

کیفیت داده‌ها پیش‌نیاز مدیریت منابع سازمان

عباس محبوب
حمیدرضا سیمه‌ساز

چکیده:

تاریخ دریافت: ۹۴/۰۶/۲۳
تاریخ پذیرش: ۹۴/۰۹/۲۴

سازمان‌ها می‌توانند با مدیریت و برنامه‌ریزی منابع خود به بهره‌وری رسیده و هزینه‌ها را کاهش دهند. منابع اطلاعاتی به‌عنوان یکی از مهم‌ترین منابع سازمانی باید بتوانند نیازهای سازمان را در برنامه‌ریزی‌ها و تصمیم‌سازی‌ها برآورده کنند. کیفیت داده‌ها مبحثی است که به شیوه‌های بهبود و مدیریت منابع سازمان و منابع اطلاعاتی مرتبط با آن‌ها می‌پردازد. استاندارد ISO 8000 با موضوع کیفیت داده‌ها و استاندارد ISO 22745 با موضوع یکپارچگی و سیستم‌های اتوماسیون صنعتی در این مقاله مورد بحث واقع می‌شود.

واژگان کلیدی:

داده‌های اصلی، مدیریت کیفیت داده‌های اصلی، کدینگ انبار، بهبود کیفیت داده‌ها، پاک‌سازی داده‌ها، الگوی شناسایی اقلام، کاتالوگ اقلام.

۱. مقدمه

بیشتر سازمان‌ها به دنبال کسب سود بیشتر و کاهش هزینه‌ها هستند. سود بیشتر و کاهش هزینه‌ها همواره با موضوع کیفیت در سازمان آمیخته است. اگر سازمان از مواد اولیه‌ای با کیفیت استفاده کند و توسط نیروی انسانی متخصص و ابزارآلات و تجهیزات با کیفیت به بهترین روش^۱، فرایندهای سازمانی را به انجام برساند به محصولی با کیفیت دست خواهد یافت. کیفیت، مفهومی است که در همه‌ی اجزای سازمان نقش پررنگ و مهمی دارد. افزون بر آن که بخش مهندسی سازمان باید از اجزای با کیفیت تشکیل شده باشد، بخش‌های پشتیبانی هم نیاز به کیفی بودن اجزای خود دارند. تمامی این موارد، سازمان‌ها را بر آن داشته است تا از ابزار کیفیت داده‌ها به‌عنوان ابزاری که صحت، دقت و سرعت را در اجرای فرایندهای سازمانی دخیل می‌کنند، سود ببرند.

داده‌ها به‌عنوان موجودیت اصلی در بحث فناوری اطلاعات در ابزار کیفیت داده‌ها استفاده می‌شوند. امروزه تمام سازمان‌ها برای اجرای فرایندهایشان از داده‌ها استفاده می‌کنند. داده‌های سازمانی یا توسط خود سازمان تولید شده‌اند یا سازمان‌های دیگری آن‌ها را ایجاد کرده و در اختیارشان قرار داده‌اند. در

چرخه حیات محصول (کالاها و خدمات)، داده‌های محصول با آن زاده می‌شوند و پس از مرگ محصول همچنان باقی می‌مانند. داده‌ها و اطلاعات محصول در طول زنجیره‌ی تأمین حرکت کرده و نقش اساسی در تصمیم‌گیری‌ها و عملکرد فرایندهای سازمانی دارند. به‌همین دلیل داده‌ها، بخشی از منابع سازمان و شاید مهم‌ترین منبع سازمان‌ها محسوب می‌شوند. کیفیت داده‌های سازمانی، اصلی مهم برای رسیدن به بهره‌وری سازمان‌هاست.

۲- داده‌ها و اطلاعات

امروزه داده‌ها و اطلاعات مربوط به موجودیت‌ها، ارزشی اگر نگوییم بیشتر، برابر با خود موجودیت‌ها یافته‌اند. داده‌ها به‌عنوان بخشی از دارایی‌های سازمان به‌حساب آمده و مبنایی برای تصمیم‌گیری‌های سازمانی هستند. با سازماندهی و مدیریت اطلاعات می‌توان تصمیمات صحیحی در جهت استفاده بهینه از موجودیت‌های سازمانی گرفت. سازمان‌ها علاوه بر گردآوری داده‌ها و اطلاعات به پردازش و فرآوری داده‌ها نیز اقدام می‌کنند. امروزه داده‌ها اغلب با سیستم‌های اطلاعاتی مرتبط هستند و طی انجام فرایندهای کاری که به‌وسیله سیستم‌های اطلاعاتی مدیریت می‌شوند، ایجاد می‌گردند. سودآوری بیشتر و پرهیز از هزینه‌های

غیرضروری یکی از اهداف اصلی سازمان‌های تولیدکننده است و این مهم محقق نمی‌شود مگر با تولید محصولات با کیفیت که با استفاده از سیستم‌های کارا و برنامه‌ریزی‌های صحیح قابل دستیابی هستند. لازمه‌ی تحقق تمام این موارد داشتن داده‌های با کیفیت است. مدیریت داده‌ها با استفاده از ابزارهای مرتبط با فناوری، پیشرفت چشمگیری داشته است؛ اما کیفیت داده‌ها به‌عنوان بخش مهمی از مقوله مدیریت داده‌ها در سازمان‌ها زیاد مورد توجه قرار نگرفته است. سازمان‌ها برای مدیریت داده‌ها هزینه می‌کنند؛ اما تمرکز اصلی آن‌ها بر ابزارهای مبتنی بر فناوری است و مدیریت محتوای داده به‌عنوان عامل مؤثر در داشتن داده‌های با کیفیت، مقوله‌ای است که باید توسط سازمان‌ها بیشتر مورد توجه قرار گیرد.

