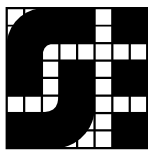


تعیین استاندارد‌های تهیه و ارائه

فراداده‌های آماری

تعیین استاندارد های تمیه و ارائه فرا داده های آماری

محمد رضا اناری (مجری) مرکز آمار ایران
سمانه قادری مرکز آمار ایران
عبدالحمید حقیقی مرکز آمار ایران
کمال نژادی مرکز آمار ایران
نیما دانش پرور پژوهشکده ی آمار



پژوهشکده ی آمار

گروه پژوهشی پردازش داده ها و اطلاع رسانی

بهار ۱۳۸۶

فهرست مطالب

مقدمه ۵

بخش اول: مبانی و کلیات

فصل اول: کلیات ۳

۱-۱ مفاهیم پایه ۴

۲-۱ داده‌های آماری ۱۲

۱-۲-۱ داده‌های خرد ۱۳

۲-۲-۱ داده‌های کلان (آمارها) ۱۶

۳-۱ فراداده‌های آماری چه هستند و چرا لازمند؟ ۲۰

۱-۳-۱ تعاریف مختلف فراداده‌ها ۲۰

۲-۳-۱ اهمیت وجود فراداده‌ها ۲۱

۴-۱ سیستم‌های اطلاع‌رسانی آماری ۲۲

۵-۱ چه کسی به فراداده‌های آماری نیاز دارد - و برای چه اهدافی ۲۷

۶-۱ چه فراداده‌های آماری مورد نیازند؟ ۳۱

۱-۶-۱ فراداده‌های مورد نیاز کاربران داده‌های آماری ۳۱

۲-۶-۱ فراداده‌های مورد نیاز تولیدکنندگان داده‌های آماری ۳۴

۳-۶-۱ فراداده‌های مرتبط با ابزارهای نرم‌افزاری ۳۵

فصل دوم: طبقه‌بندی انواع داده‌ها، فراداده‌ها و کاربران آماری ۴۱

۱-۲ طبقه‌بندی فراداده‌های آماری ۴۲

۱-۱-۲ فراداده‌های سیستم، کاربرد، اداری ۴۲

۲-۱-۲ فراداده‌های فعال، غیرفعال ۴۲

۳-۱-۲ انواع فراداده‌ها در مبادله‌ی داده‌ها و فراداده‌های آماری (SDMX) ۴۳

۴-۱-۲ طبقه‌بندی توأم فراداده‌های آماری ۴۴

۱-۴-۱-۲ فراداده‌های مربوط به نظام آماری خاص ۴۶

۲-۴-۱-۲ بازنگری کلی تهیه‌ی فراداده‌ها و دانش کلی ۵۰

۳-۴-۱-۲ مدل مفهومی شیء‌گرا برای فراداده‌های آماری ۵۱

- ۵۶ ۲-۲ طبقه‌بندی انواع کاربران فراداده‌های آماری
- ۵۹ ۳-۲ برخی از استانداردها در مورد فراداده‌های آماری
- ۵۹ ۱-۳-۲ استاندارد فراداده‌های روش‌ها و طراحی آمارگیری
- ۶۱ ۲-۳-۲ استانداردهای انتشار داده‌های خاص صندوق بین‌المللی پول
- ۶۱ ۳-۳-۲ استانداردهای مبادله‌ی فراداده‌ها و داده‌های آماری (SDMX)

۶۵ فصل سوم: منابع مختلف فراداده‌های آماری

- ۶۶ ۱-۳ منابع درون‌زای فراداده‌های آماری
- ۷۰ ۲-۳ منابع برون‌زای فراداده‌های آماری

بخش دوم: استانداردهای محتوایی فراداده‌های آماری و نحوه‌ی نمایش داده‌ها

۷۵ فصل چهارم: رهنمودهایی برای گزارش‌دهی انواع مختلف داده‌ها

- ۷۶ ۱-۴ داده‌های اولیه
- ۷۸ ۲-۴ شاخص‌ها
- ۷۹ ۳-۴ نرخ‌های رشد
- ۸۷ ۴-۴ نسبت‌ها، درصدها و نرخ‌ها

