه درستی از معیارها ارزیابی شده است یا خیر را مشخص کند. به همین دلیل از این روش در حوزه‌های مختلف نظیر شرکت‌های بیمه استفاده شده است (نادکارنی^۴ ۲۰۰۶؛ تسچن^۵ ۲۰۰۵). امروزه با بهبود مدل نقشه پردازش اطلاعات (شانکارانایارایانان و ووانگ^۶ ۲۰۰۷؛ اسکاناپیکو و همکاران ۲۰۰۵) یا با ارائه مدل کامل مبتنی بر مدیریت کیفیت داده (باسکارادا و همکاران^۷ ۲۰۰۶) تلاش‌های بیشتری برای توسعه مدیریت کیفیت داده انجام شده است. روش مدیریت کیفیت اطلاعات کلی یک روش حرفه‌ای است و کاربردهای بسیار متنوعی از این روش در زندگی واقعی و مطالعات موردی زیادی در مقاله (انگلیش ۱۹۹۹) آمده است که از مدیریت رابطه با مشتری تا بازاریابی از راه دور و حوزه سلامت را در برمی‌گیرد. تعداد زیادی از ابزارهای کیفیت داده نیز مورد مقایسه قرار گرفته‌اند. قواعد مدیریتی و داستان‌های

1. Us Department Of Defense
2. Bettchen

2. Sessions
6. Shankaranarayanan And Wang

3. Corey

4. Nadkarni
7. Baskarada

موفق بی‌شماری در وبگاه‌های مرتبط با تاثیر اطلاعات^۱ نیز آمده است که مثال‌هایی از تجارب حرفه‌ای مرتبط با حوزه استفاده از کیفیت داده برای کاهش هزینه، بهبود ارزش اطلاعاتی و اثربخشی تجاری را ارائه می‌کند.

در سال‌های زیادی برای سازمان‌دهی، نگهداشت و بهبود اطلاعات سلامت در بیست پایگاه داده مدیریت درمانی از روش موسسه کانادایی استفاده شده که بسیاری از آن‌ها منحصر به فرد هستند (لانگ و سکو ۲۰۰۵). شواهدی در سراسر کانادا وجود دارد که اثر روش موسسه کانادایی برای روش اطلاعات سلامت در حمایت از مدیریت کارآمد حوزه سلامت و توسعه سیاست‌های عمومی را نشان می‌دهد (چاپمن^۲ و دیگران ۲۰۰۶). روش موسسه کانادایی برای روش اطلاعات سلامت همچنین به‌عنوان مدل مرجع در سایر کشورها نیز استفاده می‌شود. برای مثال، این روش در کنار روش مدیریت کیفیت داده برای استقرار یک استراتژی کیفیت داده برای وزارت بهداشت در نیوزیلند^۳ استفاده شده است (کر و نوریس^۴ ۲۰۰۴) پس از اجرای بلندمدت این روش، مؤلفان داستان‌های موفق‌تری ارائه کردند که نشان می‌دهد فرایند ارزیابی در زمان استفاده از روش موسسه کانادایی برای روش اطلاعات سلامت برای دستیابی به هدف اولیه شناسایی و رتبه‌بندی مهم‌ترین مسائل بهبود کیفیت داده موفق عمل کرده است. روش موسسه کانادایی برای روش اطلاعات سلامت نه تنها نواحی مناسب را برجسته می‌کند بلکه نواحی مشکل‌دار در پایگاه داده را نیز هدف‌گیری کرده و کنترل وضعیت کیفیت داده برای داده‌های بیشمار را نیز تسهیل می‌کند. همچنین ارزیابی امتیازات می‌تواند برای توصیف کیفیت داده‌ها در یک موسسه استفاده شود. به نظر می‌رسد که برنامه‌ریزی راهبردی و مشارکتی در زمان ارائه نتایج ارزیابی تسهیل ایجاد کند. اگرچه در این حالت تعداد ارزیابی‌ها اندک است، اما پیشرفت‌های زیادی در پایگاه داده به‌دست آمده و بسیاری از این پیشرفت‌ها هنوز ناشناخته هستند. محدودیت‌های اصلی تجربه‌شده در روش موسسه

1. Info Impact

2. Chapman

3. New Zealand

4. Kerr And Norris

کانادایی برای روش اطلاعات سلامت این است که ویژگی‌های اندازه‌گیری فرایند ارزیابی، هنوز شناخته‌شده نیست و فقط داده‌های اولیه در زمان مطالعه در دسترس هستند و پس از اجرای طرح نتایجی در ارتباط با آن داده‌ها ارائه می‌شوند.

پژوهشگران روش ارزیابی کیفیت داده‌های مالی ادعا می‌کنند که کار آن‌ها نتیجه سال‌ها تجربه در حوزه کیفیت داده‌های مالی بوده و توسعه شیوه‌های تحلیل کیفیت در این روش ساده و ارزان است. آن‌ها همچنین یک سناریوی واقعی را ارائه کردند که به دلایل امنیتی تمام مدارک آن را ارائه نکردند.

در مقاله باتینی و اسکناپیکو (۲۰۰۶) تجربه بزرگ‌مقیاسی از اجرای روش جامع مدیریت کیفیت داده گزارش شده است که به سازمان‌دهی مجدد روابط دولتی تجارت در ایتالیا اشاره دارد. بررسی نحوه تعامل با دولت برای چندین رویداد تجاری نظیر آغاز یک کسب‌وکار جدید و گسترش آن شامل نوساناتی در وضعیت قانونی، ترکیب هیئت‌مدیره، مدیریت ارشد و تعداد کارکنان موردنیاز است. آژانس‌ها در تعامل خود با کسب‌وکارها، اطلاعات مشترکی را مدیریت می‌کنند که معمولاً شامل نام رئیس، مکان شرکت، نشانی شعبه‌ها و فعالیت اصلی اقتصادی است. از آنجایی که هر کسب‌وکار به صورت مستقل به هر آژانس گزارش می‌دهد، نسخه‌ها دارای سطوح متفاوتی از دقت هستند. به این منظور روش جامع برای مدیریت کیفیت داده در این حوزه اعمال شده است که نتایج زیر را موجب می‌شود:

- اجرای فعالیت‌های مبتنی بر پیوند رکوردها باهدف ایجاد ارتباط میان شناساگرهای کسب‌وکار در مدیریت‌های متفاوت؛
- مهندسی مجدد رابطه دولت با کسب‌وکار.

در سناریوی مهندسی مجدد، کسب‌وکارها فقط با یک مدیریت در تعامل هستند که مسئول ارسال مقدار یکسانی از اطلاعات به سایر مدیریت‌هاست. با در نظر گرفتن یک دوره سه‌ساله با محدودیت‌های هزینه‌ای و صرفه‌جویی در ارتباط باکیفیت داده‌ها، منافع به‌کارگیری روش جامع برای مدیریت کیفیت داده حدود ۶۰۰ میلیون

یورو تخمین زده شده است. این امر اعتبار روش را اثبات می‌کند که در آن فعالیت‌های داده محور و فرایند محور در نظر گرفته شده و برنامه کیفیت داده‌ها با بهینه‌سازی نسبت هزینه به کیفیت انتخاب شده است. سایر کاربردهای روش جامع برای مدیریت کیفیت داده در مقاله باسیل^۱ و دیگران (۲۰۰۷) گزارش شده است. مطالعات مبتنی بر روش جامع مدیریت کیفیت داده به‌عنوان روش مرجع برای ارزیابی، اندازه‌گیری و کاهش خطرهای عملیاتی مرتبط با فرایند پایه‌ای نشان می‌دهند که این روش در برخی موارد منجر به صرفه‌جویی‌های فراوان سالیانه شده است.

۲-۱۱. از کیفیت داده به سمت کیفیت اطلاعات

در این فصل به سیزده گونه از روش‌های کیفیت داده اشاره و مقایسه می‌شود. برای این منظور برخی از پژوهش‌های کیفیت داده اخیر بررسی شد. روش‌های سنجش و بهبود کیفیت داده‌ها به چندین شیوه در حال تکامل هستند:

۱. در نظر گرفتن تعداد بیشتری از انواع داده‌ها، حرکت از سمت کیفیت داده

به سوی کیفیت اطلاعات؛

۲. ربط دادن بیشتر موضوعات کیفیت داده به موضوعات فرایند کسب‌وکار؛

۳. در نظر گرفتن انواع جدید سیستم‌های اطلاعاتی مخصوصاً وب و

سیستم‌های اطلاعات دوطرفه.

تجزیه و تحلیل‌های پیشین بر این حقیقت تأکید دارند که موضوعات کیفیت اطلاعات تنها حوزه داده‌های نیمه ساختاریافته، داده‌های ساختاریافته و داده‌های چندرسانه‌ای را به‌طور کلی مورد توجه قرار داده‌اند. در واقع، روش‌ها و شیوه‌ها از مدت‌ها پیش برای این نوع داده‌ها در بخش‌های پژوهشی ایجاد شده‌اند. برای نمونه می‌توان درک زبان طبیعی برای اسناد و مدارک در مواردی که غنی‌سازی و بهبود خاصی به کمک شیوه‌های کیفیت داده روی آن‌ها رخ نداده است را نام برد.

پژوهش‌های انجام‌شده بر روی داده‌های نیمه ساختاریافته و ساختار نیافته در حوزه کیفیت داده حاکی از ارتباط تاریخی محکم بین کیفیت داده و طراحی پایگاه داده است. حتی روش‌های کامل کیفیت داده دارای سوگیری و پیشداوری هستند اما تمرکز اساسی آن‌ها بر مجموعه‌ای از داده‌های ساختاریافته است که بیشترین منابع اطلاعاتی بالغ در بیشتر سازمان‌ها را دارا هستند (باتینی و دیگران ۲۰۰۹). امروزه حرکت به سمت داده‌های نیمه ساختاریافته و فاقد ساختار به‌عنوان منابع سازمانی بسیار رایج‌تر است. مدیریت دانش، حفاظت وب و سیستم‌های اطلاعات جغرافیایی حوزه‌های پژوهشی مهمی هستند. همین‌طور شیوه‌های کیفیت داده برای داده‌های نیمه ساختاریافته و فاقد ساختار اخیراً مورد بررسی قرار گرفته‌اند (باتینی و اسکناپیکو ۲۰۰۶). بهبود شیوه‌های کیفیت داده برای داده‌های فاقد ساختار و نیمه ساختاریافته در این حوزه‌ها نیازمند درجه بالاتری از ارتباط بین‌رشته‌ای است که البته ممکن است با تأخیر بسیار زیادی انجام پذیرد.

۲-۱۲. کیفیت داده‌ها و کیفیت فرایند

ارتباط بین کیفیت داده و کیفیت فرایند به دلیل ارتباط و تنوع ویژگی‌های فرایندهای کسب‌وکار در سازمان‌ها، بخش عمده‌ای از پژوهش‌ها را دربرمی‌گیرد. تأثیرات متفاوت کیفیت داده‌ها در سه سطح مشخص به نام‌های عملیاتی، تاکتیکی و سطوح استراتژیک در پژوهش ردمن (۱۹۹۸) مورد بررسی قرار گرفته است. در این کتاب مصاحبه‌ها و خروجی‌های چندین مطالعه خاص گزارش شده است. کیفیت داده و ارتباط آن با کیفیت خدمات، محصولات، عملیات کسب‌وکار و رفتار مصرف‌کننده در قالب اصطلاحات عمومی زیادی در مقاله شنگ و میکیتین^۱ (۲۰۰۲) و شنگ (۲۰۰۳) مورد بررسی قرار گرفته‌اند. در این بررسی‌ها گزاره‌های عام مانند «کیفیت اطلاعات یک شرکت به‌صورت مثبت با عملکرد شرکت رابطه دارد» مبتنی بر شواهد تجربی استفاده شدند. همچنین این مسئله که چگونه بهبود فرایندهای تولید اطلاعات

1 Sheng And Mykytyn

به صورت مثبت بر داده‌ها و کیفیت اطلاعات تأثیر می‌گذارد نیز مورد تجزیه و تحلیل قرار گرفته است (انگلیش ۲۰۰۲).

تعدادی از مقالات به موضوعات خاصی اشاره داشته و در نتیجه نتایج محکم‌تری را نسبت به موضوعاتی ارائه می‌دهند. ورمیر^۱ (۲۰۰۰) موضوع تبادل الکترونیکی داده^۲ را بررسی می‌کند که مرتبط با تبادل داده‌ها در میان سازمان‌هایی است که از فرمت‌های استاندارد استفاده می‌کنند و چگونگی تأثیر آن بر بازدهی و اثربخشی فرایندهای کسب و کار و کیفیت داده در نظر گرفته شده است. تبادل الکترونیکی داده به صورت عمومی‌تر از زبان‌های نشانه‌گذاری به عنوان یک توانمندسازی فناوری استفاده می‌کند زیرا آن‌ها به طور بالقوه فعالیت‌های جابجایی کاغذ، خطای وارد کردن داده‌ها و وظایف وارد کردن داده‌ها را کاهش می‌دهند.

تأثیر کیفیت داده‌ها بر روی اثربخشی فعالیت‌های تصمیم‌گیری در پژوهش راقاناتان^۳ (۱۹۹۹) با تمرکز بر روی بعد دقت بررسی شده است. این تجزیه و تحلیل نشان می‌دهد که اثربخشی تصمیمات با کیفیت داده بهتر فقط زمانی که تصمیم‌گیرنده در مورد ارتباط میان متغیرهای مسئله دانش داشته باشد، بهبود می‌یابد و در غیر این صورت در جهت مخالف تنزل خواهد کرد.

تأثیر کیفیت داده‌ها در شرایط بحرانی مانند بلایا و مواقع اضطراری در پژوهش فیشر و کینگما^۴ (۲۰۰۱) مورد بحث قرار گرفته است. کاستی‌های ابعاد دقت، جامعیت، سازگاری و به هنگام بودن با رجوع به موقعیت‌های بحرانی مورد توجه قرار گرفته‌اند. مانند زمانی که ناوگان دریایی وینسنز^۵ آمریکا به ایرباس ایرانی شلیک کرد و باعث کشته شدن ۲۹۰ نفر شد.

نقش اطلاعات در زنجیره تأمین در کتاب دیدیکه^۶ (۲۰۰۵) مورد بحث قرار گرفته است. در این بررسی توان کیفی زنجیره اطلاعات برای اندازه‌گیری قابلیت فرایند تولید و همچنین برای ساخت محصول نهایی در مواردی که تهدیدات باعث

1. Vermeer

4. Fisher And Kingma

2. Electronic Data Interchange, EDI

5. Vincennes

3. Raghunathan

6. Dedeke

به هم ریختن اطلاعات، تغییر در حین تبدیل و نقص اطلاعات شده پیشنهاد می‌شود. در این بررسی یک چارچوب به نام طراحی توان کیفیت فرایند^۱ برای تشخیص، تجویز و ساخت کیفیت در میان زنجیره‌های اطلاعاتی پیشنهاد شده است.

۲-۱۳. کیفیت داده‌ها و انواع سیستم‌های اطلاعاتی جدید

با سیر تکاملی فناوری، ماهیت سیستم‌های اطلاعاتی هم‌تغییر کرده و انواع جدیدی از سیستم‌های اطلاعاتی در حال ظهور هستند. در این پژوهش موضوعاتی را تجزیه و تحلیل می‌کنیم که مربوط به وب و سیستم‌های اطلاعاتی دوطرفه هستند. نگرانی در مورد سیستم‌های اطلاعاتی باعث شده که روش‌شناسان چند مشکل عمده را مطرح نمایند:

۱. کیفیت داده‌های فاقد ساختار به‌ویژه اسناد و مدارک رو به افزایش است؛
 ۲. انواع جدیدی از ابعاد کیفیت مانند در دسترس بودن در حال تغییر است.
- پرنیسی و اسکاناپیکو^۲ (۲۰۰۳) مدلی را پیشنهاد می‌کنند که بازدهی اطلاعات را با پیوند داده‌های وب، ذکر نام هر آیتم در یک صفحه وب و ارتباط آن با صفحات و با گروهی از صفحات بالا می‌برد. این مدل درون روش طراحی کیفیت داده و مدیریت سیستم‌های اطلاعات وب به کار گرفته می‌شود. برخی نویسندگان در مورد چگونگی غنی‌سازی روش‌ها برای طراحی سیستم اطلاعات وب بحث می‌کنند. مثلاً مکا^۳ و دیگران (۱۹۹۸) و ایساکوئیز و دیگران (۱۹۹۵) با گام‌های اضافی مخصوص طراحی کیفیت داده را تشریح کرده‌اند و در این بررسی‌ها، چند بعد از جمله فراریت، کامل بودن، صحت محتوایی و معنایی مورد توجه قرار گرفته‌اند.
- موضوعات مرتبط با کیفیت مستندات وب رو به افزایش است زیرا تعداد مستنداتی که در فرمت وب مدیریت می‌شوند به‌طور پیوسته در حال افزایش هستند. مثلاً مقاله رائو^۴ (۲۰۰۳) نشان داده است که ۴۰ درصد مطالب روی شبکه در طول یک سال از بین می‌روند، ۴۰ درصد تغییر می‌یابند و تنها ۲۰ درصد به شکل اولیه

1. Design For Quality Process Capability

2. Pernici And Scannapieco

3. Mecca

4. Rao

باقی می‌مانند. مطالعات دیگری مانند لیمان و واریام^۱ (۲۰۰۳) نشان داده است که متوسط طول عمر یک صفحه وب ۴۴ روز است و هرسال چهار بار تغییر کامل وب رخ می‌دهد و در نتیجه حفظ داده‌های وب مهم و حیاتی است. اصطلاح حفظ وب نشان‌دهنده توانایی جلوگیری از دست رفتن اطلاعات در وب با ذخیره‌سازی همه ورژن‌های مهم مستندات وب است.

مقاله کاپیلو^۲ و دیگران (۲۰۰۳) یک روش برای پشتیبانی از حفظ فرایند در طول کل چرخه عمر اطلاعات (ایجاد، کسب، ذخیره و دسترسی) ارائه می‌دهد. گام‌های اصلی این روش عبارت‌اند از:

۱. هر وقت یک صفحه جدید منتشر می‌شود، داده‌ها با فراداده‌هایی که توصیف‌کننده کیفیت آن‌ها برحسب صحت، کامل بودن، سازگاری، جدید بودن و فرارایت مرتبط می‌شوند. همان‌گونه که در بخش ۲-۲ تعریف شد؛
۲. در گام جمع‌آوری اطلاعات، کاربر مقادیر قابل قبولی برای همه ابعاد کیفیت را مشخص می‌کند؛
۳. اگر داده‌های جدید الزامات کیفیت را برآورده کنند، در یک آرشیو جای داده می‌شوند و در غیر این صورت داده‌ها به صاحبانشان بازگردانده می‌شوند؛
۴. در مرحله انتشار وقتی یک صفحه جدید جایگزین یک صفحه قدیمی می‌شود، ناپایداری داده‌های قدیمی ارزیابی می‌شود، اگر داده‌های قدیمی معتبر باشند، حذف نمی‌شوند و با یک صفحه با آدرس جدید پیوند می‌خورند.

روش‌های ارزیابی کیفیت‌های خاص سایت‌های وب در پژوهش‌های آتzeni^۳ و دیگران (۲۰۰۱)، فراترالی^۴ و دیگران (۲۰۰۵) و مکا و دیگران (۱۹۹۹) بررسی شده‌اند و به‌طور خاص روی میزان دسترس‌پذیری آن‌ها مطابق با بخش ۲-۲ تمرکز دارند. نهایتاً این معیارها براساس روشی

1. Lyman And Variam

2. Capiel

3. Atzeni

4. Fraterrali

ترکیبی کمی و کیفی ارزیابی می‌شوند. در ارزیابی کمی، خطوط راهنمای ارائه‌شده توسط سازمان جهانی وب^۱ را بررسی می‌کنند. ارزیابی کیفی براساس آزمایش‌های انجام‌شده بر روی کاربران از دسترس خارج‌شده^۲ صورت می‌گیرد.

سیستم‌های دوطرفه کاملاً باز هستند و اهمیت ارزیابی و فیلتر داده‌ها در آن‌ها مطرح می‌شود. ابعاد کیفیت مرتبط با اعتماد در چنین سیستم‌هایی هنوز موضوعاتی پایان نیافته در بررسی‌ها هستند. یک‌راه حل ممکن برای حل آن تأکید بر شهرت در مقایسه با قابل اعتماد بودن هر طرف است. مثلاً سرویس رتبه‌بندی ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی، مقادیر اعتماد را با هر منبع داده در سیستم اطلاعات پیوند می‌زند.

سرویس رتبه‌بندی متمرکز بوده و فرضش بر این است که توسط یک سازمان سوم اجرا شود. معیار اعتماد برای تعیین میزان قابلیت اعتماد ارزیابی‌های کیفی انجام‌شده توسط سازمان‌ها به کار می‌روند. تمرکز معیارها براساس مطالعه دی‌سانتیس^۳ و دیگران (۲۰۰۳) بر روی شکایت کاربران از مجموعه داده‌ها که توسط آن‌ها اعلام‌شده می‌باشد. معیارهای پیشنهادشده در مطالعه گاکوسکی^۴ (۲۰۰۶) منابع داده‌ای دقیقی که داده‌ها را با منابع باکیفیت بالای دقیق مبادله می‌کنند در نظر می‌گیرند. این معیارها همچنین شامل یک مؤلفه موقتی هستند که دقت گذشته منابع داده را به حساب می‌آورد. سایر پیشنهادها شامل بررسی شیوه‌هایی است که می‌توانند به معتبر بودن مقادیر کمک کنند (گاکوسکی ۲۰۰۶).