۳. ابزارهای مدیریت داده‌ها

مدیریت داده‌ها نیازمند ابزارهای سخت‌افزاری و نرم‌افزاری مبتنی بر فناوری است. عملیات گردآوری، پردازش، پشتیبان‌گیری، بازیابی، تبادل و نگهداری از داده‌ها همگی نمودهایی از مدیریت داده‌ها هستند که با به‌کارگیری ابزارهای سخت‌افزاری و نرم‌افزاری به انجام می‌رسند. نرم‌افزارهای جامع سازمانی، سیستم‌های رایانه‌ای، زیرساخت‌های انتقال داده‌ها، تجهیزات و نرم‌افزارهای پشتیبان‌گیری و بازیابی اطلاعات نمونه‌هایی از ابزارهای مدیریت داده‌ها هستند. سازمان‌ها با صرف هزینه‌های هنگفت نسبت به تأمین ابزارهای مدیریت داده‌ها اقدام می‌کنند؛ اما بررسی و تحلیل داده‌های آن‌ها که در اثر استفاده از ابزارهایی مانند پایگاه‌های داده و نرم‌افزارهای کاربردی بسیار پر حجم هم شده‌اند نشان می‌دهد که دارای کیفیت پایینی هستند. کیفیت پایین داده‌های سازمانی باعث ناکارآمدی ابزارها و روش‌های مدیریت داده‌ها می‌شود و درنهایت علاوه بر اینکه منجر به تولید محصولات بی‌کیفیت می‌شود هزینه‌های بسیاری را هم به سازمان تحمیل می‌کند. همان‌طور که گفته شد مدیریت محتوای داده‌ها مهم‌ترین بخش مدیریت داده‌ها برای دستیابی به داده‌های با کیفیت است. داشتن روش و ابزارهای کارآمد برای مدیریت محتوای داده‌ها می‌تواند سازمان‌ها را برای رسیدن به داده‌های با کیفیت امیدوار کند.

۴. کیفیت داده‌ها

کیفیت داده‌ها^۱ با مورد استفاده‌ی آن داده‌ها در فرایندها و تصمیم‌گیری‌ها ارتباط دارد. داده‌ای با کیفیت در نظر گرفته می‌شود که برای استفاده در فرایندها و تصمیم‌گیری‌ها مناسب باشد و الزامات مورد نیاز برای استفاده‌ی مورد نظر را برآورده کند. برای سنجش کیفیت داده‌ها معیارهایی تعریف شده که برخی از مهم‌ترین آن‌ها در پی آمده است.

۴-۱ معیارهای کیفیت داده‌ها

۴-۱-۱ - صحت^۲

صحت داده‌ها به معنای آن است که داده‌ها در حوزه عملیاتی مورد نظر صحیح بوده و به‌درستی نمایش داده می‌شوند. شکل (۱) نشان‌دهنده‌ی ویژگی‌های داده‌های صحیح است.

داده‌هایی از بین ارزش‌ها یا مقادیر معتبر		داده‌هایی از بین ارزش‌ها یا مقادیر نامعتبر	
مقادیر منطبق با واقع و درست	مقادیر نامنتطبق با واقع و نادرست	مقادیر نامعتبر	مقادیر نامعتبر
نمایش درست	مقدار نادرست	مقدار نامعتبر	مقدار ندارد

شکل ۱: داده‌های صحیح

1. data quality
2. accuracy

این شکل نشان می‌دهد که داده‌های صحیح علاوه بر آنکه باید مجاز و معتبر باشند، باید به درستی و بدون ابهام نمایش داده شوند.

فرض کنید از شما پرسیده‌اند که ارتفاع میز کارتان را ارائه کنید. شما می‌توانید سه پاسخ بدهید: اول اینکه "نمی‌دانم" یعنی ارتفاع میز را اندازه‌گیری نکرده‌ام و حدود آن را هم نمی‌دانم. به این جواب، داده‌های مفقود یا نامشخص^۱ می‌گویند که به لحاظ معیار صحت داده‌ها بدترین وضعیت ممکن است. پاسخ دوم شما می‌تواند از میان مقادیر نامعتبر^۲ باشد؛ مثلاً بگویید ارتفاع میز کار "کوتاه" است. حالت سوم این است که جواب شما از میان مقادیر معتبر^۳ برای ارتفاع میز کار باشد؛ مثلاً بگویید ۷۰ سانتی‌متر. از میان سه حالت فوق کاملاً واضح است که فقط داده‌هایی که از دسته سوم یعنی مقادیر معتبر باشند، داده‌های صحیحی هستند. در میان داده‌ها یا مقادیر معتبر، می‌توان دو نوع پاسخ را از هم جدا کرد. ارزش‌های درست و ارزش‌های نادرست. مثلاً در پاسخ سؤال ارتفاع میز کار، اگر مقدار اعلام‌شده با ارتفاع واقعی میز کار برابر باشد، این داده، یک داده‌ی درست است و اگر مقدار اعلام‌شده با ارتفاع واقعی میز کار برابر نباشد، داده‌ی نادرست خواهد بود. واضح است که فقط داده‌هایی صحت دارند و این معیار کیفی برای آن‌ها رعایت‌شده است که با واقعیت منطبق باشند. برای رسیدن به صحت، باید داده‌های مفقود را یافته و ثبت کرد، داده‌های بی‌اعتبار را حذف نموده و مقادیر نادرست را اصلاح کرد.

دقت، پارامتر دیگری است که در اینجا می‌تواند مورد توجه قرار گیرد. برای مثال اگر پاسخ با یکصدم دقت مورد نظر باشد مقدار ۷۰.۰۰ سانتی‌متر دقیق‌تر از ۷۰ سانتی‌متر است.

۴-۱-۲ کامل بودن^۴

کامل بودن به معنای آن است که تا چه اندازه ویژگی‌های مورد انتظار داده فراهم‌شده‌اند.

برای مثال داده‌های مربوط به مشتری، زمانی به‌عنوان کامل در نظر گرفته می‌شود که:

• تمامی نشانی‌ها و جزئیات تماس و سایر اطلاعات

مشتری در دسترس باشد.

• داده‌های مربوط به همه‌ی مشتریان در دسترس باشد.