۹۳ فصل پنجم: رهنمودهایی برای گزارش شکل‌های متفاوت داده

- ۹۶ ۱-۵ اصطلاح‌شناسی
- ۱۰۲ ۲-۵ داده‌های تعدیل‌شده‌ی روزکاری
- ۱۰۲ ۳-۵ داده‌های تعدیل‌شده‌ی فصلی
- ۱۰۳ ۴-۵ داده‌های روند- دوره‌ای
- ۱۰۵ ۵-۵ کاربری‌های توصیه‌شده برای گزارش و ارائه‌ی شکل‌های مختلف داده‌ها
- ۱۰۵ ۱-۵-۵ اشکال داده‌ای که نمایش داده می‌شوند
- ۱۰۶ ۲-۵-۵ توصیه‌هایی در مورد تبدیل‌های تحلیلی
- ۱۰۸ ۳-۵-۵ تهیه‌ی اطلاعاتی درباره‌ی تعدیل فصلی به کاربران

۱۱۱ فصل ششم: رهنمودهایی برای گزارش‌دهی و انتشار فراداده‌ها

- ۱۱۵ ۱-۶ نیاز به استانداردهای محتوایی فراداده‌ها
- ۱۱۵ ۲-۶ کاربری‌های توصیه‌شده برای گزارش‌دهی و انتشار فراداده‌ها

۱۲۳	فصل هفتم: رهنمودهایی درباره‌ی کاربردهای گزارش‌دهی کلیدی
۱۲۴	۱-۷ تجدید نظر داده‌ها
۱۲۵	۱-۱-۷ نوع‌شناسی و اصطلاح‌شناسی
۱۲۹	۲-۱-۷ زمینه‌ی تجدید نظرها
۱۳۳	۳-۱-۷ کاربردهای مناسب برای خط مشی‌های تجدید نظر
۱۴۱	۲-۷ نمایش وقفه‌های سری‌های زمانی
۱۴۲	۱-۲-۷ روش‌هایی برای کاهش تأثیر وقفه‌های سری‌های زمانی
۱۴۳	۲-۲-۷ کاربردهای توصیه شده برای نمایش و گزارش‌دهی اطلاعات مربوط به وقفه‌های سری‌های زمانی
۱۴۴	۳-۷ خطاهای نمونه‌گیری و غیرنمونه‌گیری
۱۴۵	۱-۳-۷ خطاهای نمونه‌گیری
۱۴۹	۲-۳-۷ خطاهای غیرنمونه‌گیری
۱۵۳	۴-۷ نمایش و استفاده از اطلاعات مربوط به خطای نمونه‌گیری
۱۵۸	۵-۷ شاخص‌های پایه‌گذاری شده
۱۵۹	۱-۵-۷ استانداردهای بین‌المللی موجود
۱۵۹	۲-۵-۷ پردازش اعداد شاخص مجدد پایه‌گذاری شده
۱۵۹	۳-۵-۷ ارائه و انتشار شاخص‌های مجدد پایه‌گذاری شده
۱۶۰	۴-۵-۷ کاربردهای توصیه‌شده برای پایه‌گذاری مجدد
۱۶۰	۶-۷ نقل قول
۱۶۱	۱-۶-۷ استانداردهای موجود در ارتباط با نقل قول
۱۶۲	۲-۶-۷ کاربردهای توصیه شده برای نقل قول
۱۶۳	۷-۷ داده‌های اداری
۱۶۴	۱-۷-۷ کاربردهای توصیه شده برای ارائه و گزارش داده‌های اداری

بخش سوم: مخزن فراداده‌های آماری

۱۶۷	فصل هشتم: اجزای اصلی فراداده‌ها
۱۶۸	۱-۸ توصیف سیستم‌های تولید
۱۷۱	۲-۸ توصیف خروجی‌ها یا محصولات
۱۸۱	۳-۸ منابع قابل اشتراک و مستندات آن‌ها
۱۸۵	۴-۸ مؤلفه‌های انسانی و فنی به اشتراک گذاشته‌شده
۱۸۵	۵-۸ دانش و تجربیات مستند شده

- ۱۸۷ ۶-۸ مروری بر نظام‌های آماری
- ۱۹۱ ۷-۸ توسعه‌ی معماری فراداده‌ها
- ۲۰۰ ۸-۸ نگهداری زیربنای فراداده‌ها

۲۰۳ فصل نهم: مخزن فراداده‌ها

- ۲۰۷ ۱-۹ معماری مخزن فراداده‌های آماری
- ۲۱۱ ۲-۹ سایر استانداردها

بخش چهارم: کاربری‌های مناسب برای پیاده‌سازی وب‌گاه‌های آماری

۲۱۷ فصل دهم: بهترین کاربری‌ها در طراحی وب‌گاه برای انتشار آمار

- ۲۲۰ ۱-۱۰ کاربران یک وب‌گاه آماری
- ۲۲۲ ۲-۱۰ محتوای وب‌گاه
- ۲۲۴ ۳-۱۰ معماری وب‌گاه و عملکردهای آن
- ۲۲۸ ۴-۱۰ مسائل توسعه و پشتیبانی
- ۲۳۰ ۵-۱۰ برخی نتایج پایه
- ۲۲۳ ۶-۱۰ رهنمودهایی برای فراداده‌ها در اینترنت