1. World Wide Web

2. Disabled Users

3. De Santis

4. Gackowski

فصل سوم

مدل‌های مدیریت کیفیت داده

۳-۱. مقدمه

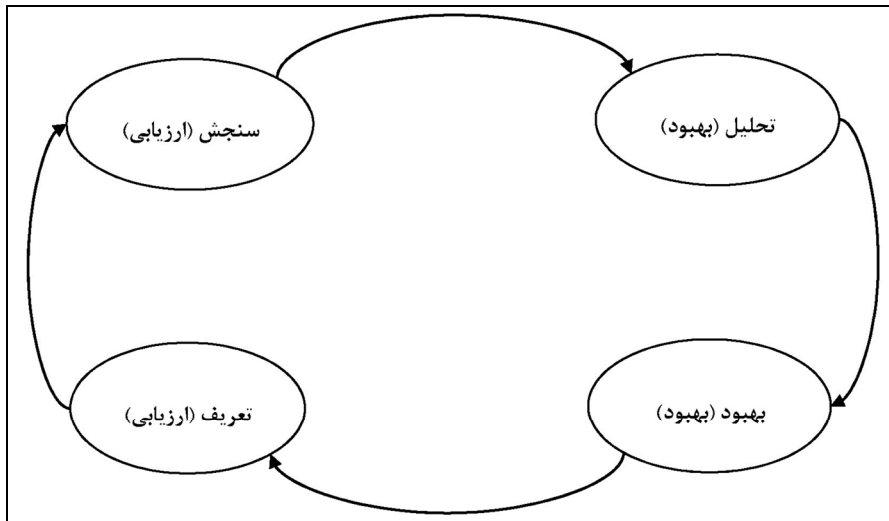
در این بخش روش‌های مختلف و چارچوب‌های متفاوت مدیریت کیفیت داده ارائه می‌شود. در این راستا توصیفی مختصر از هر روش ارائه می‌شود. اطلاعات ارائه شده ساختار مشترکی دارند که مشتمل بر سه بخش هستند:

۱. گام‌های هر روش و وابستگی متقابل آن‌ها و تصمیم‌های مهم که با نمودار و توصیف متنی بیان شده است و شامل ورودی‌ها، خروجی‌ها و گام‌ها است؛
۲. توصیف کلی و تأکید روی تمرکز هر روش و مشارکت آن در ارزیابی کیفیت داده‌ها و فرایند بهبود؛
۳. توضیحات تفصیلی که در مورد میزان کاربردی بودن هر روش.

۳-۲. روش TDQM (مدیریت جامع کیفیت داده)

روش مدیریت کیفیت داده اولین روش عمومی بود که در ادبیات موضوعی کیفیت داده منتشر شد (وانگ ۱۹۹۸). این روش در نتایج به‌دست‌آمده از پژوهش‌های دانشگاهی بدست آمده است. هدف اصلی این روش گسترش کیفیت داده و اصول مدیریت جامع کیفیت است (اوآکلان^۱ ۱۹۸۹).

1. Oaklan



گام ۱: تعریف

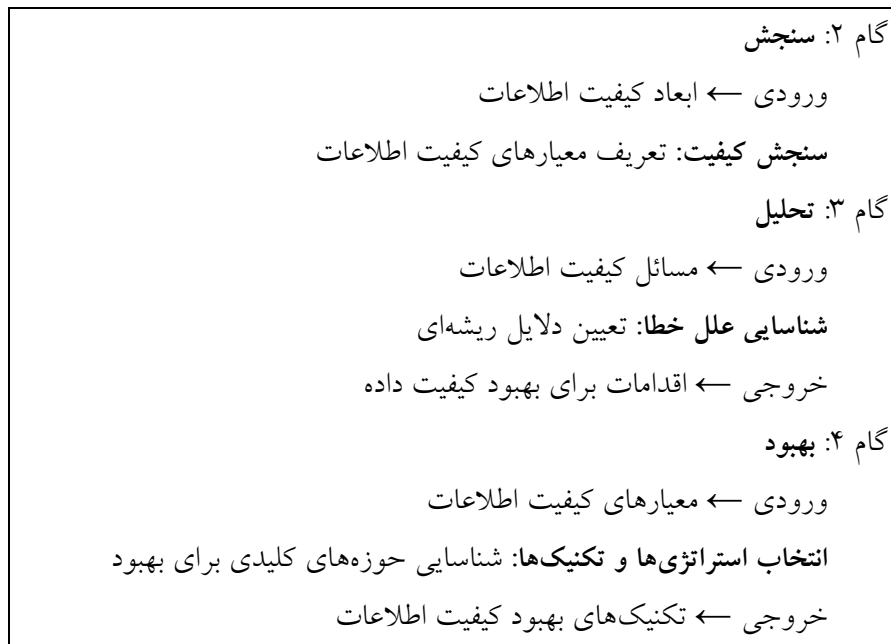
ورودی ← محصول اطلاعاتی

تحلیل داده: تعریف ویژگی‌ها در دو سطح:

- (۱) سطح بالا: توصیف عملکردها برای مصرف‌کنندگان اطلاعات
 - (۲) توصیف سطح پایین از واحدها و مؤلفه‌های پایه‌ای محصول اطلاعاتی و روابط آن‌ها
- تحلیل الزامات کیفیت داده: تعریف الزامات کیفیت اطلاعات از دیدگاه عرضه‌کنندگان پردازش اطلاعات، تولیدکنندگان و مصرف‌کنندگان مدل‌سازی فرآیند: تعریف سیستم تولید اطلاعات.
- خروجی ← طراحی فیزیکی و منطقی محصول اطلاعاتی با ویژگی‌های کیفی لازم؛

یک مدل کیفیت رابطه - موجودیتی که الزامات پردازش اطلاعات و کیفیت اطلاعات آن را تعریف می‌کنند؛

یک سیستم مدیریت اطلاعات که توصیف می‌کند چگونه پردازش اطلاعات تولید شده است.



شکل ۳-۱ گام‌های مدیریت کیفیت داده

روش مدیریت جامع کیفیت باعث جهت‌گیری تمرکز فعالیت‌های مهندسی مجدد از بازدهی به سمت اثربخشی است و این کار از طریق ارائه خطوط راهنمای روش با هدف حذف اختلاف‌ها و تفاوت‌های بین خروجی فرایندهای عملیاتی و الزامات مشتریان انجام می‌شود. باتوجه به این الزامات، مهندسی مجدد باید از مدل‌سازی فرایندهای عملیاتی شروع شود. همچنین روش مدیریت جامع کیفیت زبانی برای توصیف فرایندهای تولید اطلاعات یا پردازش اطلاعات که سازگار با اصول قبل نیز هستند، پیشنهاد می‌دهد که IP-MAP نامیده می‌شود (شانکاراناریان^۱ و دیگران ۲۰۰۰).

زبان IP-MAP به صورت متنوعی به سمت UML و برای پشتیبانی از طراحی سازمانی گسترش یافته است. IP-MAP تنها زبان برای مدل‌سازی فرایند اطلاعات بوده و نشان‌دهنده یک استاندارد عملی است. برای نمونه تجربیات عملی در ارتباط با روش مدیریت کیفیت داده در مطالعه کواک و ویکرت^۲ (۲۰۰۲) گزارش شده است.

1. Shankaranaryan

2. Kvac And Weickert

توضیحات تفصیلی: هدف روش مدیریت کیفیت داده پشتیبانی کامل از فرایند بهبود کیفیت (از تحلیل الزامات تا اجرا) است. همان‌گونه که در شکل ۳-۱ دیده می‌شود چرخه این روش شامل چهار گام تعریف، سنجش، تحلیل و بهبود است که یک فرایند پیوسته بهبود کیفیت را اعمال می‌کند.

نقش‌های مسئول برای گام‌های مختلف فرایند کیفیت همچنین در روش مدیریت کیفیت داده تعریف می‌شوند. چهار نقش متمایز وجود دارند:

۱. تأمین‌کنندگان اطلاعات که داده‌ها را برای پردازش اطلاعات ایجاد یا جمع‌آوری می‌کنند؛

۲. تولیدکنندگان اطلاعات که داده و زیرساخت سیستم‌های مربوطه را طراحی، توسعه و حفظ می‌کنند؛

۳. مصرف‌کنندگان اطلاعات که از داده‌ها در کارشان استفاده می‌کنند؛

۴. مدیران پردازش اطلاعات که مسئول مدیریت کل فرایند تولید اطلاعات در چرخه عمر اطلاعات هستند.

همچنین روش مدیریت کیفیت داده از دیدگاه اجرا هم جامع است زیرا خطوط راهنمایی در مورد چگونگی به‌کارگیری روش ارائه می‌دهد. در به‌کارگیری روش مدیریت کیفیت داده، یک سازمان باید:

۱. به‌روشنی پردازش اطلاعات را درک کند؛

۲. یک تیم پردازش اطلاعات ایجاد کند که شامل یک مدیر ارشد به‌عنوان

رهبر روش مدیریت کیفیت داده باشد، یک مهندسپردازش اطلاعات که با

روش مدیریت کیفیت داده آشنا است و اعضای که تأمین‌کننده،

تولیدکننده، مصرف‌کننده اطلاعات و مدیر پردازش اطلاعات هستند؛

۳. ارزیابی و مدیریت کیفیت اطلاعات را به همه زیرمجموعه‌های

تشکیل‌دهنده پردازش اطلاعات آموزش دهد؛

۴. بهبود مستمر پردازش اطلاعات را نهادینه کند.

روش مدیریت کیفیت داده، به ادبیات موضوعی کیفیت اطلاعات برای معیارها و شیوه‌های بهبود کیفیت اطلاعات متکی است و به صورت صریح به آن‌ها اشاره کرده (وانگ و استرانگ ۱۹۹۶) و برای مشخصه‌های ابعاد کیفیت روندی مشخص دارد. روش مدیریت کیفیت داده موضوعات کیفی را به شیوه‌های متناظر بهبود مرتبط می‌سازد. اما در مدل وانگ و استرانگ به هیچ شیوه خاصی در صنعت اشاره نشده است و هیچ گونه پشتیبانی برای تخصصی کردن شیوه‌های عمومی بهبود کیفیت ارائه نشده است.

۳-۳. روش DWQ (کیفیت انبار داده)

روش‌شناسی کیفیت انبار داده در پروژه کیفیت انبار داده‌های اروپا توسعه داده شده است (جسفیلد و دیگران ۱۹۹۸) و رابطه بین اهداف کیفیت و گزینه‌های طراحی را در انبار داده مطالعه می‌کند. این روش ذهنی بودن مفهوم کیفیت را در نظر گرفته، یک دسته‌بندی از اهداف کیفیت بر طبق گروه ذینفعان ارائه داده و بر آن‌ها متمرکز می‌شود. سپس تنوع اهداف کیفیت را در نظر می‌گیرد و فراداده‌های متناظر را تعریف می‌کند.

در روش‌شناسی کیفیت انبار داده برای طراحی فراداده‌ها در انبار داده باید سه دیدگاه لحاظ شود:

۱. دیدگاه مفهومی کسب‌وکار که روی مدل سازمان تمرکز دارد؛

۲. دیدگاه منطقی که روی طرح انبار داده تمرکز دارد؛

۳. دیدگاه فیزیکی که بیانگر لایه انتقال داده فیزیکی است.

این سه دیدگاه متناظر با سه لایه سستی انبار داده به نام‌های منابع، انبار داده و مشتریان هستند. این روش از طریق یک دیدگاه فراداده‌ای متناظر به نام سنجش کیفیت، با هر دیدگاه ارتباط پیدا می‌کند.

از دیدگاه کیفیت داده، چهار گام اصلی مشخصه‌های این روش هستند: تعریف، ارزیابی، تحلیل و بهبود (شکل ۳-۲). در نظرگیری طبقه‌بندی داده‌ها و ابعاد کیفیت

نرم‌افزار در بافت انبار داده‌ها، یکی از بخش‌های عمده این روش است. سه دسته از داده‌ها و فراداده‌ها به صورت زیر تعریف می‌شوند:

- کیفیت طراحی و مدیریت: اولی اشاره به توانایی یک مدل در ارائه کافی و مؤثر اطلاعات دارد، درحالی‌که دومی به روشی اشاره دارد که مدلی را در طول عملیات انبار داده تکمیل می‌کند؛
- کیفیت اجرای نرم‌افزار: در این روش ابعاد کیفیت استاندارد ایزو ۹۱۲۶ در نظر گرفته می‌شوند، زیرا اجرای نرم‌افزار یک وظیفه برای ایجاد انطباق با مشخصه‌های خاص انبار داده نیست؛
- کیفیت کاربرد داده: که اشاره به ابعادی دارد که نحوه استفاده و جستجوی داده در یک انبار داده را نشان می‌دهند.

روش‌های سنجش مناسب برای هر بعد شناسایی می‌شوند. فهرست این روش‌ها همواره با درجه ارتباطی که با هر ذینفع ورودی مرحله سنجش دارند تحت تأثیر قرار می‌گیرند. در گام ارزیابی کیفیت، مجموعه‌ای از اطلاعات در مورد هر بعد کیفیت داده وجود دارد که به شرح زیر است:

۱. الزامات کیفیت، محاسبه فاصله از مقادیر مورد انتظار؛
۲. سنجش کیفیت به دست آمده؛
۳. معیار استفاده شده برای محاسبه اندازه‌گیری؛
۴. وابستگی‌های علی معلولی به دیگر ابعاد کیفیت.

اطلاعات در مورد وابستگی‌های بین ابعاد کیفیت برای ردگیری و تحلیل مسائل کیفیت استفاده می‌شوند. شناسایی حوزه‌های اصلی، مرحله آخر تحلیل روش است. درواقع این مورد تنها بخشی است که با گام بهبود همپوشانی دارد، اما اطلاعات کاملی در مورد نحوه ایجاد بهبود در پایگاه داده فراهم نمی‌کند.



شکل ۳-۲ گام‌های روش‌شناسی کیفیت انبار داده

۳-۴. روش TIQM (مدیریت کیفیت جامع اطلاعات)

روش مدیریت جامع کیفیت اطلاعات برای پشتیبانی از پروژه‌های انبار داده طراحی شده است. این روش فرض یکپارچه کردن منابع داده عملیاتی به شکل یک پایگاه داده واحد و یکپارچه را داشته و در انواع یکپارچه‌سازی‌های انجام‌گرفته برای ایجاد انبار داده استفاده می‌شود. این یکپارچه‌سازی‌ها باعث حذف خطاها و ناهمگونی‌های پایگاه داده‌های منبع می‌شود. روش مدیریت کیفیت جامع اطلاعات روی فعالیت‌های مدیریتی تمرکز دارد. این فعالیت‌ها مسئول یکپارچه‌سازی منابع داده عملیاتی از طریق بحث در مورد استراتژی‌های سازمان برای انتخاب‌های مؤثر فنی هستند. برای مثال، در این روش تحلیل‌های هزینه-فایده از یک دید مدیریتی پشتیبانی می‌شوند. این روش یک دسته‌بندی تفصیلی از هزینه‌ها و منافع ارائه می‌دهد. شکل ۳-۳ گام‌های روش مدیریت کیفیت جامع اطلاعات را نشان می‌دهد. اگر از دیدگاه مدیریتی به روش مدیریت کیفیت جامع اطلاعات بنگریم سه گام اصلی وجود دارد:

- ارزیابی؛
- بهبود و مدیریت بهبود؛
- پایش.

یکی از کمک‌های ارزشمند این روش تعریف گام آخر است که خطوط راهنمایی برای مدیریت تغییر را در ساختار سازمان مطابق الزامات مدیریت کیفیت داده فراهم می‌کند. به‌علاوه از رویکرد اقتصادی، ارزیابی هزینه-فایده را برای توجیه مداخلات در کیفیت داده ارائه می‌دهند. هدف این روش دستیابی به سطح بالاتری از کیفیت داده نیست، بلکه هدف اصلی به عهده گرفتن اقدامات بهبود در صورت امکان‌پذیر بودن آن‌هاست. بنابراین اقدامات مذکور فقط هنگامی که منافع از هزینه‌ها بیشتر شود توجیه دارد.



شکل ۳-۳ گام‌های مدیریت کیفیت جامع اطلاعات

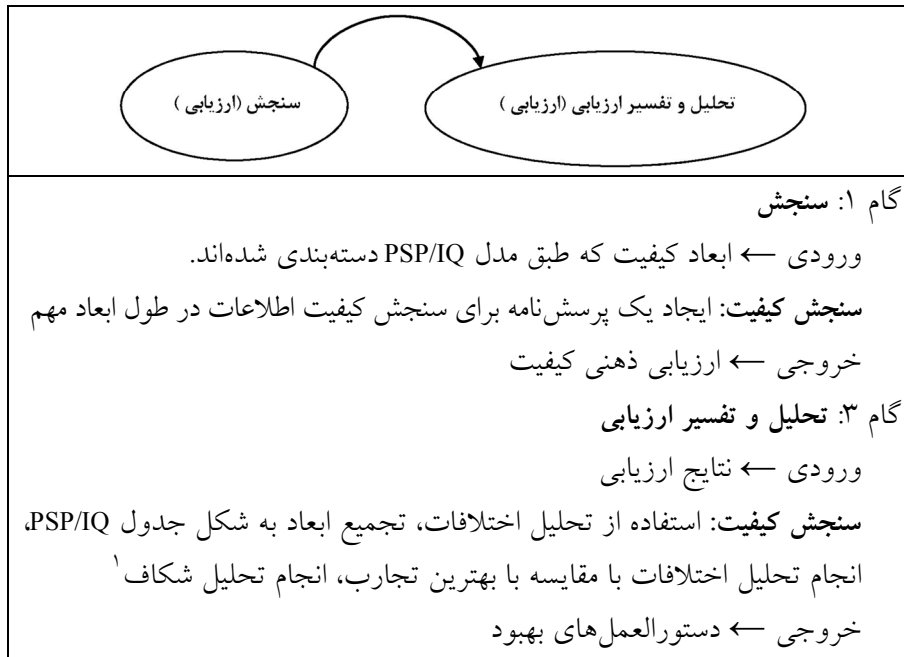
۳-۵. روش AIMQ (روشی برای ارزیابی کیفیت اطلاعات)

روش ارزیابی کیفیت اطلاعات تنها روش کیفیت اطلاعاتی است که بر روی مقایسه‌ها در راستای انتخاب بهترین‌ها تمرکز دارد (لی و دیگران ۲۰۰۲). این روش یک شیوه مستقل برای ارزیابی کیفیت است. پایه روش ارزیابی کیفیت اطلاعات یک جدول دو در دو به نام مدل PSP/IQ است (شکل ۳-۴) که ابعاد کیفیت را برپایه اهمیت‌شان از دیدگاه کاربران و مدیران دسته‌بندی می‌کند. محورهای جدول «همخوانی با مشخصه‌ها» و «همخوانی با انتظارات کاربران» هستند. در ادامه روش، چهار دسته ابعاد متمایز شناسایی می‌شوند (صحیح و دقیق، وابسته، مفید و قابل استفاده). سپس ابعاد کیفیت مطابق مقاله وانگ و استرانگ (۱۹۹۶) شناسایی شده و در راستای این دسته‌ها طبقه‌بندی می‌شوند. مقایسه‌های انجام گرفته در عمل باید اطلاعات را درون هر طبقه رتبه‌بندی کند. مدل PSP/IQ یک ورودی به روش ارزیابی کیفیت اطلاعات است که گام‌های آن به‌طور خلاصه در شکل ۳-۵ آمده است. مطالب منتشرشده در توصیف روش ارزیابی کیفیت اطلاعات عمدتاً تمرکزشان بر فعالیت‌های ارزیابی است ولی در خطوط راهنما، شیوه‌ها و ابزارهای بهبود ارائه نشده‌اند.

کیفیت اطلاعات عمدتاً با استفاده از پرسش‌نامه‌ها ارزیابی می‌شود. یک پرسشنامه اولیه آزمایشی برای شناسایی ابعاد کیفیت مرتبط و ویژگی‌های مقایسه شده به کار می‌رود. سپس یک پرسشنامه دوم از قبل شناسایی شده را به‌منظور بررسی معیارهای کیفیت اطلاعات موردتوجه قرار می‌دهند. درنهایت، این معیارها با شاخص‌ها مقایسه می‌شوند. در مقاله لی و دیگران (۲۰۰۲) فهرستی از ابعاد کیفیت استاندارد و ویژگی‌های پرسش‌نامه‌ها کمک می‌کند.

	از انتظار مشتریان فراتر می‌رود یا آن را برآورده می‌کند	همخوانی با مشخصات
اطلاعات مفید	اطلاعات صحیح و دقیق	کیفیت محصول
اطلاعات قابل استفاده	اطلاعات وابسته	کیفیت خدمت

شکل ۳-۴ مدل PSP/IQ



شکل ۳-۵ گام‌های ارزیابی کیفیت اطلاعات

۳-۶. روش CIHI (موسسه کانادایی اطلاعات سلامت)

روش موسسه کانادایی برای اطلاعات سلامت، روشی است که موسسه‌ای کانادایی برای ارزیابی و بهبود کیفیت اطلاعات سلامت به کار گرفته است (لینگ و سکو ۲۰۰۵). در این روش شکل اصلی اندازه پایگاه‌های داده و ناهمگونی آن‌هاست. روش موسسه کانادایی برای روش اطلاعات سلامت از انتخاب یک زیرمجموعه از داده‌ها برای تمرکز روی گام ارزیابی کیفیت پشتیبانی می‌کند. همچنین مجموعه وسیعی از معیارهای کیفیت را برای ارزیابی ناهمگونی‌ها پیشنهاد می‌دهد. استراتژی کیفیت داده‌ها در روش موسسه کانادایی برای اطلاعات سلامت یک رویکرد دوحله‌ای پیشنهاد می‌کند (شکل ۳-۶). گام نخست آن تعریف چارچوب کیفیت داده است و گام دوم آن یک تحلیل عمیق از داده‌هایی است که بیشترین دسترسی را دارند.