کامل بودن، بستگی به انتظاری که از آن داده‌ها می‌رود، دارد. یعنی ممکن است داده‌ها در دسترس نباشند؛ اما هنوز به‌عنوان داده‌های کامل در نظر گرفته شوند؛ چون آن داده‌ها هنوز انتظارات کاربران را برآورده می‌کنند. الزام مربوط به داده‌ها دارای دو جنبه‌ی "الزامی" و "اختیاری" است. برای مثال نشانی پست الکترونیکی مشتریان، الزامی و در دسترس است؛ اما نشانی دفاتر مشتریان چون اختیاری است در صورت در دسترس نبودن همچنان داده‌ها کامل هستند.

ممکن است داده‌ها کامل باشند؛ اما صحیح نباشند. برای مثال نشانی تمام مشتریان در دسترس است؛ اما برخی از آن‌ها نادرست هستند.

۴-۱-۳ سازگاری^۵

سازگاری داده‌ها به این معناست که داده‌ها در سراسر سازمان باید با یکدیگر هماهنگ باشند. مثال‌هایی از داده‌های ناسازگار:

• یک کارگزار غیرفعال شده اما هنوز حساب مشتری مربوطه‌اش را فعال نگه داشته است.

• یک کارت اعتباری لغو و غیرفعال شده اما وضعیت صورت حساب مشتری مربوط به آن کارت "قابل پرداخت" نشان داده می‌شود.

داده‌ها می‌توانند صحیح اما ناسازگار باشند. داده‌ها ممکن است در بخشی از سازمان هماهنگ باشند؛ اما در سراسر سازمان هماهنگ نباشند. در این صورت داده‌ها ناسازگار هستند.

داده‌ها می‌توانند کامل اما ناسازگار باشند.

۴-۱-۴ به‌هنگام بودن^۶

به‌هنگام بودن داده‌ها معیار دیگری برای سنجش کیفیت داده‌هاست. تأخیر در ارائه‌ی داده‌ها برابر با انکار داده‌هاست. به‌هنگام بودن داده‌ها بسیار مهم است. برای نمونه اگر داده‌های مربوط به کالاهای خریداری‌شده در پایان روز کاری ثبت شوند آنگاه این داده‌ها برای ثبت سفارش کالا از انبار در طی روز از کیفیت مناسب برخوردار نیستند.

1. missing value
2. invalid value
3. valid value
4. completeness
5. consistency
6. timeliness

۴-۱-۵ اصلت^۱

در دسترس بودن سوابق هرگونه ایجاد، اصلاح، گزارش، انتقال و همه تراکنش‌های قابل انجام روی داده‌ها به طوری که قابلیت ردیابی تا منشأ ایجاد داده‌ها وجود داشته باشد را اصلت داده می‌گویند. اصلت داده شاخصی برای سنجش کیفیت داده است. برای مثال اگر شماره بدنه‌ی خودرو به شماره‌ی قطعاتی که بدنه‌ی خودرو از آن‌ها ساخته شده است پیوسته باشد، داده‌ی مربوطه دارای اصلت است.

۵. بهبود کیفیت داده‌ها

همان‌طور که گفته شد، داده‌ها در تصمیم‌سازی و بهبود کیفیت محصولات سازمان‌ها نقش اساسی دارند. بنابراین اقدام برای ایجاد داده‌های باکیفیت و بهبود کیفیت داده‌های موجود باید به‌عنوان کاری بنیادین در جهت بهبود کیفیت محصولات و کاهش هزینه‌های سازمان مورد توجه ویژه قرار گیرد. برای رسیدن به این هدف، استانداردسازی ایجاد، نگهداری، پردازش و انتقال داده‌ها اهمیت به‌سزایی دارد. با توجه به اهمیت موضوع، سازمان‌های بین‌المللی مانند سازمان ناتو سعی نمودند تا با تدوین استانداردهایی نسبت به بهبود کیفیت داده‌های سازمانی اقدام نمایند. برآیند کوشش‌های چندین ساله در این مقوله منجر به تدوین استاندارد ISO 8000 با عنوان "کیفیت داده‌ها" شده است. استاندارد ISO 8000 به‌طور کلی به مقوله کیفیت داده‌ها می‌پردازد. سازمان ایزو همچنین استانداردهای دیگری در زمینه کیفیت داده‌ها تدوین کرده است که جنبه‌های عملیاتی در محیط‌های صنعتی در آن‌ها بیشتر مورد نظر بوده است. استاندارد ISO 22745 با عنوان "یکپارچه‌سازی و اتوماسیون سیستم‌های صنعتی - فرهنگ‌های فنی باز و کاربرد آن‌ها در داده‌های اصلی" و استاندارد ISO 29002 با عنوان "یکپارچه‌سازی و اتوماسیون سیستم‌های صنعتی - تبادل داده‌های مشخصاتی" دو استاندارد مرتبط با کیفیت داده‌ها هستند و الزامات پیاده‌سازی نرم‌افزارهای صنعتی مرتبط با داده‌های اصلی سازمان را تشریح می‌کنند. در ایران، سری استانداردهای IDS 306 توسط سازمان استاندارد دفاعی

ایران در این زمینه تدوین شده است.

۵-۱ استاندارد ISO 8000

اثرات و پیامدهای ناشی از "کیفیت داده‌ها" در انجام کارها، فعالیت‌های کسب‌وکار، فعالیت‌های دولتی و به‌طور کلی در جامعه دارای اهمیت است.