۲۳۵ فصل یازدهم: رهنمودهایی برای مدل‌بندی فراداده‌ها

- ۲۳۶ ۱-۱۱ منابع فراداده‌های آماری
- ۲۳۶ ۲-۱۱ فرایندهای طراحی
- ۲۳۷ ۳-۱۱ فرایندهای عملیاتی و ارزیابی
- ۲۳۷ ۴-۱۱ فرایندهای نگهداری و طراحی مجدد
- ۲۳۸ ۵-۱۱ سازمان‌دهی جریان و نگهدارنده‌های موقت فراداده‌های آماری
- ۲۴۳ ۶-۱۱ جریان‌های فراداده

۲۴۹ یافته‌ها و پیشنهادها

۲۵۵ واژه‌نامه‌ی فارسی به انگلیسی

۲۶۰ واژه‌نامه‌ی انگلیسی به فارسی

۲۶۵ مرجع‌ها

مقدمه

عبارت فراداده از ۱۵ سال پیش مورد استفاده‌ی وسیع قرار می‌گیرد و با شهرت شبکه‌ی جهانی وب عمومیت یافته است. با وجود این، مفهوم اساسی مورد استفاده در مجموعه‌های وسیع اطلاعاتی سازمان یافته شده، بوده است. فهرست‌ها یا کاتالوگ‌های کتابخانه‌ها مثالی از یک نوع فراداده‌ی پرسابقه است که در دهه‌های اخیر برای مدیریت مجموعه‌ها و ابزار کشف منابع به خدمت گرفته شده‌است. فراداده‌ها نوعی داده محسوب می‌شوند. فراداده‌ها محتوی، کیفیت و شرایط داده‌ها را توصیف می‌نمایند، به پشتیبانی سیستم‌های اطلاع‌رسانی و به انتقال داده‌ها بین سیستم‌های رایانه‌ای یا اشتراک داده‌ها با سایر سازمان‌ها کمک می‌نمایند. تعاریف مختلفی برای این واژه در منابع معتبر ارائه شده‌است که به چند مورد از آن‌ها اشاره می‌شود:

- فراداده‌ها اطلاعاتی درباره‌ی داده‌ها و فرایندهای تولید و استفاده از داده‌ها هستند (تعریف ایزو).
- فراداده‌ها داده‌هایی هستند که برای تولید مطلوب و استفاده از اطلاعاتی که معرف آن هستند، تهیه می‌شوند.

فراداده‌ها در نظام‌های تولید آمار، اعم از نظام‌های ثبتي و طرح‌های آمارگیری به‌طور دستی یا خودکار تولید می‌شوند. به عبارت دیگر فراداده‌ها بخش جدایی‌ناپذیر این نظام‌ها هستند، تولید و اطلاع‌رسانی آمار بدون آن‌ها امکان‌پذیر نیست. در واقع فراداده‌ها ارتباط معنی‌دار بین تولیدکنندگان و کاربران داده‌ها را برقرار می‌سازند. این فراداده‌ها علاوه بر کاربران، مورد نیاز کارشناسان درگیر در مراحل مختلف تولید آمار نیز هستند. فراداده‌هایی که داده‌های آماری اعم از داده‌های خرد و کلان و فرایندها و ابزارهای دخیل در تولید و کاربرد داده‌های آماری را توصیف می‌نمایند به فراداده‌های آماری موسومند. به بیان دیگر فراداده‌های آماری داده‌هایی هستند که به توصیف جنبه‌های مختلف کیفی داده‌های آماری می‌پردازند. از جمله:

محتوا: تعاریف ارقام، جمعیت‌ها، متغیرها و غیره

درستی: تشریح و توصیف انواع خطاها از جمله خطای برآورد

دسترسی: بیان این‌که چه داده‌های آماری در دسترس هستند، در کجا قرار دارند و چگونه می‌توان به آن‌ها دست یافت.