1. Gap Analysis

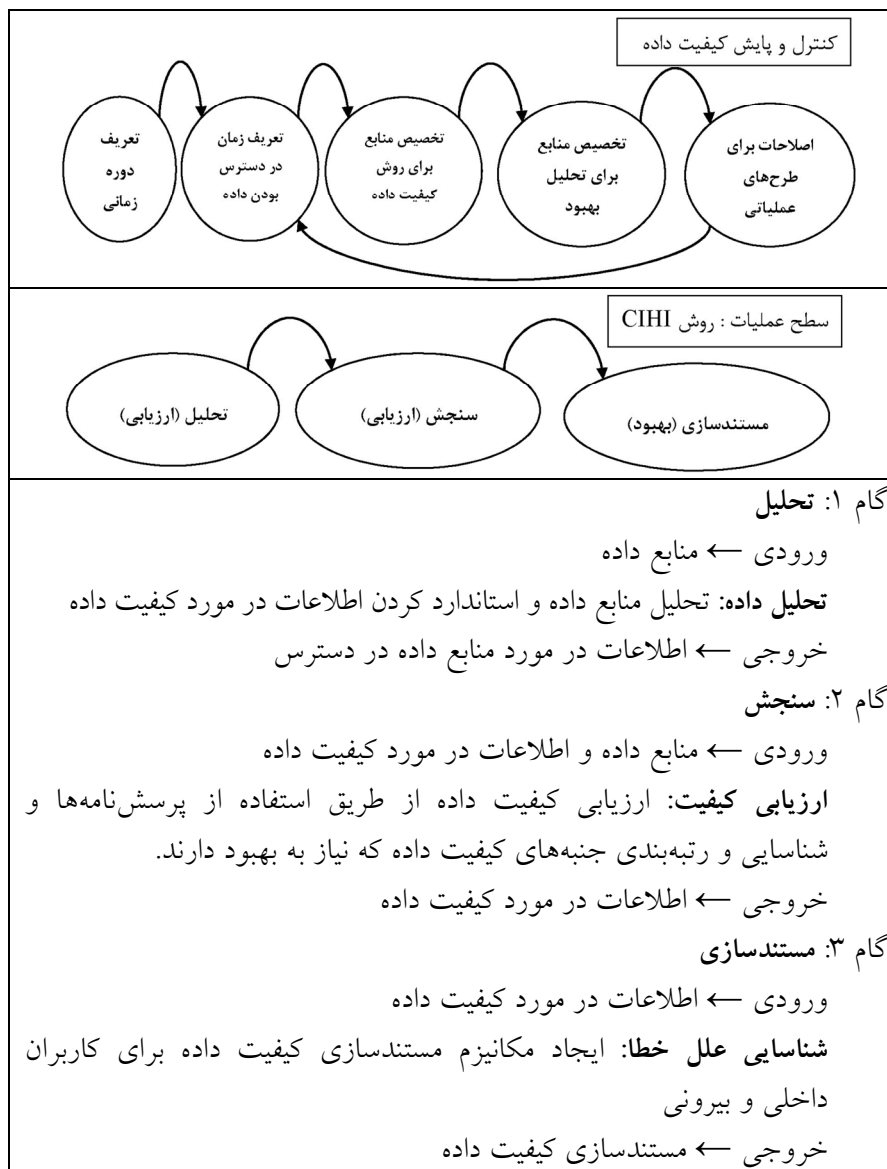
چارچوب کیفیت داده در سه مرحله تعریف می‌شود:

۱. استانداردسازی کیفیت اطلاعات داده؛
 ۲. توسعه یک استراتژی مشترک برای ارزیابی کیفیت داده؛
 ۳. تعریف یک فرایند کاری برای مدیریت داده‌های روش موسسه کانادایی برای اطلاعات سلامت که اولویت‌های کیفیت داده را شناسایی و رویه‌های بهبود مستمر داده را اجرا می‌کند.
- اجرای چارچوب موسسه کانادایی برای اطلاعات سلامت به صورت چرخه‌ای است و طبق رویکرد بهبود مستمر است. موارد زیر برای اجرای موفق لازم هستند:
- تعریف دوره زمانی برای چرخه بهبود
 - تعریف اهداف زمانی برای کیفیت‌های متفاوت مدنظر
 - تخصیص منابع خاص و ویژه برای تحلیل کیفیت داده، ارزیابی و مستندسازی
 - تخصیص منابع ویژه برای بهبود کیفیت داده
- تحلیل داده‌هایی با استفاده از موارد ذکر شده قبلی در سه مرحله انجام می‌شود:
- تحلیل کیفیت داده؛
 - ارزیابی؛
 - مستندسازی (در آن مستندات مسائل کیفی که با ارزیابی و تحلیل شناسایی شده‌اند را گزارش می‌دهند).

ارزیابی کیفیت داده‌ها براساس یک مدل سلسله مراتبی چهار سطحی صورت می‌گیرد. در سطح اول ۸۶ معیار اولیه کیفیت تعریف می‌شوند. این معیارها توسط الگوریتم‌های ترکیبی به ۲۴ مشخصه کیفیت در سطح بعدی تبدیل می‌شوند و بعداً دوباره به ۵ بعد درمی‌آیند (در سطح سوم) که شامل دقت، به موقع بودن، قابل مقایسه بودن، قابل استفاده و مرتبط بودن است. در نهایت این پنج بعد به شکل یک ارزیابی پایگاه داده در سطح چهارم درمی‌آیند.

ارزیابی اولیه ۸۶ معیار کیفیت داده، توسط پرسش‌نامه‌هایی که معیارهای آن با امتیاز ده به چهار حالت، غیرقابل کاربرد^۱، نامعلوم^۲، برآورد نشده^۳ و برآورد شده^۴

صورت می‌گیرد. سپس امتیاز در هر سطح جمع شده و ارزیابی اعتبار سنجی صورت می‌گیرد. این فرایند اعتبار سنجی اطمینان می‌دهد که تفسیر و امتیازدهی به هر معیار تا حد ممکن استاندارد است.



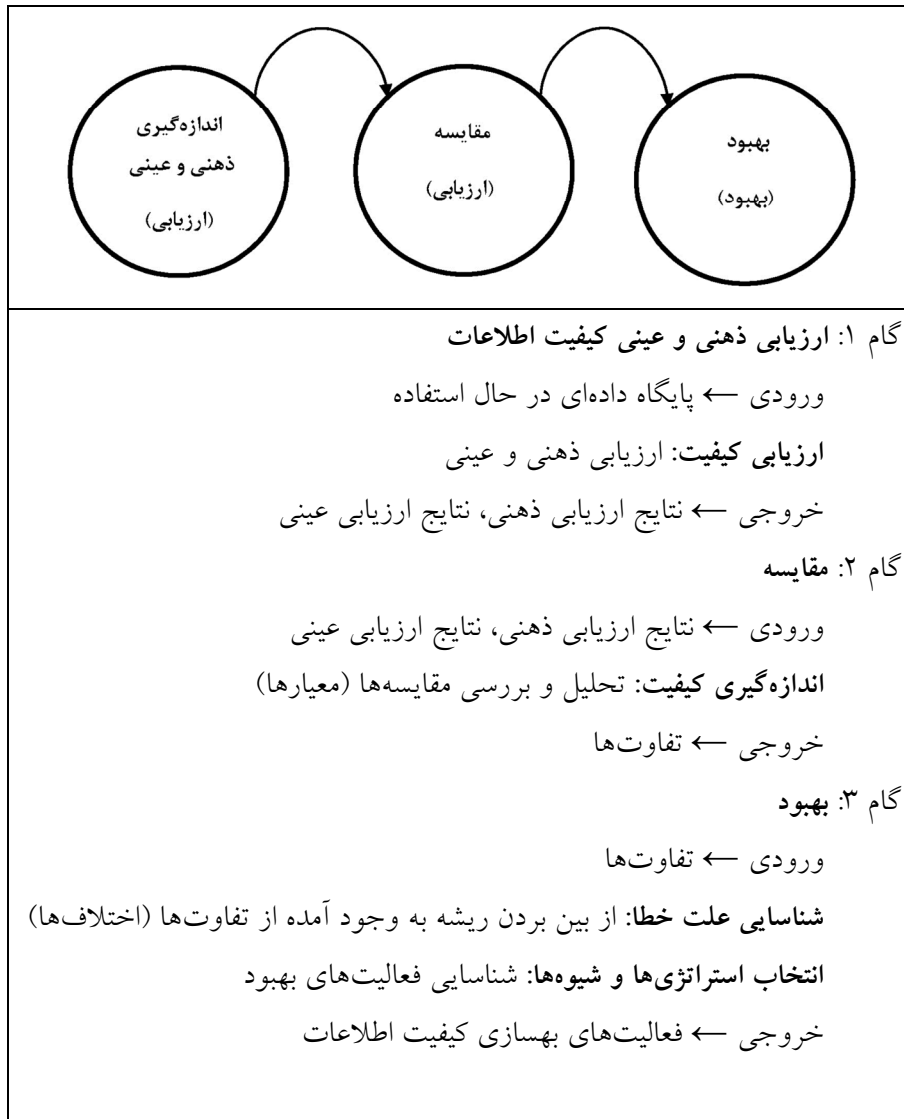
شکل ۳-۶ گام‌های روش موسسه کانادایی برای اطلاعات سلامت

۷-۳. روش DQA (ارزیابی کیفیت داده‌ها)

روش ارزیابی کیفیت داده (پپینو و دیگران ۲۰۰۲) برای ارائه اصول کلی راهنمای تعریف معیارهای کیفیت داده، طراحی شده است. در ادبیات موضوعی، معیارهای کیفیت داده عمدتاً به صورت ویژه تعریف می‌شوند تا مسائل خاصی را حل کنند و از این رو وابسته به سناریوهای در نظر گرفته شده هستند. هدف روش ارزیابی کیفیت داده شناسایی اصول سنجش عمومی کیفیت است که در پژوهش‌های پیشین مشترک باشند.

دسته‌بندی معیارهای روش ارزیابی کیفیت داده در شکل ۷-۳ خلاصه شده است. معیارهای ذهنی و معیارهای عینی شامل دسته‌بندی وابسته به وظیفه^۱ و مستقل از وظیفه^۲ هستند. اولین دسته، کیفیت داده را بدون دانش زمینه‌ای کاربردی ارزیابی می‌کند. درحالی‌که دسته دوم برای زمینه‌های کاربردی خاص تعریف می‌شود و شامل قواعد کسب‌وکار، مقررات شرکت‌ها و دولت‌ها و محدودیت‌های ارائه شده توسط مدیریت پایگاه داده است. هر دو دسته شاخص‌هایی را در چارچوب کلی زیر اندازه‌گیری می‌کنند:

- نسبت ساده
- مقادیر حداقل یا حداکثر
- متوسط وزنی



شکل ۳-۷ گام‌های ارزیابی کیفیت داده

۳-۸. روش IQM (اندازه‌گیری کیفیت اطلاعات)

هدف از روش اندازه‌گیری کیفیت اطلاعات ارائه یک چارچوب جهت کیفیت اطلاعات استفاده‌شده در داده‌های وب می‌باشد. به‌طور خاص، روش اندازه‌گیری

کیفیت اطلاعات به طراحان و مدیران وبسایت‌ها کمک می‌کند تا ابزارها و صفحات مبتنی بر وب را براساس کیفیت اطلاعات، مدیریت و نگهداری کنند. توضیحات تفصیلی: روش اندازه‌گیری کیفیت اطلاعات، دستورالعملی برای حصول اطمینان از نرم‌افزارهایی که تمامی ابعاد اساسی کیفیت اطلاعات را بررسی می‌کنند، فراهم می‌کند.

این روش مجموعه‌ای از دستورالعمل‌ها ارائه می‌دهد؛ چارچوب کیفیت اطلاعات و شاخص کیفیت را تعریف کرده و سپس دستورالعمل و چگونگی اندازه‌گیری کیفیت را شرح می‌دهد. گام‌های اصلی روش اندازه‌گیری کیفیت اطلاعات در شکل ۳-۸ توضیح داده شده‌اند. گام اول روش اندازه‌گیری معیارها را تعریف می‌کند. سپس چارچوب کیفیت اطلاعات به صورت یک فهرست از شاخص‌های مربوط به کیفیت اطلاعات شناخته شده می‌شود. محاسبه و تعریف چهارچوب، ورودی محاسبه کیفیت اطلاعات می‌باشد و شاخص کیفیت اطلاعات با روش‌ها و ابزارهای فرایند اندازه‌گیری تکمیل می‌شوند. بعضی شاخص‌ها نیازمند چندین روش اندازه‌گیری بوده و این روش استفاده از روش‌های اندازه‌گیری متعدد را هماهنگ می‌نماید.



شکل ۳-۸ گام‌های اندازه‌گیری کیفیت اطلاعات

۳-۹. روش ISTAT (اداره ملی سرشماری ایتالیا)

روش ISTAT (ایستات^۱ ۲۰۰۰؛ فالورسی و دیگران ۲۰۰۳) در اداره ملی سرشماری ایتالیا جهت به دست آوردن اطلاعات دقیق و باکیفیت بالا از شهروندان و مشاغل طراحی گردید. مسئله اساسی درباره این روش، چگونگی تضمین به دست آوردن کیفیت داده‌ها از پایگاه داده‌های متعدد عمومی و محلی است. این مسئله یک چالش خاص در مطالعات و پژوهش‌های ایتالیایی است، چراکه مدیریت عمومی به سه ناحیه جغرافیائی تقسیم شده (مرکزی، منطقه‌ای و پیرامونی) و هر مدیریت دارای داده‌های مستقل می‌باشد. روش ISTAT بر رایج‌ترین نوع تبادل اطلاعات در میان سطوح مختلف مدیریت عمومی (داده‌های خاص) تمرکز دارد. این روش به شدت بر هنجارها متمرکز بوده و هدف آن مدیریت داده‌ها می‌باشد و تا داده‌ها بتوانند کیفیت موردنیاز را تأمین نمایند.

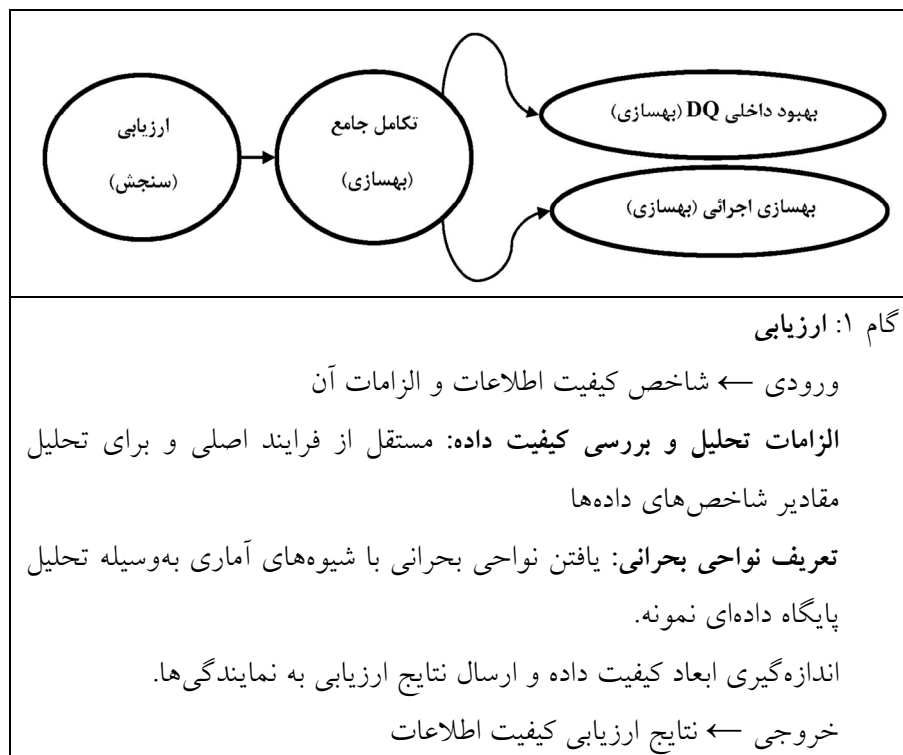
مراحل اساسی این روش (شکل ۳-۹) عبارت‌اند از:

- گام ارزیابی که در ابتدا در پایگاه داده مرکزی ISTAT انجام می‌شود. هدف آن تشخیص مشکلات مربوط به کیفیت از دیدگاه یکپارچه‌سازی داده‌ها می‌باشد؛
- گام بهسازی کلی که مسئولیت انجام ارتباط رکوردها در پایگاه داده‌های ملی و طرح راه‌حل بهسازی در فرایندهای ساخت، خرید یا انطباق راه‌حل‌های موجود را بر عهده دارد؛
- فعالیت‌های بهبود در پایگاه داده تحت مالکیت و مدیریت دولت‌های محلی. این فعالیت‌ها باید به وسیله خود دولت‌های محلی با ابزارهای ارائه شده توسط ISTAT انجام گردد؛
- فعالیت‌های بهسازی که نیازمند همکاری بخش‌های مختلف است. این فعالیت‌ها معمولاً فرایندهای جهت‌دار می‌باشند، چراکه داده‌ها را در طول

¹ Istat

اجرای فعالیت‌های عملیاتی خاص مبادله می‌کنند. در این گام پایگاه داده مرکزی ممکن است جهت اهداف هماهنگی موردنیاز باشد. این فعالیت‌ها به‌صورت مرکزی برنامه‌ریزی و هماهنگ‌سازی می‌شوند.

روش ISTAT شیوه‌ای ساده ولی تأثیرگذار را در اندازه‌گیری کیفیت ارائه می‌دهد. همچنین ابزاری برای فعالیت‌های پاک‌سازی داده‌ها در نظر می‌گیرد. در روش ISTAT، مالکین داده‌ها به‌وسیله جزئیاتی مانند کد شهروندی تعریف می‌گردند. این روش از استانداردسازی فرمت داده‌ها پشتیبانی می‌کند و بیان آن‌ها با طرح XML این اجازه را می‌دهد که پایگاه‌های داده‌ای محلی باهم ادغام گردند. مبادله اطلاعات بین شرکت‌های محلی به‌وسیله نرم‌افزارها از نو طراحی گشته و براساس انتشار و اشتراک عمل می‌کنند.



گام ۲: بهسازی جامع

ورودی ← نتایج ارزیابی کیفیت اطلاعات

شناسایی عوامل خطاها

طراحی راه‌حل بهسازی داده‌ها: ایجاد پیوند بین موارد ثبت‌شده مربوط به پایگاه داده ملی.

واگذاری مسئولیت‌های مربوط به داده: ایجاد یک مالک داده برای رشته‌های خاص کنترل و طراحی مجدد فرایندها: راه‌حل بهبود طراحی در فرایندها، نتایج ارزیابی جامع به مداخلات خاص در فرایندها

انتخاب سیاست‌ها و شیوه‌ها: ساخت و یا خرید، انطباق، طراحی ابزارهایی برای شیوه‌های مرتبط

مدیریت بهسازی: فعالیت‌های کیفیت داده برای ارائه به سازمان‌ها

خروجی ← روش بهسازی

گام ۳: بهسازی داخلی کیفیت داده (مستقلاً برای هر سازمان)

ورودی ← دستورات بهسازی از بهسازی جامع

طراحی مجدد فرایند، استانداردسازی فرمت کسب و استانداردسازی فرمت ارزی داخلی با استفاده از XML

نظارت بر بهسازی: انجام ارزیابی محلی، طراحی راه‌حل بهسازی داده‌ها و فرایندها در محدوده بحرانی

خروجی ← راه‌حل بهسازی

گام ۴: بهسازی اجرایی

ورودی ← راهکارهای ایجاد بهبود بر پایه بهبودهای عمومی

طراحی مجدد فرایند: استانداردسازی فرمت جریان اداری به‌وسیله XML، طراحی مجدد جریان ارزی با استفاده از رویداد عمومی و مشترک معماری محور

خروجی ← راهکارهای بهبود نهایی شده

۳-۱۰. روش AMEQ (اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول)

هدف اصلی روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول، یک ارزیابی دقیق برای کیفیت اطلاعات محصول در بخش ارزیابی و بهسازی^۱ بوسیله اهداف منسجم و تعیین‌شده است (سو و جین ۲۰۰۴).

این روش مخصوص ارزیابی کیفیت داده‌ها در شرکت‌های تولیدی و سازمان‌هایی است که اطلاعات محصول جزء اصلی از پایگاه‌های عملیاتی داده می‌باشد. در شرکت‌های تولیدی رابطه بین اطلاعات محصول و فرایندهای تولید به صورت یک استاندارد ساده و جامع در سراسر شرکت‌ها مطرح می‌باشد. همچنین پایگاه داده محصول، معمولاً در شرکت‌های مختلف مشابه است. این روش یک دستورالعمل شناختی برای مدل‌های اطلاعات و فرایندهای تولید مرتبط ارائه می‌کند.

روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول شامل پنج گام برای محاسبه و بهسازی کیفیت اطلاعات محصول می‌باشد (شکل ۳-۱۰). گام اول ارزیابی آمادگی فرهنگی یک سازمان است که با استفاده از شبکه مدیریت کیفیت اطلاعات پیاده‌سازی شده و نقش مهمی در مدیریت دارد. در این گام ابعاد کیفیت اطلاعات محصول براساس فعالیت‌های مختلف کسب‌وکار تعریف و طبقه‌بندی می‌شوند. مرحله دوم، اطلاعات مربوط به محصول را مشخص می‌سازد. محصول اطلاعات با یک فرایند کسب‌وکار مرتبط است و به وسیله شیوه رویکرد شیء‌گرا^۲ مدل‌سازی می‌شود.

در روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول هشت نوع موضوع مدل می‌شوند که عبارت‌اند از:

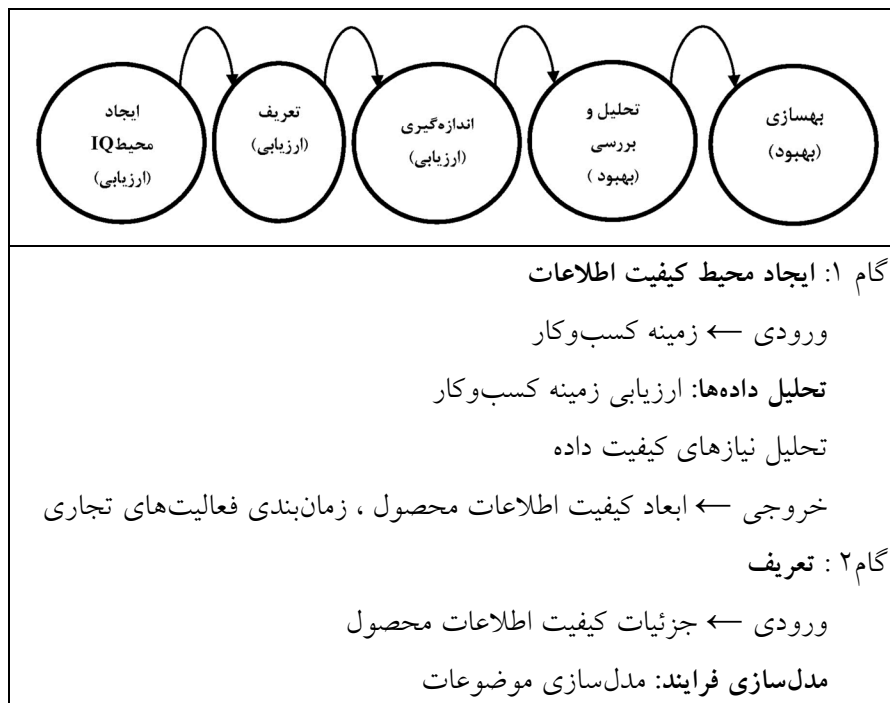
- منابع انسانی
- منابع اطلاعاتی
- فعالیت‌های شرکت

1. Product Information Quality, PIQ

2. Object-Oriented Approach, OOA

- ورودی منابع
- منابع فرایندها
- خروجی منابع
- اندازه‌گیری عملکرد
- اهداف شرکت.

در مرحله دوم، یک مدل از روش‌های اندازه‌گیری نیز تولید می‌گردد. مرحله سوم بر روی فعالیت‌های اندازه‌گیری تمرکز دارد. در مرحله چهارم تیم کیفیت اطلاعات باید به‌وسیله تحلیل و بررسی ابعاد کیفیت که نمره پایین‌تری کسب کرده‌اند به بررسی علل مشکلات ریشه‌ای و بالقوه کیفیت اطلاعات محصول بپردازند. در نهایت، گام بهبود کیفیت اطلاعات محصول می‌تواند آغاز شود. برای مراحل چهارم و پنجم، روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول ابزارهای دقیقی ارائه نمی‌کند و تنها دستورالعمل کلی ارائه می‌دهد.



مشخص نمودن محیط‌های بحرانی: مشخص کردن ابعاد کیفیت اطلاعات محصول
برای فعالیت ارزش سنجی
خروجی ← نتایج کیفیت اطلاعات محصول
گام ۳: اندازه‌گیری
ورودی ← نتایج و ابعاد کیفیت اطلاعات محصول
اندازه‌گیری کیفیت: مشخص کردن بازده اندازه‌گیری کیفیت اطلاعات محصول
برای هر فعالیت، انتخاب یک پروژه مدیریت‌پذیر کوچک، تعریف معیارها و
سؤالات درباره کیفیت اطلاعات محصول و فرایند اندازه‌گیری
خروجی ← نتایج اندازه‌گیری
گام ۴: تحلیل و بررسی
ورودی ← نتایج اندازه‌گیری
بررسی عوامل بروز اشکال: بررسی ابعاد کیفیت پائین داده‌ها، تحلیل و بررسی
مبانی و محاسبه بازده تأثیرگذار باتوجه به اهداف مالی
خروجی ← علل ریشه‌ای برای ابعاد کیفیت ضعیف
گام ۵: بهسازی
ورودی ← علل ریشه‌ای برای ابعاد کیفیت نامناسب
تعیین مسئولیت هر فرایند داده: استفاده از مدیریت منابع اطلاعات^۱ برای بهبود
بخشی کیفیت اطلاعات محصول در فعالیت‌های تجاری.
طراحی فرایند و طراحی مجدد: کاربرد مدیریت خودسازمان‌دهی^۲ در مهندسی
مجدد جریان‌های اطلاعاتی.
کنترل بهسازی: بهسازی مدیریت اهداف سرمایه‌ای^۳ برای سازگارسازی مجدد
ارزش‌های سازمانی
خروجی ← بهبود بلندمدت کیفیت اطلاعات محصول

شکل ۳-۱۰ گام‌های اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول

1. Information Resource Management, IRM
3. Enterprise Objective Management, EOM

2. Self-Organization Management, SOM

۳-۱۱. روش COLDDQ (تأثیر کیفیت پائین داده بر هزینه‌ها)

هدف اصلی از روش ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها (لوشین ۲۰۰۴) ارائه کارت امتیاز کیفیت داده جهت حمایت از ارزیابی تأثیرات کیفیت پائین داده بر هزینه‌ها می‌باشد. مانند روش مدیریت کیفیت جامع اطلاعات، این روش نیز به طبقه‌بندی هزینه‌ها، منابع و شرح آن‌ها می‌پردازد. این روش مستقیماً از تمرکز بر هزینه‌های کیفیت پائین به دلیل استفاده از شیوه‌های بهبود مستمر حاصل می‌گردد.

هدف از این روش ارزیابی کمی میزان خسارتی است که به فرایندهای کسب‌وکار سازمان در صورت مواجه شدن با اطلاعات بد و نامناسب وارد می‌شود.

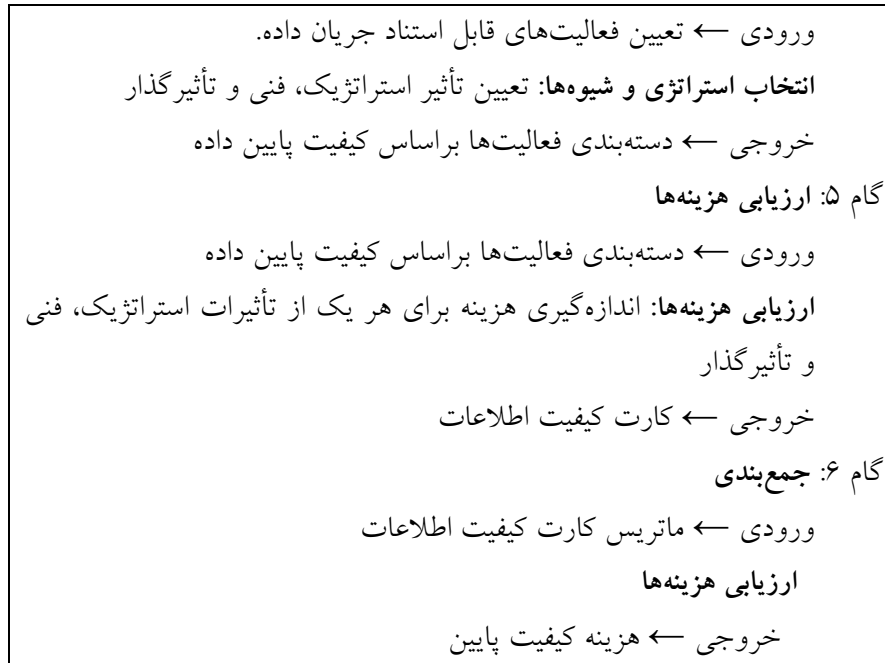
شش مرحله جهت ارزیابی هزینه‌های کیفیت پائین داده طراحی شده است (شکل ۳-۱۱). در مرحله اول از این روش، کسب‌وکار سازمان بوسیله دو مدل جریان داده‌ای مدل‌سازی می‌شود. جریان داده‌های استراتژیک که برای تصمیم‌گیری استفاده می‌شوند و جریان داده‌های عملیاتی که برای پردازش داده‌ها استفاده می‌شوند.

هر دو مدل نشان‌دهنده مجموعه‌ای از مراحل پردازش است که جریان اطلاعات را از منبع داده تا محل مصرف داده توصیف می‌نماید. تجزیه و تحلیل عینی و ذهنی فضای کسب‌وکار براساس این مدل‌ها انجام می‌شود. کاربران داخلی و خارجی، کارکنان و مشتریان به‌منظور شناسایی داده‌های ناقص بررسی می‌شوند. سپس خطاها به فعالیت‌های معیوب در مدل‌های استراتژیک و عملیاتی کسب‌وکار نسبت داده می‌شوند. این رابطه بین خطاها و فعالیت‌های پایه‌ای براساس ارزیابی هزینه ایجاد می‌شود. روش ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها یک طبقه‌بندی کامل و ارزشمند عملیاتی و تاکتیکی فراهم کرده و اثرات استراتژیک و اقتصادی را در نظر می‌گیرد. هر طبقه از هزینه‌ها ارزش اقتصادی خاصی دارند که براساس خبرگی تعیین می‌گردد. هزینه‌ها به‌عنوان ورودی در مرحله بهبود نهائی استفاده می‌گردند.

درنهایت این روش از تجزیه و تحلیل هزینه‌ها، ارزیابی و جمع‌بندی هزینه‌های پروژه‌های بهبود کیفیت پشتیبانی می‌کند. رهنمودهای این روش نیز از محاسبه بازگشت سرمایه^۱ و نقاط شکست بهبود پشتیبانی می‌کند.

1 Return On Investment, ROI





شکل ۳-۱۱ مراحل ارزیابی هزینه مؤثر برای کیفیت کم داده‌ها

۳-۱۲. روش DaQuinCIS (کیفیت داده در سیستم‌های اطلاعاتی تعاونی)

روش ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی (اسکاپانیکو و دیگران ۲۰۰۴) به کیفیت داده‌ها در سیستم‌های اطلاعاتی تعاونی مربوط است. این روش بر دو مسئله مربوط به کیفیت داده‌ها متمرکز می‌شود. اول از همه، کیفیت داده‌ها را براساس اعتماد سازمانی تعیین می‌کند. در این روش کیفیت پایین داده‌ها می‌تواند مانع هماهنگی شده و عواقب زیادی در پی داشته باشد. برای پرداختن به مسئله اول در این روش به معرفی مفهوم صدور گواهی‌نامه کیفیت داده می‌پردازیم که همراه داده‌ها در هنگام ردوبدل شدن بین سازمان‌ها استفاده می‌شود. مسئله دوم به مکانیسم انتخاب داده‌های مبتنی بر کیفیت مربوط می‌شود. ساز و کار این انتخاب‌ها براساس بالاترین کیفیت داده در بین داده‌های متعلق به سازمان‌های مختلف است. در این روش، هماهنگی یک اهرم برای بهبود کیفیت به شمار می‌رود.

این روش یک مدل ابتکاری جهت نمایش وضعیت و میزان کیفیت داده‌ها به نام داده و کیفیت داده‌ها^۱ ارائه می‌کند که شامل موارد زیر است:

۱. ساختاری برای نشان دادن داده
۲. مجموعه‌ای از ویژگی‌های کیفی داده
۳. ساختار ارائه‌دهنده ویژگی‌ها و کیفیت
۴. ارتباط بین داده‌ها و کیفیت فراداده‌ها

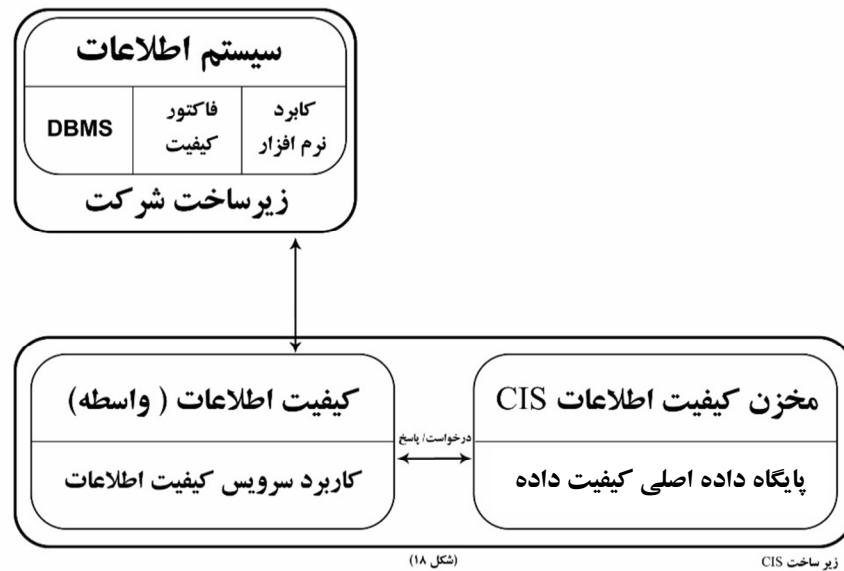
در این روش مرجعی قابل اعتماد شامل ویژگی‌های کیفیت ایجاد می‌شود. ارزش متناظر با دو بعد از کیفیت توسط یک سازمان شخص ثالث براساس پارامترهای مختلف از جمله شکایت‌های صورت گرفته توسط دیگر سازمان‌ها و درخواست‌های صادرشده از سایر سازمان‌ها تعیین می‌گردد. شکل ۳-۱۲ نشان‌دهنده مراحل اساسی از این روش به شرح زیر است:

- تجزیه و تحلیل کیفیت
- ارزیابی کیفیت
- صدور گواهینامه کیفیت
- بهبود کیفیت

این روش توسط یک سیستم شامل زیرساخت‌های داخلی و زیرساخت‌های خارجی پشتیبانی می‌گردد. مراحل این روش توسط ماژول‌های مربوط به شاخص کیفیت در هر سازمان متعلق به کشورهای مشترک‌المنافع اجرا می‌شود. ارتباط میان سازمان‌های کشورهای مشترک‌المنافع توسط خدمات اطلاع‌رسانی باکیفیت است که از یک موتور مشترک در انتشار به عنوان پیام‌رسان اصلی بین اجزای مختلف استفاده می‌گردد. در معیار کیفیت، درخواست کاربران خارجی کیفیت لحاظ می‌شود و در نهایت ماژول مناسب ارزیابی کیفیت داده‌ها و ارزش کیفی توسط کاربران تعیین می‌گردد.



شکل ۳-۱۲ گام‌های ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی



شکل ۳-۱۳ زیر ساخت‌های ارزیابی کیفیت در سیستم‌های اطلاعات تعاونی

اگر داده‌ها از نظر سطح کیفیت راضی‌کننده نباشند، یک هشدار به کاربر فرستاده می‌شود. در مقابل اگر کیفیت رضایت‌بخش باشد گواهی کیفیت مرتبط با داده برای کاربر ارسال می‌گردد. بهبود کیفیت توسط واسطه کیفیت^۱ اطلاعات انجام می‌گردد. این مازول به وسیله کالیبره شدن با مخزن اطلاعات^۲ و براساس طرح جامع، اطلاعات را تفسیر و داده‌های باکیفیت بالا را انتخاب می‌کند. هر سازمان می‌تواند درخواست مربوط به اطلاعات و داده‌های موردنیاز خود را به کلیه سازمان‌های ذی‌ربط ارسال و هم‌زمان میزان و سطح کیفیت مورد انتظار خود را تعیین نماید.

نسخه‌های مختلفی از داده‌های مشابه به‌عنوان پاسخ درخواست بررسی می‌شوند و در بین آن‌ها بهترین کیفیت انتخاب و به شرکت درخواست‌کننده ارائه می‌گردد. باتوجه به شیوه‌های مورد استفاده در گام بهسازی، روش DaQuinCIS یک الگوریتم جدید را برای تطبیق رکورد (اطلاعات ثبت‌شده) ارائه می‌دهد. تطبیق‌دهنده رکورد هم جزئی از واسطه (بررسی‌کننده) کیفیت اطلاعات است. مکانیزم تطبیق‌دهنده رکورد به‌منظور بررسی تطبیق کیفیت داده‌های صادرشده از سایر شرکت‌های همکار تعیین‌شده است.

1. Data Quality Broker 2. Data Quality Repository

۱۳-۳. روش QAFD (ارزیابی کیفیت داده‌های مالی)

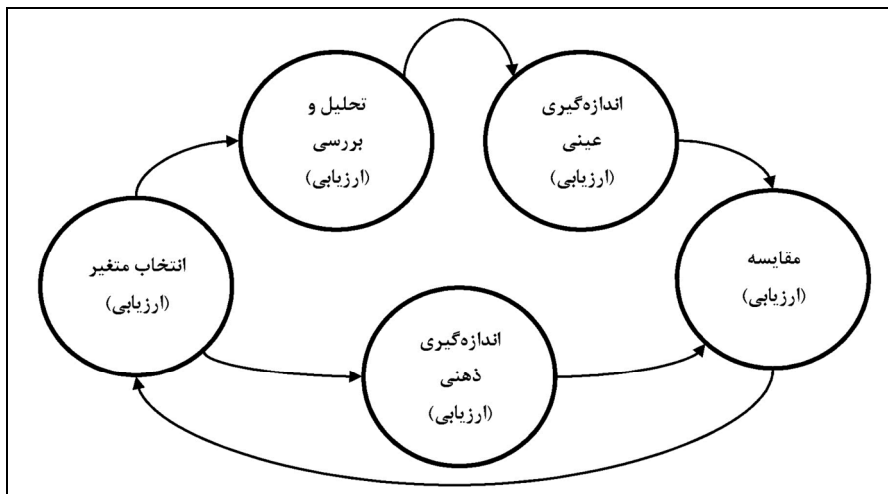
روش ارزیابی کیفیت داده‌های مالی (دی‌آمیسیس و باتینی ۲۰۰۴) برای تعریف استانداردهای کیفیت داده‌های عملیاتی مالی و به حداقل رساندن هزینه‌های ابزارهای سنجش کیفیت طراحی شده است. این روش ترکیبی از کیفیت و کمیت برای تعریف مسائل مربوط به کیفیت و انتخاب اقدامات مناسب جهت بهبود کیفیت می‌باشد. در این روش شاخص‌های مربوط به قوانین کیفیت داده‌ها، اندازه‌گیری و استراتژی‌هایی برای ارزیابی کمی و کیفی تعریف می‌شوند. به‌طورکلی این روش برای کیفیت اطلاعات مالی مناسب است.

مراحل اصلی این روش در شکل ۳-۱۴ ترسیم شده است. ابتدا این روش مناسب‌ترین متغیرهای مالی را انتخاب می‌کند که این انتخاب معمولاً براساس دانش به‌دست‌آمده از ارزیابی‌های قبلی باتوجه به میزان اثربخشی آن‌ها انجام می‌شود. متغیرها در دسته مسائل مرتبط که تحت تأثیر ریسک‌های سرمایه‌گذاران، مصرف‌کنندگان و عوامل دیگر فضای کسب‌وکار قرار دارند، گروه‌بندی می‌شوند. گام دوم این روش باهدف کشف علل اصلی خطاهاست. مناسب‌ترین ابعاد کیفیت داده‌ها در این مرحله تعیین شده و قوانین کیفیت داده به وجود می‌آیند. قوانین کیفیت داده‌ها نشان‌دهنده خواص دینامیک متغیرهاست که می‌تواند در ابعاد کیفی اندازه‌گیری شود. در مرحله سوم، ارزیابی عینی براساس شاخص‌های کمی انجام می‌شود. طراحان این روش یک مدل ریاضی برای ارزیابی عینی پیشنهاد کرده‌اند که در زمان رتبه‌بندی کلی داده‌ها با هر یک از ابعاد کیفیت ارائه می‌شود. ارزیابی ذهنی در مرحله چهارم از سه دیدگاه مختلف انجام می‌شود.

- کارشناسان کسب‌وکار
- مشتریان
- کارشناسان کیفیت داده

در این سطوح هر شخص برای ارزیابی سطح کیفیت، یک بعد کیفی را بررسی می‌نماید. ارزیابی کلی به‌دست‌آمده به‌عنوان میانگین ارزیابی ذهنی از هر طبقه کارشناسان

ثبت می‌گردد. در پایان، ارزیابی ذهنی و عینی در گام پنجم صورت می‌گیرد و برای هر یک از ابعاد، تفاوت ارزیابی ذهنی و عینی محاسبه می‌گردد. اگر تفاوت مثبت باشد، ارزیابی عینی باید براساس نقطه‌نظر کارشناسان کیفیت تجدیدنظر گردد.