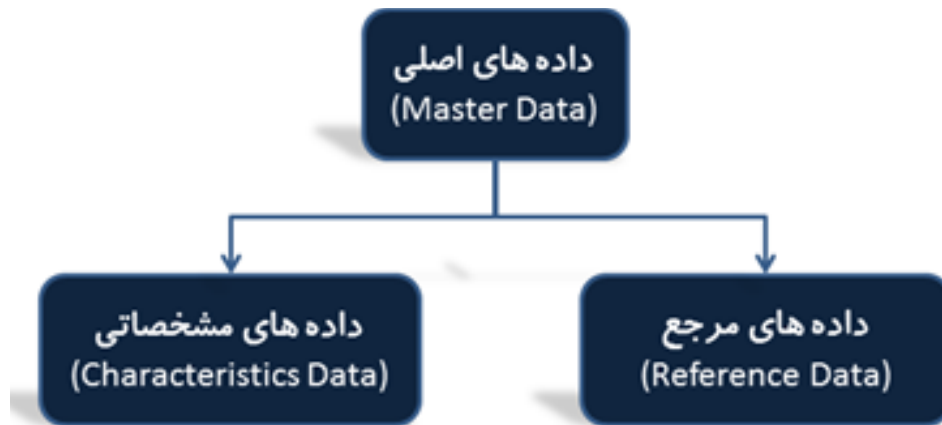
داده‌هایی با کیفیت پایین باعث نارضایتی مشتری، ازدست‌دادن عایدات و افزایش هزینه‌ها به‌دلیل صرف زمان بیشتر برای تطبیق داده‌ها می‌شوند. داده‌هایی با کیفیت پایین ممکن است باعث کاهش اعتبار یک سیستم و ریسک بالای عدم تطبیق با مقررات شود. داده‌های با کیفیت پایین باعث افزایش هزینه‌های مصرف‌کننده، افزایش مالیات، کاهش سود سرمایه‌گذاران و شکست مأموریت سازمان می‌شود. کیفیت داده‌ها وابسته به کیفیت فرایند گردآوری داده‌ها و فرایند ذخیره، نگهداری، انتقال و نمایش داده‌هاست.

استاندارد ISO 8000 به کیفیت داده‌ها می‌پردازد و پیرامون موارد زیر بحث می‌کند:

- اصول کیفیت داده‌ها،
 - مشخصاتی از داده‌ها که کیفیت آن‌ها را تعیین می‌کند،
 - فرایندهایی برای تضمین کیفیت داده‌ها.
- موارد زیر در دامنه‌ی بخش ۱۰۰ استاندارد ISO 8000 قرار دارند:
- داده‌های اصلی،
 - توصیف معماری کیفیت داده‌های اصلی.

۵-۱-۱ داده‌های اصلی

در هر سازمان، داده‌های اصلی برای شناسایی و توصیف چیزهایی که برای آن سازمان مهم هستند، به‌کار گرفته می‌شوند.



شکل ۲: داده‌های اصلی

و تحلیل مصارف و هزینه‌ها برای هزینه‌های نگهداری، تعمیر و عملیات^۱ (MRO) نیز به کار برده می‌شوند. داده‌های اصلی مواد اغلب در سیاهه‌های مواد^۲ (BOM) هم به کار گرفته می‌شوند، علاوه بر آن در طراحی‌ها، به‌عنوان کاتالوگ قطعات یا فهرست قطعات مورد استفاده هستند. یکی از گونه‌های داده‌های اصلی مواد، کاتالوگ مصور قطعات^۳ (IPC) یا فهرست قطعات یدکی است.

* داده‌های اصلی خدمات، رویه‌ها یا فرایندها: این داده‌های اصلی هنوز نسبتاً کمیاب هستند به‌جز در بخش بهداشت و درمان و صنایع تعمیرات خودروها، که صدور خودکار صورتحساب برای خدمات یا پرداخت حق بیمه متداول شده است. با این داده‌ها یک خدمت به‌صورت یک رویه یا فرایند به‌خوبی توصیف شده است.

* داده‌های اصلی دارایی‌های ثابت یا اموال: این داده‌های اصلی عموماً برای ردیابی اقلامی به‌کار می‌روند که قیمت خرید آن‌ها از مبلغ مشخصی بالاتر است یا هزینه آن‌ها طی چند سال مستهلک می‌شود. اموال معمولاً با یک شناسه‌ی یکتا مرتبط هستند (شماره سریال) و غالباً اقلام قابل جابه‌جایی هستند که گاهی لازم است مکان آن‌ها مورد بررسی و اعتبارسنجی قرار گرفته و گزارش شود. یک مسئله عمومی در مدیریت اموال این است که مشخصات اموال در طی دوره عمر آن‌ها تغییر می‌کند. تصمیم‌گیری در مورد زمان تغییر مشخصات یک قلم اموال و تعریف آن به‌عنوان نوع جدید معمولاً موضوعی چالش برانگیز است.

۱-۱-۱-۵ داده‌های مرجع

توسط مرجع آن‌ها برای استفاده در داده‌های اصلی دیگر سازمان‌ها تعیین می‌شوند.

۲-۱-۱-۵ داده‌های مشخصاتی

با توجه به مشخصات موجودیت‌هایی که آن‌ها را توصیف می‌کنند، تعیین می‌شوند.

در برنامه‌های کاتالوگ‌سازی، داده‌های اصلی برای توصیف چیزهایی مورد استفاده قرار می‌گیرند که آن‌ها را "اقلام" می‌نامند.

مثال‌هایی از داده‌های اصلی در اینجا ذکر شده است:

* داده‌های اصلی فروشندگان (تأمین کنندگان): این داده‌ها نوعاً یک فروشنده را با اشاره به مکان و وضعیت حقوقی‌اش توصیف می‌کنند. اغلب داده‌های ضروری برای توصیف یک فروشنده در قانون مشخص شده است، همچنان که یک الزام عمومی برای یک شرکت این است که بتواند همه موجودیت‌های مورد نیاز برای دادوستد را شناسایی کند.

* داده‌های اصلی مشتریان: این داده‌ها نوعاً به توصیف یک مشتری به‌عنوان موجودیتی تجاری می‌پردازند که باید کمینه شامل اطلاعات تماس مورد نیاز برای ارسال صورتحساب باشد و ممکن است شامل اطلاعات محرمانه، مانند اطلاعات کارت اعتباری باشد.