در حوزه‌ی طرح‌های آمارگیری و با در نظر گرفتن مطالبی که در بالا ذکر شد، فراداده‌ها طیف گسترده‌ای از اسناد و گزارش‌ها مانند گزارش‌های کیفیت، تعاریف و مفاهیم و طبقه‌بندی‌ها، روش‌های جانمایی، طرح نمونه‌گیری، الگوریتم‌های ادیت، جداول نهایی، نمایش خطا و ... را شامل می‌شوند. برخی از این فراداده‌ها مانند تعاریف و مفاهیم برای جلوگیری از برداشت و استفاده‌ی نادرست از آمار مورد نیاز عموم کاربران هستند. استانداردهایی توسط برخی سازمان‌ها برای تهیه و ارائه‌ی فراداده‌های آماری مانند سازمان همکاری‌های اقتصادی و

توسعه^۱ (OECD) و صندوق بین‌المللی پول در چارچوب نظام انتشار داده‌های عمومی و استاندارد انتشار داده‌های خاص طراحی شده است. مبنای این استانداردها اصول بنیادی آمارهای رسمی سازمان ملل متحد هستند. در این اصول اهمیت ویژه‌ای به پیش‌گیری از تعبیر و استفاده‌ی نادرست از آمارها توسط کاربران و صراحت و شفافیت در زمینه‌ی گزارش‌دهی کاستی‌ها و مزایای آمار منتشر شده، داده شده است. مهم‌ترین مزیت‌های تهیه‌ی چنین استانداردهایی عبارت‌اند از:

الف) ارتقای سازگاری روش‌ها از طریق استانداردسازی قالب‌ها و الگوها، اصطلاح‌شناسی و انتشار داده‌ها و فراداده‌ها در بین مؤسسات آماری ملی و بین‌المللی

ب) ارتقای مقایسه‌پذیری داده‌ها، افزایش کیفیت و مناسب بودن داده‌ها

ج) ارتقای به‌کارگیری کاربردهای مناسب در نمایش آمار و بهبود شفافیت چنین کاربردهایی برای کاربران

د) مینیمم کردن بار گزارش‌دهی به سازمان‌های بین‌المللی

ه) شناسایی حوزه‌ها و موضوعاتی که در آینده لازم است دستورالعمل‌هایی برای گزارش‌دهی و نمایش داده‌ها و فراداده‌های مربوط به آنها تهیه شود.

استانداردهای فراداده‌ای صندوق بین‌المللی پول چهار بُعد نظام‌های مورد توصیه‌ی آن را یعنی داده‌ها، صداقت، کیفیت و دسترسی توسط عموم را در بر می‌گیرد. الگوهای مربوط به رده‌های داده‌ای مانند آمار بدهی خارجی، اقتصادی مانند حساب‌های ملی و نیروی کار، جمعیتی - اجتماعی مانند فقر و آموزش در دستورالعمل‌های مربوط از جمله سند نظام انتشار داده‌های عمومی مشخص و ارائه شده و از طریق سایت تابلو اعلانات استانداردهای انتشار اطلاعات رسانی می‌شوند. وب‌گاه فوق‌الذکر در سال ۲۰۰۰ راه اندازی شده و از طریق آدرس <http://dsbb.imf.org> قابل دسترسی است. این فراداده‌ها به تفصیل در گزارش‌های نهایی طرح پژوهشی «مطالعه، شناخت و ارزیابی نظام انتشار داده‌های عمومی» که توسط دفتر تعاریف و استانداردهای آماری مرکز آمار ایران در سال ۱۳۸۳ انجام شده، شرح داده شده است.

اکثر مراکز آماری و مؤسسه‌های بین‌المللی دستورالعمل‌ها و راهنمای انتشاراتی دارند. این دستورالعمل‌ها شامل نکات فنی در زمینه‌های مختلف مانند استانداردهای نمایش داده‌ها، طرح‌بندی جدول‌ها، نمودارها، قلم‌ها و ... هستند که در نشریه‌های چاپی و الکترونیکی می‌تواند مورد استفاده قرار گیرد. این راهنماها عموماً به مسائلی مانند نمایش وقفه‌های سری‌های زمانی، تجدید نظر، نحوه‌ی ارائه‌ی خطاهای نمونه‌گیری و غیرنمونه‌گیری، نقل قول و ... می‌پردازند. توصیه‌های عملی درباره‌ی این راهنماها عبارت‌اند از:

- هر مرکز آماری باید به منظور حصول اطمینان از سازگاری در ارائه و نمایش داده‌های اطلاع‌رسانی شده در نشریات چاپی و الکترونیکی، راهنمای انتشاراتی را برای کارشناسان تهیه‌کننده‌ی گزارش‌ها و نشریه‌ها و مؤلفان داخل سازمان تهیه نماید.
- مراکز آماری باید چنین راهنماها و دستورالعمل‌هایی را از طریق وب‌گاه در اختیار عموم قرار دهند.