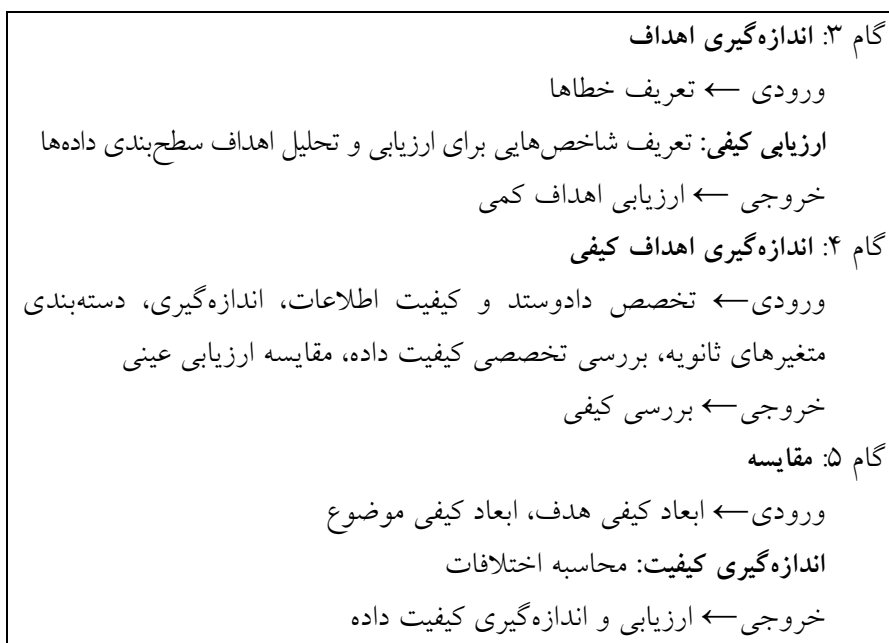


گام ۱: انتخاب متغیرها

ورودی ← متغیرهای مالی و متغیرهای زمینه‌ای
تحلیل و بررسی داده: تعریف و تشخیص بیشترین متغیرهای ثبت‌شده مالی، تقسیم متغیرها به سه گروه (کمیت، کیفیت، عددی و تاریخی).
خروجی ← گروه‌بندی متغیرهای ثانویه.

گام ۲: تحلیل و بررسی

ورودی ← گروه‌های متغیر ثانویه، شیوه‌های کیفیت داده، قوانین دادوستد، تحلیل و بررسی فرایند، مشاهده
تحلیل نیازهای کیفیت داده: تحلیل ابعاد کیفیت داده، درستی نحوی و معنایی، ثبات داخلی و خارجی، تکمیل، واحد پول، جدول‌های زمانی، یکتایی
تعریف محدوده بحرانی: کاربرد قوانین کیفیت داده
خروجی ← بررسی کمی



شکل ۳-۱۴ گام‌های ارزیابی کیفیت داده‌های مالی

۳-۱۴. روش CDQ (کیفیت داده‌های کامل)

روش جامع برای مدیریت کیفیت داده (باتینی و اسکناپیکو^۱ ۲۰۰۶؛ باتینی و دیگران ۲۰۰۸) یک روش جامع، کامل، انعطاف‌پذیر و به‌راحتی قابل پیاده‌سازی است. این روش با در نظر گرفتن شیوه‌ها و ابزارهای موجود و یکپارچه‌سازی آن‌ها در یک چارچوب می‌تواند در هر زمینه درونی و برونی سازمانی کار کند و همه نوع داده ساختاریافته، نیمه ساختاریافته و بدون ساختار را در برگیرد.

این روش از کاربران در انتخاب مناسب‌ترین شیوه‌ها و ابزارها در هر گام و در هر زمینه پشتیبانی کرده و بنابراین دارای انعطاف‌پذیری زیادی می‌باشد. درمجموع می‌توان به جمع‌بندی رسید که این روش به علت اینکه شیوه‌های اعمال‌شده را براساس اهداف مشخص‌شده در گام‌ها انتخاب می‌کند، کاربرد ساده‌ای دارد. این روش نوآورانه است چراکه انتخاب بهینه‌ای از فرایندها بهبود کیفیت را با در نظر گرفتن محدودیت‌های بودجه‌ای در شرایط مختلف فراهم می‌کند.

1. Batini And Scannapieco

این روش تا حد زیادی مستقل از زمینه کاری سازمان است. در واقع روش‌های دیگر با این فرض که اطلاعات به‌دست‌آمده در خصوص حوزه کاری سازمان قبلاً جمع‌آوری و مدل شده‌اند کار می‌کنند و تمرکز آن‌ها بر چگونگی رسیدن به کیفیت جامع داده‌ها بدون در نظر گرفتن حالت استفاده از دانش قبلی است. اما هدف اصلی روش جامع برای مدیریت کیفیت داده انجام یک ارزیابی کمی است که تعیین کند تا چه حد فرایندهای سازمان با اطلاعات بد تحت تأثیر قرار می‌گیرد.

سه مرحله اصلی این روش عبارتند از تعیین وضعیت موجود، ارزیابی و انتخاب فرایند بهبود بهینه (شکل ۳-۱۵). در مرحله اول از این روش، روابط بین واحدهای سازمانی، فرایندها، خدمات و داده‌ها تعیین می‌شوند.

این روابط با استفاده از ماتریس‌هایی از واحدهای سازمانی داده‌ها و نقش آن‌ها در فرایندهای کسب‌وکار مختلف تعریف می‌شود. همچنین در این مرحله فرایندها براساس سهمشان در تولید محصولات، قوانین حقوقی و سازمانی شرح داده می‌شوند. مرحله دوم یک هدف کیفی جدید را تعریف می‌کند که در فرایند بهبود کیفیت، ارزیابی هزینه‌ها و مزایا موردنیاز است. این مرحله به متغیرهای بحرانی تحت تأثیر کیفیت پایین می‌پردازد. از آنجاکه فعالیت‌های بهسازی پیچیده و هزینه‌بر هستند، بهتر است تمرکز روی بخشی از پایگاه داده‌ای و جریان داده باشد که باعث بروز دشواری فراوان می‌شود. درنهایت، گام سوم شامل پنج مرحله است و هدف آن شناسایی روند بهبود می‌باشد. در این مرحله پیامدهای فعالیت‌هایی با بالاترین اثربخشی شناسایی می‌شوند. هدف روش کیفی جدید با در نظر گرفتن هزینه‌ها و منافع تنظیم می‌گردد. فعالیت‌های بهسازی می‌توانند برای دستیابی به هدف کیفی جدید انجام گیرند. این روش شیوه‌های بهسازی داده محور و فرایند محور را برای پایگاه‌های تحت تأثیر کیفیت پایین پیشنهاد می‌کند.

درنهایت مجموعه‌ای از شیوه‌های بهبود سازگار به‌صورت یک فرایند بهسازی مناسب، تحلیل هزینه-فایده برتر را انتخاب می‌کند.



شکل ۳-۱۵ گام‌های روش جامع برای مدیریت کیفیت داده

۳-۱۵. خلاصه‌ای از روش‌های ذکر شده

در جدول زیر مقایسه‌ای کلی از روش‌های مدیریت کیفیت داده بر پایه موارد زیر ارائه شده است:

- تحلیل داده‌ها
- تحلیل الزامات کنترل داده
- شناسایی حوزه‌های کلیدی
- مدل‌سازی فرایند
- اندازه‌گیری کیفیت
- قابل توسعه به ابعاد و معیارها

جدول ۳-۱ جدولی از جمع‌بندی روش‌های مطالعه شده

مخفف روش	نام اصلی	تحلیل داده‌ها	تحلیل الزامات کنترل داده	شناسایی حوزه‌های کلیدی	مدل سازی فرایند	اندازه‌گیری کیفیت	قابل توسعه به دیگر ابعاد و معیارها
TDQM	مدیریت کیفیت داده	+		+	+	+	ثابت
DWQ	روش‌شناسی کیفیت انبار داده	+	+	+		+	باز
TIQM	مدیریت کیفیت جامع اطلاعات	+	+	+	+	+	ثابت
AIMQ	روشی برای ارزیابی کیفیت اطلاعات	+		+		+	ثابت
CIHI	روش موسسه کانادایی برای روش اطلاعات سلامت	+		+			ثابت
DQA	ارزیابی کیفیت داده	+		+		+	باز
IQM	اندازه‌گیری کیفیت اطلاعات	+				+	باز
ISTAT	روش ISTAT	+				+	ثابت
AMEQ	روش اندازه‌گیری مبتنی بر فعالیت و ارزیابی کیفیت اطلاعات محصول	+		+	+	+	باز
DaQuin CIS Data	کیفیت در سیستم‌های اطلاعات تعاونی	+		+	+	+	باز
QAFD	روشی برای ارزیابی کیفیت داده‌های مالی	+	+	+		+	ثابت
CDQ	روشی جامع برای مدیریت کیفیت داده	+	+	+	+	+	باز

فصل چهارم

کاربرد مفاهیم کیفیت داده در علم و فناوری

۴-۱. مقدمه

در سه فصل نخست این کتاب به مفاهیم پایه‌ای مدیریت کیفیت، کیفیت داده و مدیریت کیفیت داده پرداخته شد و چارچوب‌ها و مدل‌های استاندارد در این زمینه ارائه گردید. در این بخش به یک نمونه کاربردی از استفاده از مدل‌های مدیریت کیفیت داده در بهبود کیفیت داده‌های علم و فناوری خواهیم پرداخت. به‌منظور ارائه یک نمونه کاربردی از داده‌های حاصل از پایان‌نامه/رساله‌های دانش‌آموختگان کل کشور انتخاب‌شده و یک مدل کیفیت داده را به‌منظور ایجاد بهبود در این دسته از داده‌ها که نمایانگر پژوهش‌های علم و فناوری در کشور هستند به کار می‌گیریم.

سیستم ثبت پایان‌نامه/رساله‌های دانش‌آموختگان کل کشور به‌عنوان یک سیستم‌های حیاتی در ثبت و اشاعه مدارک علمی بوده و نقشی انکارناپذیر در توسعه پژوهش‌ها دارد. از این‌رو سطح کیفیت مطلوب سیستم تأثیر ویژه‌ای در ارتقاء کیفیت پژوهش‌های محققین خواهد داشت. نام، کد ملی، مقطع تحصیلی، رشته دانشگاهی مشخصات استاد راهنما، مشخصات پیشنهاد و پایان‌نامه از مهم‌ترین اقلام اطلاعاتی می‌باشند. به‌منظور بهبود سطح کیفیت در این سیستم تعیین وضعیت مطلوب هریک

از اقلام اطلاعاتی سیستم ثبت اقدامی بنیادین محسوب می‌گردد. در این راستا مدل‌های کیفیت داده از متون علمی مختلف استخراج و جمع‌بندی شده و وضعیت مطلوب کیفیت فراداده‌های مربوط به پیشنهاد/پایان‌نامه/رساله‌ها ترسیم خواهد شد. بررسی وضعیت موجود فرایند ثبت و اشاعه مدارک علمی، مشکلات کیفی مشاهده‌شده، سیاست‌ها و راهبردهای اشاعه از اقدامات اصلی خواهد بود.

۲-۴. معرفی فیلدهای سیستم ثبت

همان‌گونه که در مقدمه تبیین گردید در سیستم ثبت پایان‌نامه/رساله دانشجویان تحصیلات تکمیلی اقلام اطلاعاتی متفاوتی از دانشجویان دریافت می‌گردد. در این فصل ابعاد مدل کیفیت داده را بر روی اقلام مختلف فراداده پایان‌نامه‌ها و رساله‌ها ترسیم خواهیم کرد. به این منظور، در ابتدا به معرفی اقلام اطلاعاتی قابل ثبت در سیستم ثبت خواهیم پرداخت. در ادامه و در جدول زیر به معرفی این اقلام در قالب شناسنامه فیلدهای این سیستم خواهیم پرداخت.

جدول ۴-۱ شناسنامه فیلدهای فراداده‌های ثبت پایان‌نامه‌ها/رساله‌ها

نام فیلد	نام فیلد	ویژگی فیلد	توضیحات
نام	نام	صرفاً حروف	مقداری ثابت و تنها قابل تغییر با مراجعه به پروفایل کاربری
نام خانوادگی	نام خانوادگی	صرفاً حروف	مقداری ثابت و تنها قابل تغییر با مراجعه به پروفایل کاربری
رایانامه	رایانامه		مقداری ثابت و غیرقابل تغییر
شماره ملی	شماره ملی	صرفاً عدد	مقداری ثابت و غیرقابل تغییر
شماره دانشجویی	شماره دانشجویی	صرفاً عدد	نباید کمتر از ۴ و بیشتر از ۱۰ رقم باشد
مقطع تحصیلی	مقطع تحصیلی	وجود combo box	دارای قابلیت جستجو
وابستگی سازمانی	وابستگی سازمانی	وجود combo box	دارای ارتباط درونی با فیلد بعدی خود با قابلیت جستجو
موسسه آموزشی/پژوهشی	موسسه آموزشی/پژوهشی	وجود combo box	دارای ارتباط درونی با فیلد بعدی خود با قابلیت جستجو

نام فیلد	نام فیلد	ویژگی فیلد	توضیحات
دانشکده/ پژوهشکده	دانشکده/ پژوهشکده	وجود combo box	
زبان اصلی پایان نامه/ رساله	زبان اصلی پایان نامه/ رساله	وجود combo box	با قابلیت جستجو
عنوان (فارسی)	عنوان (فارسی)		نباید کمتر از ۱۰ و بیشتر از ۵۰۰ حرف باشد
عنوان (انگلیسی)	عنوان (انگلیسی)		نباید کمتر از ۱۰ و بیشتر از ۵۰۰ حرف باشد
گروه تحصیلی	گروه تحصیلی	وجود combo box	دارای ارتباط درونی با فیلد بعدی خود با قابلیت جستجو
رشته	رشته	وجود combo box	دارای ارتباط درونی با فیلد بعدی خود با قابلیت جستجو
گرایش	گرایش	وجود combo box	با قابلیت جستجو
تاریخ دفاع	تاریخ دفاع	وجود combo box	به فرمت تاریخ شمسی با قابلیت جستجو
نام خانوادگی نام خانوادگی (انگلیسی) نام خانوادگی (انگلیسی) نام خانوادگی (انگلیسی) رایانامه شماره تلفن همراه شماره ملی	نام	*	صرفاً حروف
	نام خانوادگی	*	صرفاً حروف
	نام (انگلیسی)		صرفاً حروف انگلیسی
	نام خانوادگی (انگلیسی)		صرفاً حروف انگلیسی
	رایانامه	*	
	شماره تلفن همراه		صرفاً عدد
	شماره ملی		صرفاً عدد
چکیده (فارسی)	چکیده (فارسی)		نباید کمتر از ۲۰ و بیشتر از ۳۰۰۰ حرف باشد
چکیده (انگلیسی)	چکیده (انگلیسی)		نباید کمتر از ۲۰ و بیشتر از ۳۰۰۰ حرف باشد
نوع پایان نامه/ رساله	نوع پایان نامه/ رساله		قابل انتخاب از بین گزینه‌های موجود
فهرست مطالب	فهرست مطالب		نباید کمتر از ۲۰ و بیشتر از ۲۰۰۰۰ حرف باشد
فهرست منابع فارسی	فهرست منابع فارسی		نباید کمتر از ۲۰ و بیشتر از ۲۰۰۰۰ حرف باشد

نام فیلد	نام فیلد	ویژگی فیلد	توضیحات
فهرست منابع غیرفارسی	فهرست منابع غیرفارسی		نباید کمتر از ۲۰ و بیشتر از ۲۰۰۰ حرف باشد
کلیدواژه‌ها	کلیدواژه‌ها		حداقل تعداد کلیدواژه‌ها به فارسی و انگلیسی برای هر زبان دو مورد است با حداکثر سه مورد
فایل pdf پایان‌نامه / رساله	فایل pdf پایان‌نامه / رساله		قالب فایل pdf مجاز است
فایل word پایان‌نامه / رساله	فایل word پایان‌نامه / رساله		قالب فایل docx, doc مجاز است
دیگر فایل‌های پایان‌نامه / رساله	دیگر فایل‌های پایان‌نامه / رساله		

در ویرایش جدید سیستم ثبت، فیلدهای سیستم به‌خوبی طراحی شده‌اند و مشکلات کیفی قبلی تا حد زیادی با روش‌های مختلف برطرف شده‌اند. با اینهمه می‌توان برخی نکات زیر را در طراحی‌های آینده موردتوجه قرار داد:

- در قسمت پروفایل کاربری برای ورود در فیلدهای نام و نام خانوادگی و نام پدر ورود کاراکترهای ؟، /، (،)، % و ... مجاز است؛
- در قسمت نام «فیلد شماره دانشجویی» وارد کردن مقادیر حروف و کاراکترهایی مانند ؟، /، (،)، %، \$ و ... مجاز است؛
- در قسمت استاد راهنما / مشاور، فیلدهای نام و نام خانوادگی به‌صورت فارسی و انگلیسی وارد کردن کاراکترهای ؟، /، (،)، % و ... مجاز است؛
- در فیلد چکیده (فارسی) قابلیت ورود حروف انگلیسی وجود دارد؛
- در فیلد چکیده (انگلیسی) قابلیت ورود حروف فارسی وجود ندارد؛
- برای انتخاب فایل پایان‌نامه در فیلد «فایل word پایان‌نامه / رساله» این امکان برای دانشجو وجود دارد که اگر پایان‌نامه / رساله خود را با نرم‌افزار دیگری تهیه کرده باشد، می‌تواند فهرست مطالب و همچنین فهرست منابع خود را در این قسمت بارگذاری کرده و دیگر بخش‌ها را در فیلد «دیگر فایل‌های پایان‌نامه / رساله» بارگذاری کند.

لازم به ذکر است که موارد بالا برخی نکات بدست آمده در بررسی اولیه است. در انتهای مطالعه کاربردی حاضر و بررسی جایگاه ابعاد مختلف فراداده‌ها بر روی فیلدهای سیستم ثبت می‌توان به الگوی جامعی در خصوص نحوه کنترل کیفی فراداده‌های ثبت شده در سیستم دست یافت و درعین حال می‌توان از روش‌ها و اصولی برای جلوگیری از بروز مشکلات کیفی استفاده کرد. در ادامه به معرفی مدل کیفیت فراداده انتخاب شده و ابعاد مختلف آن به منظور نگاشت آن بر روی اقسام داده‌ای پایان‌نامه‌ها/ رساله‌ها خواهیم پرداخت.

۴-۳. معرفی مدل کیفیت فراداده انتخابی

همان‌گونه که در فصل سوم این گزارش به تفصیل اشاره گردید، مدل‌ها و چارچوب‌های مختلفی برای کیفیت داده در متون علمی وجود دارد. یکی از مهم‌ترین این چارچوب‌ها که در مقالات مختلفی مورد توجه قرار گرفته و در فصل سوم نیز به آن اشاره گردید، مدل مدیریت کیفیت جامع داده است.

مدل مدیریت کیفیت داده به لحاظ ساختاری و مدیریتی تشابه زیادی به مدل بهبود مستمر دمی‌نگ دارد که در فصل اول به آن اشاره شده است. در این روش، هر ۴ مرحله یعنی برنامه‌ریزی، اجرا، کنترل و اقدام مورد نیاز است به گونه‌ای که سازمان بتواند اقدامات لازم برای تضمین کیفیت خدمات خود و کاهش خطا در فرایندهای تولید داده و اطلاعات را انجام دهد.

مدل مدیریت کیفیت داده چهار سطح مختلف از ابعاد را در معیارهای مختلف مورد توجه قرار داده است که به شرح زیر هستند:

۱. کیفیت ذاتی داده‌ها^۱

کیفیت ذاتی داده‌ها نه تنها دقت و عینیت را در بردارد بلکه باورپذیری و اعتبار عمومی را نیز شامل می‌شود. صحت به معنی ارائه مقادیر واقعی مشخصه‌های مورد نظر یک مفهوم در حوزه عملیاتی مورد نظر است. مشتریان داده‌ها برخلاف باور سنتی

مهندسان فناوری اطلاعات فقط صحت و عینیت را موردتوجه قرار نمی‌دهند بلکه باورپذیری و اعتبار عمومی داده‌ها نیز برایشان مهم است. برای مثال صحت و عینیت داده‌های آماری زمانی باورپذیرتر می‌شود که به مرجع ارائه آمار نیز اشاره شود.

۲. کیفیت زمینه‌ای داده‌ها^۱

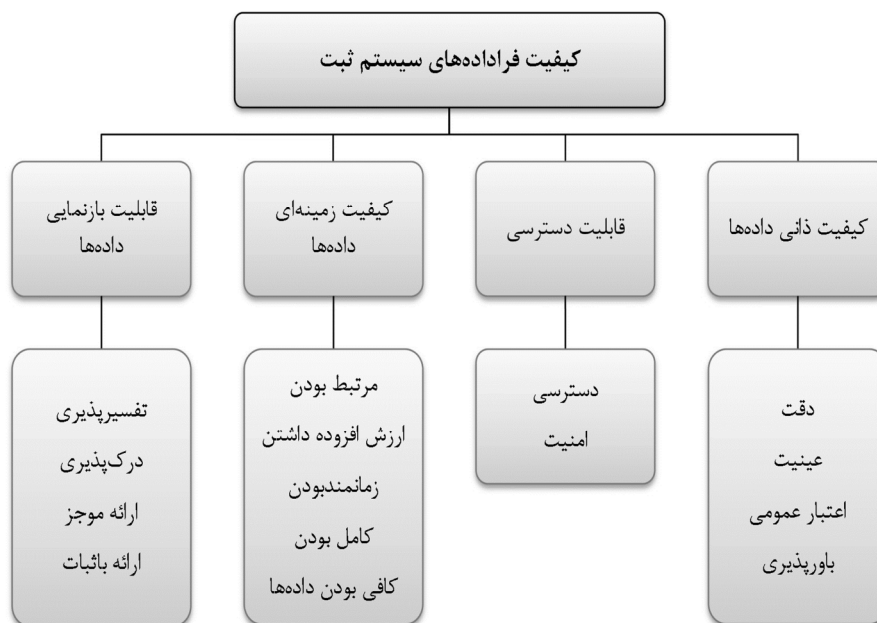
کیفیت داده‌ها باید در زمینه کار با داده‌ها، موردتوجه قرار گیرد. این بعد شامل ابعاد فرعی مانند دارا بودن ارزش افزوده، مرتبط بودن، زمانمند بودن، کامل بودن و کافی بودن داده‌هاست. از آنجا که فعالیت‌هایی که با داده‌ها انجام می‌گیرد در طول زمان متغیر بوده و نیز کاربران داده‌ها نیز متفاوت هستند، ارزیابی کیفیت زمینه‌ای داده‌ها کاری دشوار است. یک روش برای انجام این ارزیابی، تعیین پارامترهای ابعاد ارزیابی هر کار است تا مشتری بتواند بر مبنای آن مشخص کند که چه نوع کاری در حال انجام است و ویژگی داده‌های مورد استفاده چگونه باید باشد.