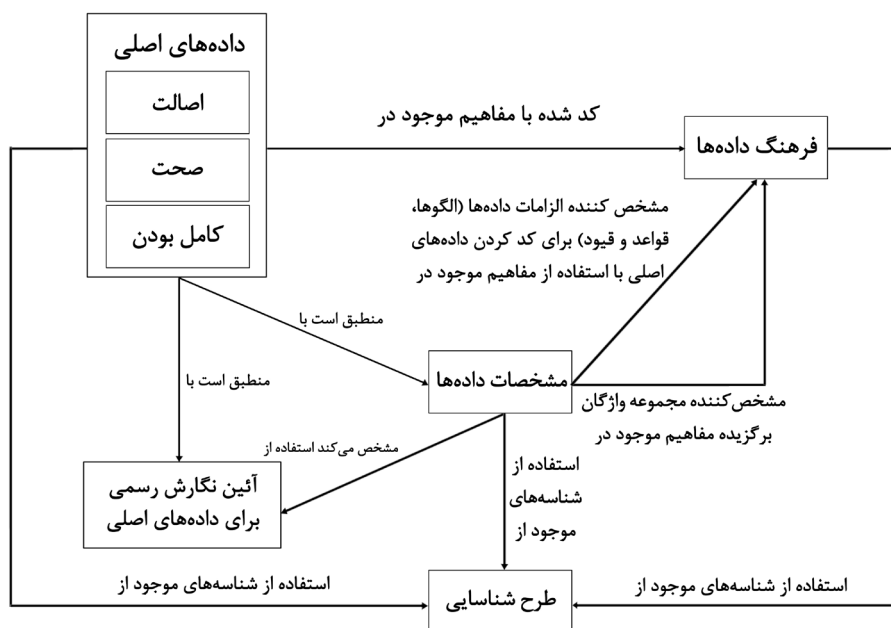
* داده‌های اصلی مواد یا اقلام: این داده‌ها نوعاً به توصیف اقلام فیزیکی که ردیابی و انبار می‌شوند یا به‌طور منظم خریداری می‌شوند، می‌پردازند. با وجود اینکه این اقلام اغلب محدود به موارد خاصی هستند که برای مثال اقلام خریداری شده تحت قرارداد مانند مواد تولیدی را می‌توان نام برد، اما برای بهبود تجزیه

1. maintenance and Repair Operations
2. Bill Of Materials
3. Illustrated Parts catalog

* داده‌های اصلی مکان‌ها: با وجود اینکه در خدمات تحویل به ندرت می‌بینیم که فهرست جداگانه‌ای از داده‌های اصلی مکان‌ها مورد استفاده قرار گرفته باشد، با این وجود فهرستی از داده‌های اصلی مکان فروشندگان و مشتریان می‌تواند سبب بهبود داده‌ها شود. مدل داده‌ها برای داده‌های اصلی مکان‌ها اساساً ساده بوده و یک مکان فیزیکی را با شماره مرجع سراسری (مانند کد پستی) معرفی می‌کند. در عمل شاید که توضیحات دیگری مانند یک آدرس پستی برای انجام خدمات تحویل نیاز باشد.

* داده‌های اصلی نقاط تماس، کارکنان یا منابع انسانی^۱ (HR): این داده‌های اصلی نوعاً یک فرد را توصیف می‌کنند. عموماً این داده‌ها شامل اطلاعات مرتبط با روابط سازمان استخدای کارکنان هستند؛ اما بهتر است به جای اینکه به‌عنوان داده‌های اصلی به‌کار بسته شوند، مانند داده‌های تراکنشی استفاده شوند.

۳-۱-۵ معماری کیفیت داده‌های اصلی



شکل ۳: مدل معماری کیفیت داده‌های اصلی در ISO 8000

درج مقدار مشخصه‌ی "اندازه‌ی قابل قبول پایه‌ی میز تحریر"، ابتدا توصیف‌گر اندازه‌ی "بیشینه"، سپس یک داده‌ی مقداری، سپس واحد اندازه‌ی "سانتی‌متر" آورده شود.

آیین نگارش رسمی داده‌های اصلی، مشخص کردن واژگان، کلمات و جملات معتبر از لحاظ نوشتاری و ترکیبی در یک زبان رسمی است.

مثال: در زبان رسمی فارسی، سند "دستور خط فارسی" مصوب فرهنگستان زبان و ادب فارسی ویرایش ۱۳۸۴، می‌تواند به‌عنوان آیین نگارش رسمی داده‌های اصلی قرار گیرد. برای نمونه ممکن است در این آیین نگارش الزام شده باشد که واحد اندازه‌ی "سانتی‌متر" به این صورت نوشته شود و به‌کاربردن "سانتیمتر" و

در معماری کیفیت داده‌های اصلی استاندارد ISO 8000، اصالت، صحت و کامل بودن سه معیار برای سنجش کیفیت داده‌های اصلی هستند.

فرهنگ داده‌ها، مجموعه‌ای از واژگان استانداردشده‌ی مربوط به مفاهیم است. این مفاهیم عناصر توصیف‌کننده‌ی داده‌های اصلی هستند.

مثال: واژگان "سانتی‌متر"، "طول"، "پایه‌ی میز تحریر"، "بیشینه" و تعاریف آن‌ها در فرهنگ داده‌ها هرکدام با کد یکتای منحصربه‌فرد ثبت شده‌اند. مشخصات داده‌ها، مشخص‌کننده‌ی الزاماتی برای کد کردن داده‌های اصلی^۴ با استفاده از عناصر فرهنگ مفاهیم است.

مثال: الزام‌شده است در زمان ایجاد کاتالوگ برای

1. Human Resources
2. data dictionary
3. data specification
4. formal syntax for master data

«CM» و غیره نادرست هستند.

طرح شناسایی، شیوه‌ی کدگذاری مفاهیم مختلف حوزه مسئله مانند عناصر فرهنگ داده‌ها، قالب‌های از پیش تعیین‌شده‌ی مشخصات داده‌ها و کاتالوگ‌های داده‌های اصلی را مشخص می‌کند.