^۱ Organization for Economic Cooperation and Development

اگر چه در سال‌های گذشته مرکز آمار ایران استانداردهای انتشاراتی را در قالب راهنمای تدوین نشریات آماری تدوین نمود اما در آن راهنما کم‌تر به استانداردهای محتوایی پرداخته شده و بیش‌تر نحوه‌ی نمایش داده‌ها و طراحی جداول را شامل می‌شود. از طرف دیگر در نشریات مرکز و به تبع آن در درگاه ملی آمار از الگوهای متفاوتی برای ارائه و نمایش فراداده‌ها استفاده شده است. به‌عنوان مثال در برخی نشریات تنها خطای نمونه‌گیری برآوردها ارائه شده و در برخی دیگر به خطاهای غیرنمونه‌گیری در برآوردها مانند بی‌پاسخی پرداخته شده است. در برخی زمینه‌ها مانند نحوه‌ی نمایش خطای نمونه‌گیری در جداول و نمایش وقفه در سری‌های زمانی نیز، استاندارد خاصی وجود ندارد. علاوه براین در صورت ارائه‌ی داده‌های خرد یا بانک‌های اطلاعاتی به کاربران نیز، فراداده‌های مربوط به آن‌ها نامشخص هستند. این طرح پژوهشی به منظور ارائه‌ی راهکار در زمینه‌های بالا و شناسایی ابعاد فراداده‌های آماری، نحوه‌ی مدل‌بندی و مبادله‌ی آن‌ها از طریق شبکه‌ی جهانی و استانداردهای مربوط انجام شد. فصل‌بندی و بخش‌بندی مطالب این گزارش براساس شرح خدمات مندرج در طرح پژوهشی انجام شده است. این بخش‌بندی گرچه در ابتدا جدا از یکدیگر به نظر می‌رسد اما ارتباط تنگاتنگی با یکدیگر دارند. در ادامه مطالب مندرج در این گزارش به اختصار شرح داده می‌شود:

در بخش اول این گزارش منابع، تعاریف و طبقه‌بندی‌های مختلف مربوط به فراداده‌های آماری، و ضرورت و اهمیت تولید و استفاده از آن‌ها شرح داده شده و برخی استانداردهای ملی و بین‌المللی در زمینه‌ی فراداده‌های آماری شرح داده می‌شوند. این بخش، پیش‌نیاز سایر بخش‌ها محسوب می‌شود. تکامل فناوری در طی سال‌های گذشته، خصوصاً فناوری مبتنی بر وب امکانات فنی زیادی را برای نگهداری فراداده‌ها در مخزن فراداده‌های آماری، بهبود روش‌های مورد استفاده در این زمینه و مبادله‌ی داده‌ها فراهم کرده است. هشت سازمان بین‌المللی، از جمله بانک جهانی و صندوق بین‌المللی پول فعالیت جدیدی را از سال ۲۰۰۰ با عنوان ابتکار مبادله‌ی داده‌ها و فراداده‌های آماری آغاز نموده‌اند. استاندارد نسخه‌ی یک این ابتکار در سال ۲۰۰۴ برای مبادله‌ی داده‌ها و فراداده‌های آماری از طریق شبکه جهانی اینترنت نهایی شد. مبنای اصلی این استاندارد زبان نشانه‌گذاری توسعه‌پذیر (XML) است.

بخش دوم این گزارش که شامل ۶ فصل است بیش‌تر به استانداردهای محتوایی فراداده‌های آماری و نحوه‌ی نمایش داده‌ها پرداخته است. دستورالعمل‌ها و توصیه‌های مربوط به نمایش و گزارش‌دهی که در این طرح پژوهشی بررسی شده براساس دامنه‌ی وسیعی از استانداردهای آماری بین‌المللی موجود است که در حوزه‌های آماری مختلف (اقتصادی، اجتماعی، جمعیت‌شناسی) توسط سازمان‌های بین‌المللی با همکاری مراکز آماری ملی تهیه شده است. این استانداردها توسط سازمان همکاری‌های اقتصادی و توسعه، برای کشورهای عضو آن سازمان تدوین شده است.