۳. قابلیت بازنمایی داده‌ها^۲

قابلیت بازنمایی داده‌ها درباره چارچوب ارائه (ارائه کوتاه و باثبات) و معنای داده‌ها (تفسیرپذیری و درک‌پذیری) است. این دو جنبه نشان می‌دهند که داده‌ها چقدر برای کاربران داده‌ها خوب ارائه شده‌اند.

۴. قابلیت دسترسی^۳

به‌طور سنتی دسترسی، شاخص جداگانه‌ای از ابعاد کیفیت شناخته می‌شد اما تحقیق وانگ و همکارانش نشان داد که شاخصی کیفی و موردتوجه کاربران داده‌هاست. دسترسی خود شامل بر دو بعد فرعی دسترسی به داده‌ها و نیز امنیت دسترسی به داده‌هاست.



شکل ۴-۱ ابعاد کیفیت فراداده‌ها در سیستم ثبت

۴-۴. نگاشت اولیه ابعاد کیفی فراداده روی ارقام اطلاعاتی سیستم ثبت

در این بخش به ازای هریک از ابعاد کیفی مدل مدیریت کیفیت داده که در بخش قبلی مورد شناسایی قرار گرفت، ارتباط میان ارقام اطلاعاتی سیستم ثبت پایان‌نامه/ رساله با آن ابعاد موردبررسی قرار می‌گیرد. جدول‌های ۴-۲ تا ۴-۵ به ازای هر یک از چهار گروه شکل ۴-۱ این ارتباط را نمایان می‌سازد.

در خصوص جدول بالا شایان ذکر است در برخی از ابعاد به ازای هر فیلد از پایان‌نامه/ رساله آن بعد قابل تبیین و تفسیر خواهد بود. برای نمونه در بعد دقت در فیلد «نام» مواردی بایستی موردتوجه قرار گیرد که با فیلد «کد ملی» متفاوت است. از سوی دیگر برخی ابعاد مانند باورپذیری یا اعتبار عمومی به‌کل فراداده‌های پایان‌نامه/رساله‌ها تفسیر شده‌اند.

جدول ۴-۲ طرح کیفی فراداده‌های فیلدهای ثبت پایان‌نامه‌ها در سیستم ثبت (کیفیت ذاتی)

کیفیت ذاتی داده‌ها ^۱				نام فیلد	بعد
اعتبار عمومی ^۵	باورپذیری ^۴	عینیت ^۳	دقت ^۲		
معتبر بودن در سیستم ثبت پایان‌نامه: تمامی فیلدهای سیستم دارای اعتبار هستند و هر کاربری با مشاهده آن‌ها می‌تواند به سوالات پاسخ دهد	در اطلاعات مربوط به متن پایان‌نامه/رساله قواعد نگارشی به‌درستی رعایت شده باشد و متن از گویایی کامل برخوردار باشد تا کاربران سیستم گنج به‌راحتی بتوانند از این اطلاعات استفاده نمایند. علاوه در انتخاب واژگان کلیدی و عنوان مناسب توسط نویسنده این موضوع رعایت شده باشد. همچنین در مرحله بعدی از فرایند اشاعه یعنی نمایه‌سازی و ویرایش این موضوع می‌تواند موردتوجه قرار گیرد.	مطابقت با مستندات هویتی (مانند کارت ملی)	مطابقت با متن پایان‌نامه/رساله	نام	
		مطابقت با مستندات هویتی (مانند کارت ملی)	مطابقت با متن پایان‌نامه/رساله	نام خانوادگی	
			در زمان ایجاد پروفایل کاربری برای ادامه ثبت‌نام به رایانامه داده شده یک ایمیل حاوی لینک ورود به قسمت ثبت پایان‌نامه‌ها فرستاده می‌شود. حتماً باید فرمت زیر را رعایت کند: xxx@xxx.com	رایانامه	
			مطابقت با الگوریتم شماره ملی مطابقت با وب‌سرویس تطبیق کد ملی و نام و نام خانوادگی با ثبت‌احوال	شماره ملی	
		مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل	نباید کمتر از ۴ و بیشتر از ۱۰ رقم باشد	شماره دانشجویی	

1. Intrinsic

2. Accuracy

3. Objectivity

4. Believability

5. Reputation

نام فیلد	بعد	کیفیت ذاتی داده‌ها ^۱			اعتبار عمومی ^۵
		دقت ^۲	عینیت ^۳	باورپذیری ^۴	
مقطع تحصیلی		مطابقت با متن پایان‌نامه/ رساله	مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل		معتبر بودن در سیستم ثبت پایان‌نامه: تمامی فیلدهای سیستم دارای اعتبار هستند و هر کاربری با مشاهده آن‌ها می‌تواند به سوالات پاسخ دهد.
وابستگی سازمانی		مطابقت با متن پایان‌نامه/ رساله	مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل	در اطلاعات مربوط به متن پایان‌نامه/ رساله قواعد نگارشی به‌درستی رعایت شده باشد و متن از گویایی کامل برخوردار باشد	
موسسه آموزشی/پژوهشی		مطابقت با متن پایان‌نامه/ رساله		تا کاربران سیستم گنج به‌راحتی بتوانند از این اطلاعات استفاده نمایند.	
دانشکده/پژوهشکده		مطابقت با متن پایان‌نامه/ رساله		بعلاوه در انتخاب واژگان کلیدی و عنوان مناسب توسط نویسنده این موضوع رعایت شده باشد.	
زبان اصلی پایان‌نامه		مطابقت با متن پایان‌نامه/ رساله		همچنین در مرحله بعدی از فرایند اشاعه یعنی نمایه‌سازی و ویرایش این موضوع می‌تواند موردتوجه قرار گیرد.	
عنوان فارسی		نباید کمتر از ۱۰ و بیشتر از ۵۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله		
عنوان انگلیسی		نباید کمتر از ۱۰ و بیشتر از ۵۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله		
گروه تحصیلی		مطابقت با متن پایان‌نامه/ رساله	مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل		
رشته		مطابقت با متن پایان‌نامه/ رساله	مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل		

نام فیلد	بعد	کیفیت ذاتی داده‌ها ^۱			اعتبار عمومی ^۵
		دقت ^۲	عینیت ^۳	باورپذیری ^۴	
استاد راهنما/ استاد مشاور	گرایش	مطابقت با متن پایان‌نامه/ رساله	مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل	در اطلاعات مربوط به متن پایان‌نامه/ رساله قواعد نگارشی به‌درستی رعایت شده باشد و متن از گویایی کامل برخوردار باشد تا کاربران سیستم گنج به‌راحتی بتوانند از این اطلاعات استفاده نمایند. به‌علاوه در انتخاب واژگان کلیدی و عنوان مناسب توسط نویسنده این موضوع رعایت شده باشد. همچنین در مرحله بعدی از فرایند اشاعه یعنی نمایه‌سازی و ویرایش این موضوع می‌تواند موردتوجه قرار گیرد.	معتبر بودن در سیستم ثبت پایان‌نامه: تمامی فیلدهای سیستم دارای اعتبار هستند و هر کاربری با مشاهده آن‌ها می‌تواند به سؤالات پاسخ دهد.
	تاریخ دفاع	مطابقت با متن پایان‌نامه/ رساله	مطابقت با اطلاعات مندرج در پایگاه علمی محل تحصیل		
	نام	نباید کمتر از ۲ و بیشتر از ۲۵ حرف باشد مطابقت با متن پایان‌نامه/ رساله	مطابقت با متن پایان‌نامه/ رساله		
	نام خانوادگی	نباید کمتر از ۲ و بیشتر از ۵۰ حرف باشد مطابقت با متن پایان‌نامه/ رساله	مطابقت با متن پایان‌نامه/ رساله		
	نام (انگلیسی)	نباید کمتر از ۲ و بیشتر از ۲۵ حرف باشد مطابقت با متن پایان‌نامه/ رساله			
	نام خانوادگی (انگلیسی)	نباید کمتر از ۲ و بیشتر از ۵۰ حرف باشد مطابقت با متن پایان‌نامه/ رساله			
	رایانامه	حتماً باید فرمت زیر را رعایت کند: xxx@xxx.com			
	شماره تلفن همراه	باید به‌صورت عدد و با فرمت صحیح باشد، یعنی ۰۹ شروع شود.			
	شماره ملی	مطابقت با الگوریتم شماره ملی			

نام فیلد	بعد	کیفیت ذاتی داده‌ها ^۱			اعتبار عمومی ^۵
		دقت ^۲	عینیت ^۳	باورپذیری ^۴	
چکیده (فارسی)		نباید کمتر از ۲۰ و بیشتر از ۳۰۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله		معتبر بودن در سیستم ثبت پایان‌نامه: تمامی فیلدهای سیستم دارای اعتبار هستند و هر کاربری با مشاهده آن‌ها می‌تواند به سوالات پاسخ دهد.
چکیده (انگلیسی)		نباید کمتر از ۲۰ و بیشتر از ۳۰۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله	در اطلاعات مربوط به متن پایان‌نامه/ رساله قواعد نگارشی به‌درستی رعایت شده باشد و متن از گویایی کامل برخوردار باشد تا کاربران سیستم گنج به‌راحتی بتوانند از این اطلاعات استفاده نمایند. به‌علاوه در انتخاب واژگان کلیدی و عنوان مناسب توسط نویسنده این موضوع رعایت شده باشد. همچنین در مرحله بعدی از فرایند اشاعه یعنی نمایه‌سازی و ویرایش این موضوع می‌تواند موردتوجه قرار گیرد.	
نوع پایان‌نامه/ رساله					
فهرست مطالب		نباید کمتر از ۲۰ و بیشتر از ۲۰۰۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله		
فهرست منابع فارسی		نباید کمتر از ۲۰ و بیشتر از ۲۰۰۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله		
فهرست منابع غیرفارسی		نباید کمتر از ۲۰ و بیشتر از ۲۰۰۰۰ حرف باشد	مطابقت با متن پایان‌نامه/ رساله		
کلیدواژه‌ها		کامل بودن فیلدها	مطابقت با متن پایان‌نامه/ رساله		
فایل pdf پایان‌نامه/ رساله		عدم پذیرش فایل با فرمتی دیگر	کامل بودن محتوا		
فایل word پایان‌نامه/ رساله		عدم پذیرش فایل با فرمتی دیگر	کامل بودن محتوا		
دیگر فایل‌های پایان‌نامه/ رساله					

جدول ۴-۳ طرح کیفی فیلدهای فراداده‌های ثبت پایان‌نامه‌ها در سیستم ثبت (کیفیت زمینه‌ای)

نام فیلد	بعد	کیفیت زمینه‌ای داده‌ها			
		مرتبط بودن ^۱	ارزش افزوده ^۲ داشتن	زمانمند بودن	کافی بودن ^۳ داده‌ها
نام	نامی	تمامی اطلاعاتی که باید در اختیار سیستم ثبت پایان‌نامه قرار گرفته شود به صورت فیلدهایی تعبیه شده است.	اطلاعات دریافتی در حال حاضر ارزش افزوده کافی را از منظر کاربران سیستم (گنج) دارد و هیچ اطلاعات اضافی دریافت نمی‌شود.	به روز بودن در سیستم ثبت پایان‌نامه: تمامی فیلدها به جز فیلدهای نام، نام خانوادگی، رایانامه و شماره ملی که در مرحله ورود اطلاعات شخصی و تحصیلی به صورت از پیش تعیین شده و باتوجه به اطلاعات پروفایل کاربری است می‌تواند وارد مراحل ثبت پایان‌نامه شود و اطلاعات خود را در فیلدهای مشخص شده وارد نماید. وارد کردن اطلاعات در هر فیلدی که با ستاره مشخص شده است برای فرد اجباری است و فیلدهایی که ستاره دار نمی‌باشند، اختیاری است. به این ترتیب برای هر فرد که در سیستم پروفایل کاربری دارد یکسری اطلاعات قبل از شروع به مراحل ثبت پایان‌نامه د اختیار سیستم قرار می‌گیرد.	صرفاً حروف را می‌پذیرد
نام خانوادگی	نام خانوادگی	به منظور ارزیابی میزان مرتبط بودن داده‌های هر رکورد کلیه نیازمندی‌ها در سیستم‌های آتی را شناسایی نمود تا در صورت نیاز فیلدی اضافه یا کم نمود. برای نمونه اگر در آینده قرار است بانک اطلاعاتی پژوهشگران تقویت و تکمیل شود آدرس Google Scholar هر نویسنده یا استاد وی اخذ شود (یک فیلد اضافه شود).	فیلدهای دیگر به سیستم ثبت جهت افزایش ارزش افزوده بایستی در جلسات کارشناسی تصمیم‌گیری شود	هر فرد، هر فردی که دارای پروفایل کاربری است می‌تواند وارد مراحل ثبت پایان‌نامه شود و اطلاعات خود را در فیلدهای مشخص شده وارد نماید. وارد کردن اطلاعات در هر فیلدی که با ستاره مشخص شده است برای فرد اجباری است و فیلدهایی که ستاره دار نمی‌باشند، اختیاری است. به این ترتیب برای هر فرد که در سیستم پروفایل کاربری دارد یکسری اطلاعات قبل از شروع به مراحل ثبت پایان‌نامه د اختیار سیستم قرار می‌گیرد.	صرفاً حروف را می‌پذیرد
رایانامه	رایانامه				پذیرفتن حروف و عدد
شماره ملی	شماره ملی				پذیرفتن عدد
شماره دانشجویی	شماره دانشجویی				پذیرفتن عدد
مقطع تحصیلی	مقطع تحصیلی				
وابستگی سازمانی	وابستگی سازمانی				
موسسه آموزشی/پژوهشی	موسسه آموزشی/پژوهشی				
دانشکده/پژوهشکده	دانشکده/پژوهشکده				
زبان اصلی پایان‌نامه	زبان اصلی پایان‌نامه				
عنوان فارسی	عنوان فارسی				پذیرفتن حروف و عدد

1. Relevancy

2. Value Added

3. Amount Of Data

کیفیت زمینه‌ای داده‌ها					بعد	
نام فیلد	مرتبط بودن ^۱	ارزش افزوده ^۲ داشتن	زمانمند بودن	کامل بودن	کافی بودن ^۳ داده‌ها	
استاد راهنما/ استاد مشاور	عنوان انگلیسی	تمامی اطلاعاتی که باید در اختیار سیستم ثبت پایان‌نامه قرار گرفته شود به صورت فیلدهایی تعبیه شده است.	اطلاعات دریافتی در حال حاضر ارزش افزوده کافی را از منظر کاربران سیستم (گنج) دارد و هیچ اطلاعات اضافی دریافت نمی‌شود.	به روز بودن در سیستم ثبت پایان‌نامه: تمامی فیلدها به جز فیلدهای نام، نام خانوادگی، رایانامه و شماره ملی که در مرحله ورود اطلاعات شخصی و تحصیلی به صورت از پیش تعیین شده و با توجه با اطلاعات پروفایل کاربری کامل شده‌اند، می‌توانند قبل از تأیید نهایی کاربر به روز شوند و در آن‌ها تغییرات ایجاد شود. البته فیلدهای شماره ملی و رایانامه از این قاعده مستثنا هستند و کاربر در آن‌ها نمی‌تواند تغییری را اعمال کند.	جامعیت در سیستم ثبت پایان‌نامه: می‌توان اشاره کرد که با ثبت نام هر دانشجو و اختصاص پروفایل کاربری به هر فرد، هر فردی که دارای پروفایل کاربری است می‌تواند وارد مراحل ثبت پایان‌نامه شود و اطلاعات خود را در فیلدهای مشخص شده وارد نماید. وارد کردن اطلاعات در هر فیلدی که با ستاره مشخص شده است برای فرد اجباری است و فیلدهایی که ستاره دار نمی‌باشند، اختیاری است. به این ترتیب برای هر فرد که در سیستم پروفایل کاربری دارد یکسری اطلاعات قبل از شروع به مراحل ثبت پایان‌نامه د اختیار سیستم قرار می‌گیرد.	پذیرفتن حروف و عدد
	گروه تحصیلی					
	رشته					
	گرایش					
	تاریخ دفاع					
	نام	بودن داده‌های هر رکورد کلیه نیازمندی‌ها در سیستم‌های آتی را شناسایی نمود تا در صورت نیاز فیلدی اضافه یا کم نمود. برای نمونه اگر در آینده قرار است بانک اطلاعاتی پژوهشگران تقویت و تکمیل شود آدرس Google Scholar هر نویسنده یا استاد وی اخذ شود (یک فیلد اضافه شود).				پذیرفتن حروف و عدد
	نام خانوادگی					پذیرفتن حروف و عدد
	نام (انگلیسی)					پذیرفتن حروف و عدد
	نام خانوادگی (انگلیسی)					پذیرفتن حروف و عدد
	رایانامه					پذیرفتن حروف و عدد
شماره ملی	شماره تلفن همراه					پذیرفتن حروف و عدد
	شماره ملی					پذیرفتن حروف و عدد

کیفیت زمینه‌ای داده‌ها					نام فیلد
کافی بودن داده‌ها ^۳	کامل بودن	زمانمند بودن	ارزش افزوده داشتن ^۲	مرتبط بودن ^۱	
	جامعیت در سیستم ثبت پایان‌نامه: می‌توان اشاره کرد که با ثبت نام هر دانشجو و اختصاص پروفایل کاربری به هر فرد، هر فردی که دارای پروفایل کاربری است می‌تواند وارد مراحل ثبت پایان‌نامه شود و اطلاعات خود را در فیلدهای مشخص شده وارد نماید. وارد کردن اطلاعات در هر فیلدی که با ستاره مشخص شده است برای فرد اجباری است و فیلدهایی که ستاره‌دار نمی‌باشند، اختیاری است. به این ترتیب برای هر فرد که در سیستم پروفایل کاربری دارد یکسری اطلاعات قبل از شروع به مراحل ثبت پایان‌نامه د اختیار سیستم قرار می‌گیرد.	به روز بودن در سیستم ثبت پایان‌نامه: تمامی فیلدها به جز فیلدهای نام، نام خانوادگی، رایانامه و شماره ملی که در مرحله ورود اطلاعات شخصی و تحصیلی به صورت از پیش تعیین شده و باتوجه با اطلاعات پروفایل کاربری کامل شده‌اند، می‌توانند قبل از تأیید نهایی کاربر به روز شوند و در آن‌ها تغییرات ایجاد شود. البته فیلدهای شماره ملی و رایانامه از این قاعده مستثنا هستند و کاربر در آن‌ها نمی‌تواند تغییری را اعمال کند.	اطلاعات دریافتی در حال حاضر ارزش افزوده کافی را از منظر کاربران سیستم (گنج) دارد و هیچ اطلاعات اضافی دریافت نمی‌شود. در خصوص اضافه شدن فیلدهای دیگر به سیستم ثبت جهت افزایش ارزش افزوده بایستی در جلسات کارشناسی تصمیم‌گیری شود	تمامی اطلاعاتی که باید در اختیار سیستم ثبت پایان‌نامه قرار گرفته شود به صورت فیلدهایی تعبیه شده است. به منظور ارزیابی میزان مرتبط بودن داده‌های هر رکورد کلیه نیازمندی‌ها در سیستم‌های آتی را شناسایی نمود تا در صورت نیاز فیلدی اضافه یا کم نمود. برای نمونه اگر در آینده قرار است بانک اطلاعاتی پژوهشگران تقویت و تکمیل شود آدرس Google Scholar هر نویسنده یا استاد وی اخذ شود (یک فیلد اضافه شود).	چکیده (فارسی)
پذیرش صرفاً حروف انگلیسی					چکیده (انگلیسی)
					نوع پایان‌نامه / رساله
پذیرفتن حروف و عدد					فهرست مطالب
پذیرفتن حروف و عدد					فهرست منابع فارسی
پذیرش صرفاً حروف و اعداد انگلیسی					فهرست منابع غیرفارسی
پذیرفتن حروف و عدد					کلیدواژه‌ها
پذیرش فایل					فایل pdf پایان‌نامه / رساله
پذیرش فایل					فایل word پایان‌نامه / رساله
پذیرش فایل					دیگر فایل‌های پایان‌نامه / رساله

جدول ۴-۴ طرح کیفی فیلدهای فراداده‌های ثبت پایان‌نامه‌ها در سیستم ثبت (کیفیت بازنمایی)

قابلیت بازنمایی داده‌ها				بعد	
ارائه بایدها	ارائه موجز	درک پذیری	تفسیرپذیری	نام فیلد	
طراحی سیستم ثبت باید به گونه‌ای باشد که داده‌ها و فراداده‌های ثبت شده در آن با تغییرات آتی در سخت افزارها و به روز شدن نرم افزارها ارائه و نمایش با مشکل مواجه نشود.	فرمت کاراکترهای تایپ شده بایستی به گونه‌ای باشد که در نمایش در سیستم‌های بعدی مانند نمایه سازی و ویرایش و همچنین گنج فرمت مطلوب کاربران را داشته باشد.	سهولت درک در سیستم ثبت پایان نامه: تمامی فیلدهای موجود در سیستم به راحتی برای کاربر قابل تشخیص و قابل فهم می باشد.	فرمت و ساختار کلیه فیلدهای سیستم ثبت باید به گونه‌ای باشد که به راحتی و بدون اشکال در سیستم گنج انتقال داده شود و برای کاربران سیستم گنج که از مرورگرها و سیستم عامل های مختلف قابل نمایش باشد.	نام	استاد راهنما/ استاد مشاور
				نام خانوادگی	
				رایانامه	
				شماره ملی	
				شماره دانشجویی	
				مقطع تحصیلی	
				وابستگی سازمانی	
				موسسه آموزشی/پژوهشی	
				دانشکده/پژوهشکده	
				زبان اصلی پایان نامه	
				عنوان فارسی	
				عنوان انگلیسی	
				گروه تحصیلی	
				رشته	
				گرایش	
				تاریخ دفاع	
				نام	
				نام خانوادگی	
				نام (انگلیسی)	
				نام خانوادگی (انگلیسی)	
				رایانامه	
				شماره تلفن همراه	
				شماره ملی	

قابلیت بازنمایی داده‌ها				بعد
ارائه اثبات	ارائه موجز	درک پذیری	تفسیر پذیری	نام فیلد
				چکیده (فارسی)
				چکیده (انگلیسی)
				نوع پایان‌نامه / رساله
				فهرست مطالب
طراحی سیستم ثبت بایستی به‌گونه‌ای باشد که داده‌ها و فرداده‌های ثبت‌شده در آن با تغییرات آتی در سخت‌افزارها و به‌روز شدن نرم‌افزارها ارائه و نمایش با مشکل مواجه نشود.		سهولت درک	فرمت و ساختار	فهرست منابع فارسی
		در سیستم	کلیه فیلدهای	فهرست منابع غیرفارسی
		ثبت	سیستم ثبت	کلیدواژه‌ها
		پایان‌نامه:	به‌گونه‌ای باشد	فایل pdf پایان‌نامه / رساله
		تمامی	که به‌راحتی و	فایل word پایان‌نامه / رساله
		فیلدهای	بدون اشکال در	دیگر فایل‌های پایان‌نامه / رساله
		موجود در	سیستم گنج	
		سیستم	انتقال داده شود	
		به‌راحتی برای	و برای کاربران	
		کاربر	سیستم گنج که	
		قابل تشخیص	از مرورگرها و	
		و قابل فهم	سیستم‌عامل‌های	
		می‌باشد.	مختلف	
			قابل نمایش باشد.	