۵-۲ استاندارد ISO/TS 22745

استاندارد ISO/TS 22745 سیستمی را برای فناوری توصیفی مشخص می‌کند که شامل موارد زیر است:

- فرهنگ فنی باز^۱،
- راهنماهای شناسایی^۲،
- داده‌های اصلی^۳،
- طرح شناسایی^۴،
- دستورالعمل‌هایی برای نگهداری از فرهنگ فنی باز،
- واسطه‌هایی برای پرس‌وجوی اطلاعات از فرهنگ فنی باز، شامل پرس‌وجوی عناصر فرهنگ فنی باز مرتبط با یک مفهوم مفروض.

۵-۲-۱ فرهنگ فنی باز

هر فرهنگ فنی باز، مجموعه مفاهیم تعریف‌شده‌ای است که از طریق اجماع بدنه‌های مختلف نظیر، "ایزو" و "آی ای سی" تشکیل شده و در حال توسعه هستند. هر فرهنگ فنی باز شامل اصطلاحات، تعاریف و تصاویر مفاهیم مورد استفاده برای توصیف افراد، سازمان‌ها، مکان‌ها، کالاها و خدمات است. مثال: "پیچ سرشش‌گوش"، "گام پیچ"، "قطر اسمی پیچ"، "بیشینه ۱۰ میلی‌متر"، "میلی‌متر" و "بیشینه" نمونه‌هایی از مفاهیم تعریف‌شده در فرهنگ فنی باز هستند.

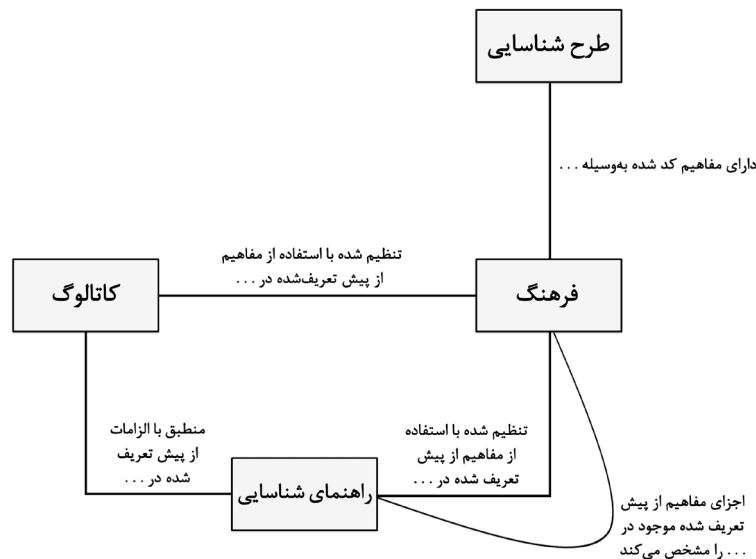
۵-۲-۲ راهنمای شناسایی

راهنمای شناسایی عبارت است از قالب و الزامات داده‌ها برای یک کاتالوگ به‌گونه‌ای که نیازهای یک دریافت‌کننده داده را برآورده کند.

۵-۲-۳ کاتالوگ

کاتالوگ شامل فهرستی از توصیفات اقلام است. در این استاندارد، هر قلم به وسیله ارتباطش با یک کلاس و مجموعه‌ای از زوج‌های خصیصه- ارزش توصیف می‌شود.

۵-۲-۴ معماری کیفیت داده‌ها



شکل ۴: مدل معماری کیفیت داده‌های اصلی در ISO/TS 22745

1. open technical dictionary
2. identification guide
3. master data
4. identification scheme

نکته: استاندارد ISO/TS 29002 استاندارد مکملی برای استاندارد ISO/TS 22745 است و در بسیاری موارد از استاندارد ISO/TS 22745 به استاندارد ISO/TS 29002 ارجاع داده شده است.

۵-۳ استاندارد IDS 306

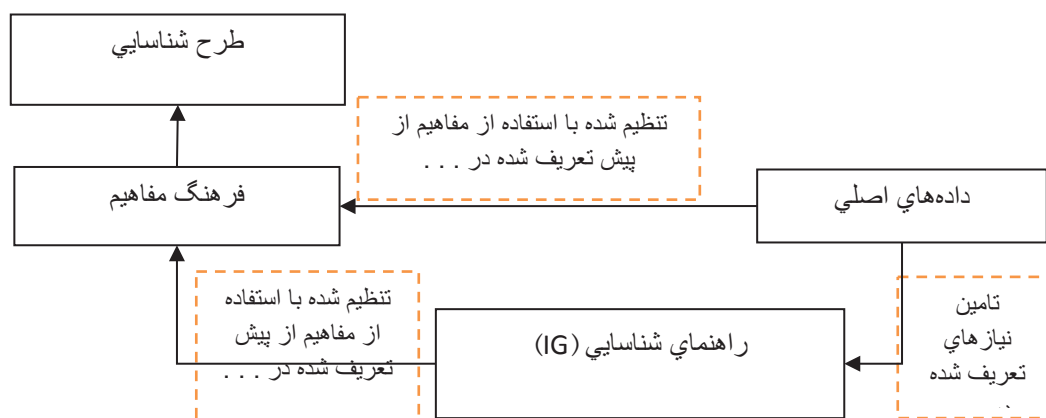
استاندارد ملی IDS 306 که توسط سازمان استاندارد دفاعی ایران تدوین شده از استانداردهای بین‌المللی، ISO 8000، ISO/TS 22745 و ISO/TS 29002 و آموزه‌های سازمان ناتو سرچشمه گرفته است.