پروژه‌های اندکی در زمینه‌ی پیاده‌سازی و اجرای مخزن فراداده‌های آماری مشاهده شده است. در بخش سوم به معرفی مخزن فراداده‌های آماری پرداخته شده است. مخزن فراداده‌های آماری، یک مخزن فراداده‌های آماری مرکزی منطقی است که امکان پرس و جو، ویرایش و مدیریت فراداده‌ها را فراهم می‌کند. چنین سیستمی مکانیسمی را برای دستیابی به اطلاعات درباره‌ی محصولات آماری و علاوه بر آن طرح، توسعه و تحلیل فراهم می‌نماید. فراداده‌ها به ندرت ناقص یا گم شده یا پراکنده هستند. بیشتر وقت‌ها تنها منبع برای فراداده‌ها، کارشناسان موضوعی هستند. مدیریت مؤثر و کارای فراداده‌های آماری، سودمندی داده‌های آماری را به شدت افزایش می‌دهد. از آنجایی که فراداده‌ها نیز نوعی از داده‌ها هستند، می‌توانند درست مانند داده‌هایی که توصیف می‌نمایند و در پایگاه داده‌ها ذخیره و بازیابی می‌شوند، در یک مخزن ذخیره و بازیابی شوند. مخزن فراداده‌های آماری نباید جدای از پایگاه داده‌های آماری مراکز آماری طراحی شود. علاوه بر این در هنگام طراحی مخزن فراداده‌ها، باید عملکردها و وظایف بسیاری برای مخزن فراداده‌های آماری در نظر گرفته شوند. مخزن فراداده‌های آماری در درجه‌ی اول ابزار استاندارد برای محققین و تحلیل‌گران در یافتن داده‌ها و توصیف‌های آماری هستند. لغت‌نامه‌ی داده‌ها، طرح رکوردها، پرسش‌نامه‌ها، طرح‌های نمونه‌گیری و انحراف استاندارد از انواع اطلاعاتی هستند که به طور مستقیم در چنین مخازنی وجود دارند.

در بخش چهارم، نقش فراداده‌ها در اینترنت و ویژگی‌های مناسب برای وب‌گاه‌های آماری بررسی شده است. این بخش عمدتاً شامل توصیه‌ها و کاربردهای مناسب برای پیاده‌سازی وب‌گاه‌های آماری و مدل‌بندی فراداده‌های آماری است.

امید است مطالب ارائه‌شده در بخش دوم این گزارش بتواند در تکمیل راهنمای انتشاراتی مرکز آمار ایران با تأکید بر استانداردهای محتوایی سودمند واقع شود و مطالب ارائه شده در بخش سوم و چهارم در طراحی و ارتقا درگاه ملی آمار مورد استفاده قرار گیرد.

بخش اول

مبانی و کلیات

فصل اول

کلیات

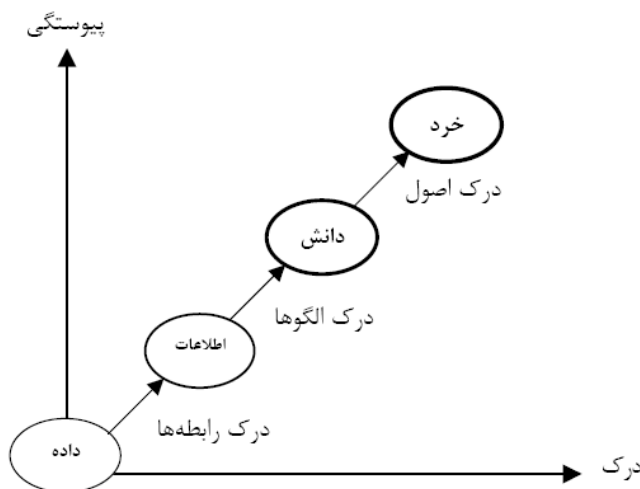
۱- کلیات

در این فصل با در نظر گرفتن ساختار داده‌ها، مفاهیم اولیه در مورد انواع داده‌ها بیان شده است. تفاوت بین داده‌ها و اطلاعات، مشخص و از دیدگاه‌های مختلف به شرح آن پرداخته شده و فراداده‌ها به عنوان نوعی از داده‌ها معرفی و اهمیت و لزوم وجود آن بیان شده است. همچنین ساختار یک نظام آماری ارائه و نقش فراداده‌ها در قسمت‌های مختلف آن نشان داده شده است.

۱-۱- مفاهیم پایه [۴ و ۵]

الف) داده‌ها، اطلاعات، دانش و خرد^۱

بدون تردید تمدن بشری مدیون پیشرفت امکانات ثبت، ذخیره‌سازی و پردازش داده‌ها است، که به همراه خود دیگر زمینه‌های تمدن بشری را به این حد تکامل رسانده است. امروزه افزایش سریع حجم پایگاه داده‌ها به شکلی است که توانایی انسان برای درک این داده‌ها بدون ابزارهای پرقدرت میسر نمی‌باشد. در این وضعیت، تصمیم‌گیری‌ها به جای تکیه بر اطلاعات بر درک مدیران و کاربران تکیه دارند، چرا که تصمیم‌گیرندگان ابزار قوی‌ای برای استخراج اطلاعات با ارزش را در دست ندارند. در واقع شرایط فعلی توصیف‌کننده‌ی حالتی است که ما از لحاظ داده‌ها غنی، اما از لحاظ اطلاعات ضعیف هستیم. حال با توجه به رقابت شدید در عرصه‌های علمی، اجتماعی، اقتصادی، سیاسی و نظامی استفاده‌ی موثر از داده‌ها توسط مدیران، یک هدف عمده برای بهبود وضعیت موجود محسوب می‌شود.