جدول ۴-۵ طرح کیفی فیلدهای فراداده‌های ثبت پایان‌نامه‌ها در سیستم ثبت (قابلیت دسترسی)

قابلیت دسترسی		بعد	
امنیت ^۲	دسترسی ^۱	نام فیلد	
امنیت در سیستم ثبت پایان‌نامه: داشتن رمز عبور و رایانامه که هر کاربر تنها برای پروفایل کاربری برای دسترسی به سیستم ثبت پایان‌نامه خود استفاده می‌کند، نمونه‌ای از داشتن امنیت است.	طراحی لایه‌های امنیتی در سیستم ثبت و همچنین سیستم گنج که به‌طور مستقیم با کاربران نهایی در ارتباط هستند باید به‌گونه‌ای باشد که امکان دسترسی به اطلاعات شخصی دانشجویان تحصیلات تکمیلی برای افراد غیرمجاز مقدور نباشد.	نام	
		نام خانوادگی	
		رایانامه	
		شماره ملی	
		شماره دانشجویی	
		مقطع تحصیلی	
		وابستگی سازمانی	
		موسسه آموزشی/پژوهشی	
		دانشکده/پژوهشکده	
		زبان اصلی پایان‌نامه	
		عنوان فارسی	
		عنوان انگلیسی	
		گروه تحصیلی	
		رشته	
		گرایش	
		تاریخ دفاع	
		استاد راهنما/ استاد مشاور	نام
			نام خانوادگی
			نام (انگلیسی)
			نام خانوادگی (انگلیسی)
			رایانامه
			شماره تلفن همراه
			شماره ملی
امنیت در سیستم ثبت پایان‌نامه: داشتن رمز عبور و رایانامه که هر کاربر تنها برای پروفایل کاربری برای دسترسی به سیستم ثبت پایان‌نامه خود استفاده می‌کند، نمونه‌ای از داشتن امنیت است.	طراحی لایه‌های امنیتی در سیستم ثبت و همچنین سیستم گنج که به‌طور مستقیم با کاربران نهایی در ارتباط هستند باید به‌گونه‌ای باشد که امکان دسترسی به اطلاعات شخصی دانشجویان تحصیلات تکمیلی برای افراد غیرمجاز مقدور نباشد	چکیده (فارسی)	
		چکیده (انگلیسی)	
		نوع پایان‌نامه/ رساله	
		فهرست مطالب	
		فهرست منابع فارسی	
		فهرست منابع غیرفارسی	
		کلیدواژه‌ها	
		فایل pdf پایان‌نامه/ رساله	
		فایل word پایان‌نامه/ رساله	
		دیگر فایل‌های پایان‌نامه/ رساله	

۴-۵. نتیجه‌گیری و جمع‌بندی

در فصل حاضر به معرفی ابعاد مختلف مدیریت کیفیت داده در چهار حوزه کیفیت ذاتی، زمینه‌ای، بازنمایی و دسترسی، و همچنین چگونگی جایگاه آن‌ها بر روی کیفیت نهایی اقلام اطلاعاتی سیستم ثبت مورد بررسی قرار گرفت. در واقع به ازای هر قلم اطلاعاتی سیستم ثبت، وضعیت مطلوب آن قلم در این چهار بعد مورد ارزیابی قرار گرفت. همان‌گونه که از جدول‌های ۴-۲ تا ۴-۵ مشاهده می‌شود، هریک از اقلام اطلاعاتی در سیستم ثبت بر پایه مدل TDQM وضعیت مطلوب توصیف شده در این جدول‌ها را خواهد داشت. در این فصل مدل TDQM به‌عنوان یک مدل مطلوب، پرکاربرد و متناسب برای بومی شدن و نهادینه شدن در سیستم ثبت معرفی گردد. چهار مرحله اصلی این مدل شامل برنامه‌ریزی، اندازه‌گیری، تجزیه و تحلیل و بهبود به‌عنوان چارچوب اصلی این مدل مورد اشاره قرار گرفت. می‌توان به کمک گروه‌های خبرگان، هریک از ابعاد کیفی فراداده که در جدول‌های بالا ارائه شده را دقیق‌تر تعریف و برای عملیاتی ساختن پیاده‌سازی آن‌ها، اقدامات لازم را انجام داد.

منابع

منابع فارسی

ارشادی، محمدجواد و احترامی، تینا، ۱۳۹۵. طرح‌ریزی و استقرار نظام تضمین کیفیت در سامانه‌های گردآوری و ثبت، سازمان‌دهی، و اشاعه اطلاعات پایان‌نامه‌ها/ رساله‌های دانش‌آموختگان داخل کشور: طرح پژوهشی در پژوهشگاه علوم و فناوری اطلاعات ایران، تهران.

ارشادی، محمدجواد، رجبی، تقی، شیرانی، فرهاد و رضایی، نساء، ۱۳۹۴. کاربرد تکنیک تحلیل ریشه‌ای در حل مشکلات کیفی سامانه‌های اطلاعاتی تحقیقاتی: مطالعه موردی سامانه اشاعه اطلاعات پایان‌نامه‌ها/ رساله‌های دانش‌آموختگان داخل کشور (گنج). نشریه مدیریت اطلاعات.

علیدوستی، سیروس و عصاره، حسن، ۱۳۸۴. طراحی فرایند کنترل کیفیت نمایه‌سازی: طرح پژوهشی، پژوهشگاه اطلاعات و مدارک علمی ایران، تهران.

منابع لاتین

Abiteboul, S., Buneman, P. & Suciu, D., 2000. Data on the Web: from relations to semistructured data and XML. Morgan Kaufmann publisher.

Aiken, P.H., 1996. Data reverse engineering: slaying the legacy dragon. McGraw-Hill Companies.

Amicis, F.D., 2004. A methodology for data quality assessment on financial data. Studies in Communication Sciences, 4(2), pp.115-137.

Arenas, M., Bertossi, L. & Chomicki, J., 1999. Consistent query answers in inconsistent databases. In Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (pp. 68-79). ACM.

Arisawa, H., Kambayashi, Y., Kumar, V., Mayr, H.C. & Hunt, I. eds., 2002. Conceptual Modeling for New Information Systems Technologies: ER Workshops, HUMACS, DASWIS, ECOMO, and DAMA, Yokohama Japan, November 27-30. Revised Papers (Vol. 2465). Springer Science & Business Media.

Atzeni, P. & D'Antonellis, V., 1993. Relational database theory. Benjamin publisher.

- Ballou, D., Wang, R., Pazer, H. & Tayi, G.K., 1998. Modeling information manufacturing systems to determine information product quality. *Management Science*, 44(4), pp.462-484.
- Ballou, D.P. & Pazer, H.L., 1985. Modeling data and process quality in multi-input, multi-output information systems. *Management science*, 31(2), pp.150-162.
- Baskarada, S., Koronios, A. & Gao, J., 2006. Towards a capability maturity model for information quality management: a TDQM approach.
- Batini, C. & Scannapieco, M., 2016. *Data and information quality*. Cham, Switzerland: Springer International Publishing.
- Batini, C. & Scannapieco, M., 2016. *Data and Information Quality: Dimensions, Principles and Techniques*. In *Data and Information Quality* (pp. E1-E1). Springer, Cham.
- Batini, C. & Scannapieco, M., 2016. Methodologies for information quality assessment and improvement. In *Data and Information Quality* (pp. 353-402). Springer, Cham.
- Batini, C., Cappiello, C., Francalanci, C. & Maurino, A., 2009. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3), p.16.
- Batini, C., Cappiello, C., Francalanci, C. & Maurino, A., 2009. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, 41(3), p.16.
- Bertolazzi, P., De Santis, L. & Scannapieco, M., 2003. Automatic record matching in cooperative information systems. In *Proceedings of the International Workshop on Data Quality in Cooperative Information Systems (DQCIS)* (p. 9).
- Bettschen, P., 2005. Master Data Management (MDM) enables IQ at Tetra Pak. In *IQ*.
- Bilke, A., Bleiholder, J., Naumann, F., Böhm, C., Draba, K. & Weis, M., 2005. Automatic data fusion with HumMer. In *Proceedings of the 31st international conference on very large data bases* (pp. 1251-1254). VLDB Endowment.
- Bovee, M., Srivastava, R.P. & Mak, B., 2003. A conceptual framework and belief-function approach to assessing overall information quality. *International journal of intelligent systems*, 18(1), pp.51-74.
- Buneman, P., 1997. Semistructured data. In *Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems* (pp. 117-121). ACM.
- Buneman, P., Khanna, S. & Wang-Chiew, T., 2001. A characterization of data provenance, Why and where?. In *International conference on database theory* (pp. 316-330). Springer, Berlin, Heidelberg.
- Caldiera, V.R.B.G. & Rombach, H.D., 1994. Goal question metric paradigm. *Encyclopedia of software engineering*, 1, pp.528-532.

- Cali, A., Calvanese, D., De Giacomo, G. & Lenzerini, M., 2004. Data integration under integrity constraints. *Information Systems*, 29(2), pp.147-163.
- Calvanese, D., De Giacomo, G. & Lenzerini, M., 1999. Modeling and querying semi-structured data. *Networking and Information Systems Journal*, 2, pp.253-273.
- Cappiello, C., Francalanci, C. & Pernici, B., 2003. Preserving Web Sites: a Data Quality Approach. In *IQ* (pp. 331-344).
- Cappiello, C., Francalanci, C., Pernici, B., Plebani, P. & Scannapieco, M., 2003. Data quality assurance in cooperative information systems: a multi-dimension quality certificate. In *International Workshop on Data Quality in Cooperative Information Systems (DQCIS 2003-ICDT 2003)* (pp. 47-54).
- Chapman, A., Richards, H. & Hawken, S., 2006. Data and Information Quality at the Canadian Institute for Health Information. In *ICIQ* (pp. 326-340).
- Chengalur-Smith, I.N., Ballou, D.P. & Pazer, H.L., 1999. The impact of data quality information on decision making: An exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering*, 11(6), pp.853-864.
- Corey, D.J., Cobler, L., Lonsdale, T. & Haynes, K., 1996. Data Quality Assurance Activities in the Military Health Services System. In *IQ* (pp. 127-153).
- Dasu, T. & Johnson, T., 2003. *Exploratory data mining and data cleaning* (Vol. 479). John Wiley & Sons.
- De Amicis, F., Barone, D. & Batini, C., 2006, November. An Analytical Framework to Analyze Dependencies Among Data Quality Dimensions. In *ICIQ* (pp. 369-383).
- De Michelis, G., Dubois, E., Jarke, M., Matthes, F., Mylopoulos, J., Papazoglou, M., Pohl, K., Schmidt, J., Woo, C. & Yu, E., 1997. Cooperative information systems: a manifesto. *Cooperative Information Systems: Trends & Directions*, pp.315-165.
- De Michelis, G., Dubois, E., Jarke, M., Matthes, F., Mylopoulos, J., Papazoglou, M., Pohl, K., Schmidt, J., Woo, C. & Yu, E., 1997. Cooperative information systems: a manifesto. *Cooperative Information Systems: Trends & Directions*, pp.315-165.
- Dedeke, A. 2005. Building quality into the information supply chain. *Advances in Management Information Systems-Information Quality Monograph (AMIS-IQ) Monograph*. R. Wang, E. Pierce, S. Madnick, and Fisher C.W., Eds.
- English, L.P., 1999. *Improving data warehouse and business information quality: methods for reducing costs and increasing profits* (Vol. 1). New York: Wiley.
- English, L.P., 2002. Process Management and Information Quality: How Improving Information Production Processes Improves Information (Product) Quality. In *IQ* (pp. 206-209).
- English, L.P., 2009. *Information quality applied: Best practices for improving business information, processes and systems*. Wiley Publishing.

- Eppler, M. & Helfert, M., 2004, November. A classification and analysis of data quality costs. In International Conference on Information Quality (pp. 311-325).
- Eppler, M.J. & Muenzenmayer, P., 2002, November. Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. In IQ (pp. 187-196).
- Falorsi, P.D. & Righi, P., 2008. A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology*, 34(2), pp.223-234.
- Fellegi, I.P. & Holt, D., 1976. A systematic approach to automatic edit and imputation. *Journal of the American Statistical association*, 71(353), pp.17-35.
- Fisher, C.W. & Kingma, B.R., 2001. Criticality of data quality as exemplified in two disasters. *Information & Management*, 39(2), pp.109-116.
- Fraternali, P., Lanzi, P.L., Matera, M. & Maurino, A., 2004. Model-driven Web usage analysis for the evaluation of Web application quality. *J. Web Eng.*, 3(2), pp.124-152.
- Gackowski, Z.J., 2006. Redefining Information Quality & its Measuring: The Operations Management Approach. In ICIQ (pp. 399-419).
- Hammer, M. & Champy, J., 2009. *Reengineering the Corporation: Manifesto for Business Revolution*, A. Zondervan.
- Hammer, M., 1990. Reengineering work: don't automate, obliterate. *Harvard business review*, 68(4), pp.104-112.
- Hernández, M.A. & Stolfo, S.J., 1998. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data mining and knowledge discovery*, 2(1), pp.9-37.
- Inmon, W. H. 1992. *Building the Data Warehouse*, New York, NY: Wiley.
- Isakowitz, T., Stohr, E.A. & Balasubramanian, P., 1995. A methodology for structured hypermedia design. *Communications of the ACM*, 38(8), pp.34-44.
- Isakowitz, T., Stohr, E.A. & Balasubramanian, P., 1995. A methodology for structured hypermedia design. *Communications of the ACM*, 38(8), pp.34-44.
- ISTAT. 2004. Guidelines for the data quality improvement of localization data in public administration (in Italian). Available at: [www.istat.it].
- Jarke, M., Gellersdörfer, R., Jeusfeld, M.A., Staudt, M. & Eherer, S., 1995. A deductive object base for meta data management. *Journal of Intelligent Information Systems*, 4(2), pp.167-192.
- Jarke, M., Lenzerini, M., Vassiliou, Y., & Vassilioudis, P., 1995. *Fundamentals of Data Warehouses*. Springer Verlag.
- Jeusfeld, M.A., Quix, C. & Jarke, M., 1998. Design and analysis of quality information for data warehouses. In International Conference on Conceptual Modeling (pp. 349-362). Springer, Berlin, Heidelberg.
- Kerr, K. & Norris, T., 2004. The Development of a Healthcare Data Quality Framework and Strategy, In IQ (pp. 218-233).

-
- Kettinger, W.J. & Grover, V., 1995. Toward a theory of business process change management. *Journal of Management Information Systems*, 12(1), pp.9-30.
- Kovac, R. & Weickert, C., 2002. Starting with Quality using TDQM in a Start-Up Organization. In *IQ* (pp. 69-78).
- Lee, Y.W., Strong, D.M., Kahn, B.K. & Wang, R.Y., 2002. A methodology for information quality assessment. *Information & management*, 40(2), pp.133-146.
- Lenzerini, M., 2002, June. Data integration, A theoretical perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 233-246). ACM.
- Lenzerini, M., 2002, June. Data integration: A theoretical perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 233-246). ACM.
- Long, J. & Seko, C. April 2005, A cyclic-hierarchical method for database data-quality evaluation and improvement. In *Advances in Management Information Systems-Information Quality Monograph (AMISIQ)*.
- Long, J. A., & Seko, C. E. 2005, A cyclic-hierarchical method for database data-quality evaluation and improvement. *Information quality*, 1, 52.
- Loshin, D., 2001. *Enterprise knowledge management: The data quality approach*. Morgan Kaufmann.
- Luisa De Biagi, Roberto Puccinelli, Massimiliano Saccone, Luciana Trufelli. 2014. *Research product repositories: strategies for data and metadata quality control*. National Research Council of Italy.
- Lyman, P. & Varian, H.R., 2003. How much information? SIMS. University of California at Berkeley, CA, US. Available at: [<http://www.sims.berkeley.edu/how-much-info-2003>].
- Madnick, S.E. & Wang, Y.Y.R., 1992. Introduction to the TDQM research program. Massachusetts Institute of Technology.
- Mecca, G., Atzeni, P., Masci, A., Sindoni, G. & Merialdo, P., 1998. The Araneus Web-based management system. In *ACM SIGMOD Record* (Vol. 27, No. 2, pp. 544-546). ACM.
- Mecca, G., Merialdo, P., Atzeni, P., Crescenzi, V. & Crescenzi, V., 1999. The (Short) Araneus Guide to Web-Site Development. In *WebDB (Informal Proceedings)* (pp. 13-18).
- Motro, A. & Anokhin, P., 2006. Resolution of data inconsistencies in the integration of heterogeneous information sources. *Information fusion*, 7(2), pp.176-196.
- Muthu, S., 1999. Business process reengineering: a consolidated methodology/Subramanian Muthu, Larry Whitman, and S. Hossein Cheraghi. In *Proceedings of the 4th Annual International Conference on Industrial Engineering Theory, Applications and Practice*.-November 17-20.
- Nadkarni, P., 2006. Delivering Data On Time: The Assurant Health Case. In *ICIQ* (pp. 341-355).

- Naumann, F., 2003. Quality-driven query answering for integrated information systems (Vol. 2261). Springer.
- Nelson, R.R., Todd, P.A. & Wixom, B.H., 2005. Antecedents of information and system quality: an empirical examination within the context of data warehousing. *Journal of management information systems*, 21(4), pp.199-235.
- Oakland, J.S., 1989. TQM-the new way to manage. In *Proceedings of the 2nd International Conference on Total Quality Management*. Springer. Butterworth-Heinemann.
- OMB, U., 2002. Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Part IX: Office of Management and Budget.
- Pernici, B. & Scannapieco, M., 2003. Data quality in web information systems. In *Journal on Data Semantics I* (pp. 48-68). Springer, Berlin, Heidelberg.
- Pipino, L.L., Lee, Y.W. & Wang, R.Y., 2002. Data quality assessment. *Communications of the ACM*, 45(4), pp.211-218.
- Raghunathan, S., 1999. Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. *Decision Support Systems*, 26(4), pp.275-286.
- Rahm, E., Thor, A., Aumueller, D., Do, H.H., Golovin, N. & Kirsten, T., 2005. iFuice-Information Fusion utilizing Instance Correspondences and Peer Mappings. In *WebDB* (pp. 7-12).
- Rao, R., 2003. From unstructured data to actionable intelligence. *IT professional*, 5(6), pp.29-35.
- Redman, T.C., 1996. *Data quality for the information age* (Artech House computer science library). Artech House. Boston.
- Redman, T.C., 1998. The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), pp.79-82.
- Scannapieco, M. & Catarci, T., 2002. Data quality under a computer science perspective. *Archivi & Computer*, 2, pp.1-15.
- Scannapieco, M., Pernici, B. & Pierce, E.M., 2002. Towards a Methodology for Quality Improvement Based on the IP-MAP Framework. In *IQ* (pp. 279-291).
- Scannapieco, M., Pernici, B. & Pierce, E.M., 2002. Towards a Methodology for Quality Improvement Based on the IP-MAP Framework. In *IQ* (pp. 279-291).
- Scannapieco, M., Virgillito, A., Marchetti, C., Mecella, M. & Baldoni, R., 2004. The DaQuinCIS architecture: a platform for exchanging and improving data quality in cooperative information systems. *Information systems*, 29(7), pp.551-582.
- Sessions, V., 2007, November. Employing the TDQM methodology: An assessment of the SC SOR. In *Proceedings of the 12th International Conference on Information Quality* (pp. 519-537).
- Shankaranarayanan, G. & Wang, R.Y., 2007, November. Current state and perspectives. In *Proceedings of the 12th International Conference on Information Quality* (pp. 510-517).