۵-۳-۱ معماری داده‌ها

بانک اطلاعاتی محصول، مانند هر بانک اطلاعاتی دیگر حاوی اطلاعات مرتبط با موضوعی است که برای آن توسعه می‌یابد. اطلاعات مربوط به اقلام در این بانک ذخیره می‌شود. در حوزه نیروهای مسلح، منظور از اقلام و محصولات دفاعی، تمامی موجودیت‌هایی هستند که در این حوزه به‌عنوان محصول یا قلم تدارک‌شده^۱، استفاده شده و یا مورد معامله قرار می‌گیرند. هر قلم^۲ در ذهن کسانی که با آن سروکار خواهند داشت، تداعی‌کننده‌ی مفهومی مشخص خواهد بود. مفهوم، خود شامل اجزایی است که به‌وسیله آن اجزا شناسایی می‌شود. این اجزای اطلاعاتی از تداعی شدن ماهیت‌های مختلف در ذهن، تحت عنوان یک مفهوم جلوگیری می‌کنند. اجزای اصلی مفهوم عبارتند از: اصطلاح، تعریف و تصویر. این اجزا، مؤلفه‌های اصلی برای شناسایی مفهوم هستند. سایر اجزا عبارتند از: کومه نوشت، نماد نوشتاری و نماد تصویری. اگر اجزای مفهوم به درستی انتخاب و تعریف نشوند، شناسایی مفهوم دچار ابهام خواهد شد. در ارتباط با مفاهیم و اجزای اطلاعاتی مرتبط با آن‌ها، رعایت موارد زیر الزامی است:

- برای هر مفهوم یا جزو اطلاعاتی مرتبط با آن، یک شناسه^۳ کاراکتری تعیین شده (با توجه به طرح شناسایی) و به آن تخصیص داده شود.

- در سراسر حوزه نیروهای مسلح برای شناسایی اقلام از شناسه‌های تخصیص یافته به مفهوم استفاده شود.
- هر مفهوم باید دست‌کم دارای یک اصطلاح و یک تعریف باشد (داشتن تصویر برای این منظور الزامی نیست؛ اما برای شناسایی کامل‌تر مفهوم، استفاده از تصاویر توصیه می‌شود).
- همه‌ی طرف‌های درگیر با مفهوم که قصد شناسایی و به‌کارگیری اقلام را دارند، باید برای شناسایی اقلام مورد نظرشان از اجزای مفهوم موجود در فرهنگ مفاهیم استفاده کنند.
- مفاهیم و اجزای اطلاعاتی مرتبط با آن‌ها در یک پایگاه اطلاعاتی (فرهنگ مفاهیم) که سازمان مدرسکا، مسئول نگهداری آن است، درج شده و دسترسی به آن‌ها برای همه نهادهایی که با شناسه‌ها سروکار دارند امکان‌پذیر باشد.
- سازمان‌ها برای توسعه داده‌های اصلی و شناسایی محصولاتشان، باید از اطلاعات مفاهیم و اجزای آن، که در فرهنگ مفاهیم گردآوری شده‌اند، استفاده کنند.
- فرهنگ مفاهیم به صورت متمرکز مدیریت شده و توسعه یابد.



شکل ۵ _ مدل معماری کیفیت داده‌های اصلی در IDS 306-4

1. Item of Supply (IOS)
2. Item
3. Identifier

۶- پاک‌سازی و کیفی سازی داده ها

بی‌کیفیتی داده‌ها ریشه اصلی شکست‌ها و ناکامی‌های سازمان است. پاک‌سازی داده‌ها عامل کیفیت داده‌ها بوده و بر روی داده‌های اصلی تمرکز دارد. داده‌های اصلی داده‌هایی هستند که اشخاص، سازمان‌ها، مکان‌ها، کالاها، خدمات، فرایندها، رویه‌ها و قوانین و مقررات را تعریف و تشریح می‌کنند. تعریف متداول پاک‌سازی به طور نمونه تنها شامل حذف عناصر نامطلوب مانند چرک و خاک است اما در پاک‌سازی داده‌ها، شامل ساختاردهی به داده‌ها و همچنین افزودن داده‌های مفقود و از دست رفته است.

در هر کسب‌وکاری، فهرست اصلی فروشنده‌ها، خریداران، کارکنان، دارایی‌ها، کاتالوگ کالاها و خدمات فروخته شده و کاتالوگ مواد و خدمات خریداری شده را خواهید یافت. این رکوردهای داده‌های اصلی در پایگاه داده‌ها ذخیره شده و به‌طور معمول با یک کد مشخص می‌شوند. اگر این کدها شفاف و بی‌ابهام نباشند، شما با نرم‌افزارهای مدیریت کننده‌ی داده‌های اصلی مشکل پیدا می‌کنید.

۶-۱ چرا پاک‌سازی داده‌ها

پاک‌سازی داده‌ها زمانی که داده‌ها جوابگوی نیازهای ما نباشند لازم است. روشن است که گردآوری داده‌ها برای برآورده شدن نیازهای خاصی انجام شده است. لازمه داده‌های باکیفیت آن است داده‌های گردآوری شده در یک فرایند، قابل استفاده مجدد در سایر فرآیندهای سازمانی باشند. در واقع مشخصه‌ی دیگر داده‌های اصلی آن است که در سایر فرآیندهای سازمانی قابل استفاده مجدد باشند.

داده‌های اصلی نه تنها برای این‌که قابل استفاده مجدد هستند، اهمیت دارند، بلکه بخاطر اینکه در تبادلات تجاری به‌عنوان داده‌های مرجع مورد استفاده قرار می‌گیرند نیز مهم‌اند. بنابراین داشتن داده‌های اصلی جامع و به‌روز بسیار مهم است.

پاک‌سازی داده‌های اصلی جهت اطمینان از این‌که داده‌های اصلی پاسخگوی نیازها هستند، ضروری است. پاک‌سازی داده‌ها جهت اطمینان از این‌که رکوردهای داده‌های اصلی به‌روز هستند، لازم است و شامل همه

داده‌های مورد نیاز جهت پشتیبانی از همه مأموریت‌های یک کسب و کار است. به طور خلاصه، پاک‌سازی داده‌ها، فرایندی است که کیفیت داده‌های اصلی را ایجاد و حفظ می‌کند.