شکل ۱: سلسله مراتب رسیدن از داده به خرد

داده: داده، خام است، به صورت ساده وجود دارد و به خودی خود دارای هیچ معنی و مفهومی نیست. به هر حالتی ممکن است وجود داشته باشد (قابل استفاده و غیر قابل استفاده). در واقع داده‌ها نمایشی از واقعیت‌ها، معلومات، مفاهیم، رویدادها یا پدیده‌ها برای برقراری ارتباط، تفسیر یا پردازش، توسط انسان یا ماشین هستند.

اطلاعات: اطلاعات همان داده‌هایی هستند که به وسیله‌ی وابستگی‌های رابطه‌ای معنی و مفهوم گرفته است. در واقع اطلاعات همان داده‌های پردازش شده است. می‌توان گفت اطلاعات داده‌هایی هستند که پس از گردآوری، پردازش شده و به اطلاعات تبدیل شده‌اند. بین داده‌ها و اطلاعات همانند خبر و اطلاع رابطه وجود دارد. خبری که دریافت می‌شود، پس از ارزیابی به اطلاع تبدیل می‌شود. داده‌ها نیز پردازش می‌شوند تا اطلاعات را پدید آورند. به بیان دیگر اطلاعات حاصل تکامل داده‌ها است. وقتی انسان تصویری را مشاهده می‌نماید، این تصویر به عنوان یک داده از طریق دستگاه ورودی یعنی چشم وارد سیستم می‌شود و از طریق اعصاب به مغز می‌رود، مغز پس از دریافت تصویر آن را با تصویرهای موجود در حافظه مقایسه می‌کند. چنانچه تصویر دریافتی با یک تصویر موجود در حافظه مطابقت داشته باشد، اطلاع حاصل می‌شود. به این ترتیب بین داده‌ها و اطلاعات یک شکاف وجود دارد که اندازه‌ی این شکاف با حجم داده‌ها ارتباط مستقیم دارد. هرچه داده‌ها حجیم‌تر باشند، این شکاف بیشتر خواهد بود و هرچه حجم داده‌ها کمتر و روش‌ها و ابزار پردازش داده‌ها کارتر باشد، فاصله‌ی بین داده‌ها و اطلاعات کمتر است. از دیدگاه دیگر اطلاعات همان داده‌های پردازش شده به منظور پاسخ به پرسش‌های «چه، چرا، چه کسی، کی و کجا» است.

دانش: دانش به مجموعه‌ای از اطلاعات گفته می‌شود که به منظور استفاده‌ی خاصی گردآوری شده‌اند. می‌توان گفت که دانش فرایندی تشخیصی است. هنگامی که شخصی اطلاعاتی را به حافظه‌ی خود می‌سپارد در واقع وی در حال گردآوری دانش است. این دانش درک خوبی به وی می‌دهد ولی نمی‌توان از این دانش‌ها برای استنباط در مورد دانش‌های بیش‌تر استفاده کرد. از یک دیدگاه دانش، پاسخ‌هایی در جواب سوال‌هایی با عبارت پرسشی «چگونه» است.

برای مثال دانش‌آموزان مدرسه‌ی ابتدایی تمامی جدول ضرب ۱۰ در ۱۰ را یاد می‌گیرند. آن‌ها می‌توانند با دانشی که دارند حاصل ضرب ۲ در ۲ را انجام دهند اما اگر از آن‌ها حاصل ضرب ۱۰۲۶ در ۳۰۰ پرسیده شود نمی‌توانند پاسخ صحیحی دهند زیرا دانش آن‌ها در حد همان جدول ضرب ۱۰ در ۱۰ است. برای پاسخ به چنین پرسشی نیاز به درک درست و توانایی بیش‌تری دارند که در سطوح بالاتر از دانش این توانایی وجود دارد.