- Shankaranarayanan, G., Wang, R.Y. & Ziad, M., 2000. IP-MAP: Representing the Manufacture of an Information Product. *IQ*, 2000, pp.1-16.
- Sheng, Y.H. & Mykytyn Jr, P.P., 2002. Information Technology Investment and Firm Performance: A Perspective of Data Quality. In *IQ* (pp. 132-141).
- Sheng, Y.P., 2003. Exploring the Mediating and Moderating Effects of Information Quality on Firms, Endeavor on Information Systems. In *IQ* (pp. 344-353).
- Stoica, I., Morris, R., Liben-Nowell, D., Karger, D. R., Kaashoek, M. F., Dabek, F., & Balakrishnan, H. 2003, Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Transactions on Networking (TON)*, 11, 17-32.
- Storey, V.C. & Wang, Y.Y.R., 1994. Modeling quality requirements in conceptual database design. Total Data Quality Management Research Program, Sloan School of Management, Massachusetts Institute of Technology.
- Su, Z. & Jin, Z., 2007. A methodology for information quality assessment in the designing and manufacturing processes of mechanical products. In *Information Quality Management: Theory and Applications* (pp. 190-220). IGI Global.
- Vassiliadis, P., Vagena, Z., Skiadopoulos, S., Karayannidis, N. & Sellis, T., 2001. Towards the modeling, design, control and execution of ETL processes. *Information Systems*, 26(8), pp.537-561.
- Vermeer, B.H., 2000, January. How important is data quality for evaluating the impact of EDI on global supply chains?. In *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on* (pp. 10-pp). IEEE.
- Wand, Y. & Wang, R.Y., 1996. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), pp.86-95.
- Wang, R.Y. & Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), pp.5-33.
- Wang, R.Y. & Strong, D.M., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), pp.5-33.
- Wang, R.Y., 1998. A product perspective on total data quality management. *Communications of the ACM*, 41(2), pp.58-65.
- Wang, R.Y., Kon, H.B. & Madnick, S.E., 1993, April. Data quality requirements analysis and modeling. In *Data Engineering, 1993. Proceedings. Ninth International Conference on* (pp. 670-677). IEEE.
- Wang, Y.R. & Madnick, S.E., 1990. A polygen model for heterogeneous database systems: The source tagging perspective.
- Wayne, W.E., 2004. Data quality and the bottom line: achieving business success through a commitment to high quality data. The Data warehouse Institute (TDWI) report, Available at [www.dw-institute.com].
- Zachman, J.A., 2006. The Zachman Institute for Framework Advancement. Available at: [http://www.zifa.com], [Accessed March 7, 2017].

نمایه

۲۸، ۳۲، ۱۰۲	دسترس پذیری	
۲۸، ۲۹، ۳۲، ۶۸، ۷۰، ۱۴۵، ۱۴۸، ۱۵۸	ذاتی	ابعاد کیفیت
۲۸، ۲۹، ۷۰، ۸۲، ۱۰۵	متنی	داده
۲۸، ۲۹، ۳۲	نمایشگری	
۴۲، ۴۵، ۶۱، ۶۳، ۶۴، ۶۶، ۸۰، ۸۲، ۹۲، ۹۴، ۹۸، ۱۳۷	داده محور	استراتژی
۴۲، ۴۴، ۴۵، ۶۱، ۶۲، ۶۳، ۶۴، ۶۷، ۸۰، ۹۴، ۹۸، ۱۳۷	فرایند محور	کیفیت داده
۱۲، ۲۶، ۲۷، ۲۸، ۳۰، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۳۹، ۴۰، ۴۱، ۴۳، ۴۴، ۴۷، ۴۹، ۵۲، ۵۴، ۵۶، ۶۱، ۶۴، ۶۸، ۷۰، ۷۲، ۷۳، ۷۶، ۷۷، ۸۱، ۸۲، ۸۸، ۹۱، ۹۲، ۹۴، ۹۸، ۱۰۰، ۱۰۳، ۱۰۵، ۱۰۶، ۱۰۹، ۱۱۰، ۱۱۱، ۱۱۲، ۱۱۳، ۱۱۴، ۱۱۵، ۱۱۶، ۱۱۷، ۱۱۹، ۱۲۰، ۱۲۱، ۱۲۲، ۱۲۳، ۱۲۴، ۱۲۵، ۱۲۶، ۱۲۷، ۱۲۸، ۱۲۹، ۱۳۰، ۱۳۲، ۱۳۳، ۱۳۴، ۱۳۶، ۱۳۷، ۱۳۸، ۱۴۰، ۱۴۲، ۱۴۵، ۱۴۶، ۱۴۸، ۱۴۹، ۱۵۰، ۱۵۱، ۱۵۲، ۱۵۳، ۱۵۴، ۱۵۷، ۱۵۸		اطلاعات
۱۵، ۱۶، ۱۸، ۶۵، ۱۴۶	آمار	
۵۵، ۶۴، ۹۰، ۱۴۲، ۱۴۳	جستجو	اطلاعات
۴۶، ۴۹، ۵۹، ۶۲، ۶۹، ۷۴، ۸۰، ۹۴، ۱۰۳، ۱۱۰، ۱۱۱، ۱۱۶، ۱۱۷، ۱۳۱، ۱۴۰	معیار	
۵۵، ۵۸، ۵۹، ۶۲، ۶۴، ۶۶، ۷۵، ۸۳، ۸۸، ۹۲، ۱۰۹، ۱۱۰، ۱۱۱، ۱۴۰		انبار داده

۲۶، ۵۹، ۷۰، ۷۳، ۸۸، ۹۴، ۱۰۳، ۱۰۸، ۱۱۶، ۱۳۱، ۱۴۵	اجرا	چرخه بهبود مستمر
۱۴، ۱۸، ۲۳، ۲۵، ۲۶، ۱۴۵	اقدام	
۱۴، ۲۶، ۹۶، ۱۲۱، ۱۴۵، ۱۵۸	برنامه‌ریزی	
۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۶، ۲۷، ۲۸، ۳۳، ۳۴، ۴۱، ۴۲، ۴۴، ۴۵، ۶۰، ۶۱، ۶۳، ۶۷، ۶۸، ۸۸، ۹۶، ۱۲۴، ۱۲۷، ۱۳۹، ۱۴۵	کنترل	
۲۷، ۲۸، ۳۴، ۳۸، ۴۲، ۴۳، ۵۶، ۹۰، ۱۴۱، ۱۴۶	فناوری	کیفیت دسترس‌پذیری داده
۲۶، ۳۱، ۷۰، ۷۱، ۷۳، ۹۷، ۱۴۶، ۱۵۷	امنیت	
۲۶، ۳۶، ۴۳، ۴۸، ۵۵، ۵۶، ۶۹، ۷۰، ۷۱، ۷۳، ۹۰، ۱۰۲، ۱۱۵، ۱۴۶، ۱۵۷، ۱۵۸	دسترسی	
۲۹، ۷۰، ۷۱، ۱۴۵، ۱۴۶، ۱۴۸	باورپذیری	
۲۹، ۵۳، ۶۱، ۷۰، ۸۳	تکرار	کیفیت ذاتی داده
۱۵، ۲۹، ۴۸، ۵۶، ۶۹، ۷۱، ۷۳، ۸۳، ۸۴، ۹۴، ۹۶، ۹۸، ۱۰۲، ۱۱۵، ۱۴۵، ۱۴۷، ۱۴۸	دقت	
۲۹، ۱۴۵، ۱۴۶، ۱۴۸	عینیت	
۲۹، ۱۰۳، ۱۴۸	معتبر بودن	
۷۰، ۷۱، ۱۴۶، ۱۵۲، ۱۵۳، ۱۵۴	ارزش افزوده	کیفیت محتوای داده
۱۵۲	به‌روز بودن	
۲۸، ۱۰۱، ۱۰۲، ۱۴۶، ۱۵۱، ۱۵۲	کامل بودن	
۲۹، ۷۰، ۷۱، ۷۳	مربوط بودن	
۲۹	مقدار داده‌ها	کیفیت نمایشگری داده
۷۰، ۱۴۶، ۱۵۵	تفسیرپذیری	
۲۸، ۵۰، ۶۲، ۷۲، ۱۰۰، ۱۰۲	سازگاری	
۲۹، ۱۵۵، ۱۵۶	سهولت درک	
۳۸، ۱۰۵	مختصر	

۱۸، ۲۲، ۲۴، ۲۵، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۳۷، ۳۸، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴، ۴۶، ۴۹، ۵۰، ۵۱، ۵۲، ۵۳، ۵۴، ۵۵، ۵۶، ۵۷، ۵۸، ۵۹، ۶۰، ۶۲، ۶۴، ۶۵، ۶۶، ۶۸، ۷۱، ۷۵، ۷۶، ۷۷، ۷۸، ۷۹، ۸۰، ۸۳، ۸۴، ۸۵، ۸۷، ۹۰، ۹۲، ۹۳، ۹۴، ۹۵، ۹۶، ۹۸، ۱۰۱، ۱۰۲، ۱۰۴، ۱۰۵، ۱۰۸، ۱۱۰، ۱۱۲، ۱۱۵، ۱۱۷، ۱۱۸، ۱۲۱، ۱۲۳، ۱۲۵، ۱۲۶، ۱۲۸، ۱۲۹، ۱۳۲، ۱۳۴، ۱۳۶، ۱۹۸، ۱۴۰، ۱۴۱، ۱۴۷، ۱۴۸، ۱۵۵، ۱۵۹، ۱۶۰	داده
۳۲، ۶۹، ۱۲۰، ۱۲۱، ۱۳۱، ۱۳۲، ۱۴۶	شاخص
۳۰، ۷۱، ۱۴۲، ۱۴۵، ۱۵۸	فراداده
۱۴، ۱۵، ۱۶، ۱۷، ۱۸، ۱۹، ۲۰، ۲۱، ۲۲، ۲۳، ۲۴، ۲۵، ۲۶، ۲۷، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵، ۳۶، ۴۰، ۴۱، ۴۲، ۴۳، ۴۴، ۴۵، ۴۶، ۴۷، ۴۸، ۴۹، ۵۰، ۵۱، ۵۲، ۵۳، ۵۴، ۵۵، ۵۶، ۵۷، ۵۸، ۵۹، ۶۱، ۶۸، ۶۹، ۷۰، ۷۱، ۷۲، ۷۳، ۷۴، ۷۵، ۷۷، ۷۸، ۷۹، ۸۰، ۸۱، ۸۲، ۸۳، ۸۴، ۸۵، ۸۶، ۸۷، ۸۸، ۹۰، ۹۱، ۹۲، ۹۳، ۹۴، ۹۶، ۹۸، ۹۹، ۱۰۰، ۱۰۱، ۱۰۲، ۱۰۴، ۱۰۵، ۱۰۶، ۱۰۷، ۱۰۸، ۱۰۹، ۱۱۰، ۱۱۴، ۱۱۵، ۱۱۶، ۱۱۷، ۱۱۸، ۱۱۹، ۱۲۰، ۱۲۱، ۱۲۲، ۱۲۳، ۱۲۴، ۱۲۵، ۱۲۶، ۱۲۷، ۱۲۸، ۱۲۹، ۱۳۰، ۱۳۱، ۱۳۲، ۱۳۳، ۱۳۴، ۱۳۵، ۱۳۶، ۱۳۷، ۱۳۸، ۱۳۹، ۱۴۰، ۱۴۱، ۱۴۲، ۱۴۵، ۱۴۸، ۱۴۹، ۱۵۲، ۱۵۴، ۱۵۸، ۵۸، ۶۱، ۶۲، ۶۹، ۸۴، ۹۱، ۹۷، ۱۲۷، ۱۳۴، ۱۳۶، ۱۴۰	مدیریت
۵۸، ۶۱، ۶۲، ۶۹، ۸۴، ۹۱، ۹۷، ۱۲۷، ۱۳۴، ۱۳۶، ۱۴۰	کیفیت
مالی	

واژه‌نامه فارسی به انگلیسی

Statistical	آمارى
Dimensions	ابعاد
Value	ارزش
Evaluating	ارزشیابی
Assessment	ارزیابی
Execution	اجرا
Integration	ادغام
Relational	ارتباطی
Added	افزوده
Security	امنیت
Data warehouse	انبار داده
Warehousing	انبارداری
Do	انجام
Measurement	اندازه‌گیری
Objectivity	اهمیت
Standard	استاندارد
Standardization	استانداردسازی
Fishbone	استخوان ماهی
Strategy	استراتژی
Information	اطلاعات
Reputation	اعتبار
Economic	اقتصادی
Act	اقدام
Electronic	الکترونیکی

Port	انتقال برنامه
Representational	بازنمایی
Database	بانک اطلاعاتی
Believability	باورپذیری
Unstructured	بدون ساختار
Schemaless	بدون طرح
Met	برآورد شده
Check	بررسی
Return	برگشت
Inter-Relation	برون ارتباطی
Timeliness	به موقع بودن
Budget	بودجه
Localization	بومی سازی
International	بین‌المللی
Linkage	پیوند
Info Impact	تأثیر اطلاعات
Interchange	تبادل
Analysis	تحلیل و بررسی
Cooperative	تعاونی
Completeness	تکمیل
Capability	توانایی
Distributed	توزیع شده
Development	توسعه
Production	تولید
Product	تولید، محصول
Consistency	ثبات
Comprehensive	جامع
Look-Up	جستجو
Session	جلسه
Total	جمع
Chart	چارت
Cycle	چرخه
Null	خالی

Creative	خلاق یا مبتکر
Self-Describing	خود توصیفی
Self-Organization	خود سازمان‌ده
Automotive	خودرو
Data	داده
Data-Driven	داده محور
Domain	دامنه
Processing	در حال پردازش
Intra-Relation	درون ارتباطی
Accessibility	دسترسی‌پذیری
Access	دسترسی
Defense	دفاع
Accuracy	دقت
Inherent	ذاتی
Intrinsic	ذاتی
Subjective	ذهنی
Media	رسانه
Record	رکورد
Currency	رواج
Method	روش
Methodology	روش‌شناسی
Process	روند
Approach	رویکرد
Language	زبان
Manufacturing	ساخت
Structured	ساختار
Investment	سرمایه‌گذاری
Health	سلامت
Criteria	شاخص
Metric	شاخص، متریک
Enterprise	شرکت
Manufacturer	شرکت تولید کننده
Gap	شکاف

Object-Oriented	شیء‌گرا
Industrial	صنعتی
Required	ضروری
Design	طرح
Plan	طرح
Schema	طرح
Brainstorming	طوفان مغزی
Phrase	عبارت
Consensus	عمومیت
Objective	عینی
Metadata	فراداده
Process-Driven	فرایند محور
Technology	فن آوری
Maintainability	قابلیت نگهداری
Power	قدرت
Users	کاربران
Application	کاربرد
Pragmatic	کاربردی
Scorecard	کارت امتیازی
Broker	کارگزار
Canadian	کانادایی
Dirty	کثیف
Business	کسب و کار
Quantification	کمی‌سازی
Control	کنترل
Quality	کیفیت
Missing	گمشده
Matrix	ماتریکس
Financial	مالی
Activity-Based	مبتنی بر فعالیت
Unified	متحد
Variability	متغیر بودن
Contextual	متنی

Constraints	محدودیت‌ها
Manufactured Product	محصول تولیدی
Benchmark	محک
Environment	محیط
Conciseness	مختصر
Repository	مخزن
Modeling	مدل‌سازی
Management	مدیریت
Related	مربوط
Sorted	مرتب شده
Relevancy	مرتبط بودن
Independent	مستقل
Documentary	مستند
Characteristics	مشخصات
Transaction	معامله
Disabled	معلول
Semantic	معنایی
National	ملی
Resource	منابع
Applicable	مناسب
Reengineering	مهندسی مجدد
Material	مواد
Institute	مؤسسه
Unknown	ناشناس
Survey	نظرسنجی، تحلیل
Map	نقشه
Diagram	نمودار
Entity	نهاد
Volatility	نوسان
Man Power	نیروی انسانی
Semi-Structured	نیمه ساختاری
Goal	هدف
Cost-Effect	هزینه-اثر

Peer	همتا
Neighborhood	همسایه
Cooperation	همکاری
Dependent	وابستگی
Task	وظیفه
Edit	ویرایش
Attribute	ویژگی
Monolithic	یکپارچه

واژه‌نامه انگلیسی به فارسی

Access	دسترسی
Accessibility	دسترسی پذیری
Accuracy	دقت
Act	اقدام
Activity-Based	مبتنی بر فعالیت
Added	افزوده
Analysis	تحلیل و بررسی
Applicable	مناسب
Application	کاربرد
Approach	رویکرد
Assessment	ارزیابی
Attribute	ویژگی
Automotive	خودرو
Believability	باورپذیری
Benchmark	محک
Brainstorming	طوفان مغزی
Broker	کارگزار
Budget	بودجه
Business	کسب و کار
Canadian	کانادایی
Capability	توانایی
Characteristics	مشخصات
Chart	چارت
Check	بررسی
Completeness	تکمیل

Comprehensive	جامع
Conciseness	مختصر
Consensus	عمومیت
Consistency	ثبات
Constraints	محدودیت‌ها
Contextual	متنی
Control	کنترل
Cooperation	همکاری
Cooperative	تعاونی
Cost-Effect	هزینه-اثر
Creative	خلاق یا مبتکر
Criteria	شاخص
Currency	رواج
Cycle	چرخه
Data	داده
Data warehouse	انبار داده
Database	بانک اطلاعاتی
Data-Driven	داده محور
Defense	دفاع
Dependent	وابستگی
Design	طرح
Development	توسعه
Diagram	نمودار
Dimensions	ابعاد
Dirty	کثیف
Disabled	معلول
Distributed	توزیع شده
Do	انجام
Documentary	مستند
Domain	دامنه
Economic	اقتصادی
Edit	ویرایش
Electronic	الکترونیکی

Enterprise	شرکت
Entity	نهاد
Environment	محیط
Evaluating	ارزشیابی
Execution	اجرا
Financial	مالی
Fishbone	استخوان ماهی
Gap	شکاف
Goal	هدف
Health	سلامت
Independent	مستقل
Industrial	صنعتی
Info Impact	تأثیر اطلاعات
Information	اطلاعات
Inherent	ذاتی
Institute	مؤسسه
Integration	ادغام
Interchange	تبادل
International	بین‌المللی
Inter-Relation	برون ارتباطی
Intra-Relation	درون ارتباطی
Intrinsic	ذاتی
Investment	سرمایه گذاری
Language	زبان
Linkage	پیوند
Localization	بومی سازی
Look-Up	جستجو
Maintainability	قابلیت نگهداری
Man Power	نیروی انسانی
Management	مدیریت
Manufactured Product	محصول تولیدی
Manufacturer	شرکت تولید کننده
Manufacturing	ساخت

Map	نقشه
Material	مواد
Matrix	ماتریکس
Measurement	اندازه‌گیری
Media	رسانه
Met	برآورد شده
Metadata	فرا داده
Method	روش
Methodology	روش‌شناسی
Metric	شاخص، متریک
Missing	گمشده
Modeling	مدل‌سازی
Monolithic	یکپارچه
National	ملی
Neighborhood	همسایه
Null	خالی
Objective	عینی
Objectivity	اهمیت
Object-Oriented	شیء‌گرا
Peer	همتا
Phrase	عبارت
Plan	طرح
Port	انتقال برنامه
Power	قدرت
Pragmatic	کاربردی
Process	روند
Process-Driven	فرایند محور
Processing	در حال پردازش
Product	تولید، محصول
Production	تولید
Quality	کیفیت
Quantification	کمی‌سازی
Record	رکورد

Reengineering	مهندسی مجدد
Related	مربوط
Relational	ارتباطی
Relevancy	مرتبط بودن
Repository	مخزن
Representational	بازنمایی
Reputation	اعتبار
Required	ضروری
Resource	منابع
Return	برگشت
Schema	طرح
Schemaless	بدون طرح
Scorecard	کارت امتیازی
Security	امنیت
Self-Describing	خود توصیفی
Self-Organization	خود سازمان‌ده
Semantic	معنایی
Semi-Structured	نیمه ساختاری
Session	جلسه
Sorted	مرتب شده
Standard	استاندارد
Standardization	استانداردسازی
Statistical	آماري
Strategy	استراتژی
Structured	ساختار
Subjective	ذهنی
Survey	نظرسنجی، تحلیل
Task	وظیفه
Technology	فن آوری
Timeliness	به موقع بودن
Total	جمع
Transaction	معامله
Unified	متحد

Unknown	ناشناس
Unstructured	بدون ساختار
Users	کاربران
Value	ارزش
Variability	متغیر بودن
Volatility	نوسان
Warehousing	انبارداری

Models and Techniques for Data Quality Management: A Practical Approach on Science and Technology Data

By:

Mohammad Javad Ershadi

Mohammad Mahdi Ershadi

Edited by:

Hamid Keshavarz



**Iranian Research Institute
for Information Science and Technology
(I r a n D o c)**



**Chapar
Publication**

2019