۶-۲ مزایای داده‌های اصلی با کیفیت

ممکن است تصور شود، تا زمانی که هزینه‌های پاک‌سازی داده‌ها بالا است، مزایا در مقابل هزینه‌ها به چشم نمی‌آیند. بهتر است منافع مستقیم و غیر مستقیم پاک‌سازی و مدیریت کیفیت داده‌ها بررسی بیشتری شود و با یک تحلیل دقیق هزینه منفعت مشخص شود مزایا به هزینه‌ها برتری دارند یا خیر. ملموس‌ترین منفعت پاک‌سازی داده‌های اصلی، شناسایی موارد تکراری موجود و بالقوه است. منافع مالی ناشی از حل مشکل موارد تکراری (دوگانگی) به‌راحتی در فعالیتهای مختلف مانند طراحی سیستم، مهندسی، ساخت، عملیات و نگهداری اندازه‌گیری می‌شود.

از نظر مالی، کیفیت داده‌های اصلی، عامل مهم و تعیین کننده‌ای در صحت تجزیه و تحلیل‌های مالی است. برنامه‌های کاربردی برنامه‌ریزی منابع سازمانی¹ (ERP) نمای خوبی از انسجام سازمان ارائه می‌کنند، اما صحت و دقت این نما به کیفیت داده‌های اصلی بستگی دارد. مشهور است که نقطه ضعف هر ERP، کیفیت داده‌های اصلی است. این موضوع به میزان زیادی از عجله در بارگذاری داده‌ها در زمان مورد نظر، عدم تعهد و التزام به کار، کمبود بودجه یا تجربه ناشی می‌شود. البته هیچ‌یک از این‌ها در صورت داشتن توانایی در ایجاد و مدیریت یک فرهنگ مفاهیم، یک یا چند طرح طبقه‌بندی و از همه مهم‌تر بیان الزامات داده‌ها (راهنمای شناسایی) رخ نخواهد داد.

از مهمترین خروجی‌های پروژه پاک‌سازی داده‌ها ایجاد داده‌های قابل انتقال است. داده‌های قابل انتقال حرکت داده از یک کاربرد به کاربرد دیگر را تسهیل می‌کند که به مقدار قابل توجهی هزینه کاهش و قابلیت اطمینان یکپارچگی داده‌ها افزایش می‌یابد. بدون داده‌های قابل انتقال، هزینه یکپارچگی داده‌ها بین دو شرکت بالا خواهد بود.

1. Enterprise Resource Planning



خروجی دیگر پاک‌سازی داده‌ها، بهبود کیفیت کاتالوگ فروش محصولات است. با توجه به اینکه امروزه عملیات جست‌وجو و سنجش کیفیت محصول با استفاده از تجهیزات رایانه‌ای انجام می‌شود، توصیفات با کیفیت محصولات در کاتالوگ‌های فروش (کاتالوگ الکترونیکی) مزیت رقابتی خوبی برای جذب مشتریان بالقوه و نگهداری مشتریان موجود ایجاد می‌کند.

فراتر از منافع مستقیم پاک‌سازی داده‌ها، مزایای غیر مستقیمی هم وجود دارد که به پاک‌سازی و مدیریت کیفیت داده‌های اصلی مرتبط هستند. مهم‌ترین آن‌ها ایجاد یک پایگاه دانش مشترک و استفاده از آن در برقراری ارتباط مؤثر میان عملیات مختلف سازمانی مانند حسابداری، تدارکات و مهندسی در داخل یک سازمان و سازمان‌های وابسته به آن است. مشارکت در توسعه بانک مفاهیم مشترک، روش سودمندی برای رسیدن به وحدت رویه است.

پاک‌سازی داده‌ها مانند هر فرایند دیگری تا وقتی که می‌تواند به‌طور موفقیت‌آمیزی در مقیاس کوچک به‌وسیله‌ی هر شخصی با مهارت و دانش مختصر به انجام برسد، ساده است. اما مهارت و دانش زیادی برای موفقیت در آن نیاز است. این موضوع خیلی شبیه آشپزی است که در آن تفاوت زیادی بین آشپزی برای یک خانواده، آشپزی حرفه‌ای برای صدها نفر مهمان و طراحی، ایجاد و مدیریت یک فرایند تولید غذا به‌صورت مداوم، مؤثر، اقتصادی با میلیون‌ها آیتم کیفی وجود دارد. آشپزی مقایسه خوبی است، شما ابتدا نیاز به توصیف محصول نهایی دارید، سپس با انتخاب مواد درست، فرایند درست و ابزار درست مواجه هستید. البته که این، تمام داستان نیست، شما علاوه بر آن‌ها به تجربه و مهارت نیاز دارید که این‌ها استعدادها و هدایایی طبیعی نیستند. مانند بیشتر فرایندها، تخصصی کردن و صنعتی کردن به شرکت‌ها کمک کرده است که ابزارهایی را برای افزایش سرعت و بهبود کیفیت و کاهش هزینه پاک‌سازی داده‌ها توسعه دهند.

۷. نتیجه‌گیری

فرایندی که پاک‌سازی و کیفی‌سازی داده‌های اصلی سازمان را تأمین می‌کند با نام فرایند "شناسایی و کاتالوگ‌سازی اقلام" معرفی می‌شود. این فرایند به‌طور ویژه به موضوع مدیریت داده‌های اصلی سازمان و شناسایی منابع فیزیکی می‌پردازد. لزوم اجرای این فرایند به‌عنوان یک فرایند پیش‌نیاز برای فرایندهای مرتبط با برنامه‌ریزی منابع سازمان از بدیهیات است.

۸. مراجع

1. TECHNICAL SPECIFICATION ISO/TS 8000-100: 2009(E) ;First edition 2009-08-01;Data quality-Part 100: Master data: Overview.
2. TECHNICAL SPECIFICATION ISO/TS 22745-1: 2009(E) ; First edition 2009-03-15;Industrial automation systems and integration - Open technical dictionaries and their application to master data -Part 1: Overview and fundamental principles.
۳. استاندارد دفاعی ایران ؛ ۳۰۶-۴ ؛ ۱۳۹۰ ؛ سامانه شناسه‌گذاری اقلام و محصولات دفاعی - مشخصات بانک اطلاعاتی محصول.
4. Peter R. Benson ; Managing a Data Cleansing (cleaning) or Cataloging Project.