درک^۱: درک، فرایندی احتمالاتی و درون‌یاب^۲ است. هم‌چنین می‌توان گفت فرایندی شناختی و تحلیلی است. در واقع درک، فرایندی است که مرتبط با چگونگی به‌دست آوردن دانش است و یا این‌که چگونه می‌توان با آمیختن و ترکیب دانش‌های از قبل به‌دست آمده، دانش جدیدی را کسب کرد. چنانچه ملاحظه می‌شود درک و دانش به

۱- Understanding

۲- Interpolative and Probabilistic

هم نزدیک هستند اما تفاوت این دو مفهوم مانند تفاوت یادگیری و بخاطر سپردن (حفظ کردن) است. افرادی که دارای درک مناسبی هستند می‌توانند فعالیت‌های واگذار شده به خود را به خوبی انجام دهند. زیرا آن‌ها توانایی این که با استفاده از دانش پیشین، دانش جدیدی را به دست آورند را دارا هستند. درک، پاسخ به پرسش‌هایی است که با عبارت پرسشی «چرا» مطرح می‌شوند.

خرد: فرایندی غیر قطعی، غیر احتمالاتی و برون‌یاب است. همان‌طور که از شکل ۱ پیداست جایگاه این سطح در بالاترین سطح از آگاهی‌های حاصل از داده‌ها قرار دارد. به‌طور خاص می‌توان گفت سطوحی از اعمال و تصمیم‌گیری‌های انسانی (وجدان، دستورالعمل‌های اخلاقی و...) از این قسمت نشأت می‌گیرد. در این سطح آن‌چه را که قبلاً درک نشده و احتیاج به کاوش‌های فلسفی دارد مورد بررسی قرار می‌گیرد. بر خلاف سطح قبلی در این سطح سؤال‌هایی پرسیده می‌شود که به سادگی نمی‌توان به آن‌ها پاسخ گفت و در برخی از موارد هیچ مورد مشابهی از قبل وجود نداشته و باید با یک تفکر منطقی، استنباط‌های فلسفی و نتایج حاصل از سطوح قبل به سؤال‌های مطرح شده پاسخ داد. بنابراین خرد مرحله‌ای است که باید خوب و بد، درست و نادرست را از هم تمیز داد و در مورد آن‌ها قضاوت کرد. جالب توجه است که مراحل قبلی را می‌تواند رایانه انجام دهد ولی در این مرحله توانایی ندارد. این مرحله تنها مختص انسان‌هاست. به منظور روشن شدن مطلب به مثال زیر توجه کنید:

داده: حقیقت یا بیانی از یک واقعه بدون ارتباط آن با مابقی اشیاء است. برای مثال باران می‌بارد.

اطلاعات: تاکید بر درک ارتباطها بر اساس ترتیب و یا مطلوب‌تر از آن یافتن ارتباط علت و معلولی (در صورت امکان). برای مثال: درجه حرارت به ۱۵ درجه رسید و لذا باریدن باران شروع شد.

دانش: ارائه کننده‌ی الگوی ارتباطها و پیوستگی‌ها و به‌طور کلی در سطحی بالاتر ارائه‌کننده‌ی پیش‌بینی در مورد آن‌چه که در آینده اتفاق خواهد افتاد. به‌عنوان مثال: چنان‌چه میزان رطوبت خیلی زیاد شود و درجه حرارت نیز بالا رود آن‌گاه اتمسفر قادر نخواهد بود تحمل این رطوبت را داشته باشد و لذا بارش باران شروع خواهد شد.

خرد: خرد بیش از آن که به صورت اصولی از دانش حاصل شده باشد حاصل یک تفسیر و کاوش کلی از یافته‌ها به‌همراه استنباط‌های لازم است. برای مثال (فلاسفه می‌گویند) باران می‌بارد به دلیل این که باران می‌بارد. لذا باید به دنبال دلایل بارش بود و عوامل مختلفی را باید در نظر گرفت. در این مرحله باید درکی از عوامل مشترک که باعث بارش می‌شود را در نظر گرفت که از آن جمله می‌توان به تبخیر، هوای جاری، درجه‌ی حرارت، دگرگونی و تغییرات هوا، بارش اشاره کرد.

ب) داده‌های ادراکی^۱ و داده‌های نمادین^۲

انسان‌ها همواره اطلاعات و دانش را در ذهنشان ایجاد و پردازش کرده‌اند. آن‌ها خیلی پیش از اختراع رایانه‌ها، داده‌ها را تولید و پردازش کرده‌اند. در حقیقت انسان‌ها از ۲ نوع داده‌ها استفاده می‌کنند (شکل ۲ را ببینید). از یک سو انسان از طریق حواس خود یعنی دیدن، شنیدن، لمس کردن، مزه و بو کردن قادر به درک دنیای بیرون است. این حواس برای واقعیت، تولید و درک داده‌هایی استفاده می‌شوند که مغز انسان قادر به تغییر آن‌هاست. از طریق این داده‌ها ذهن انسان قادر خواهد بود تا مرحله به مرحله یک تصویر یا مدل ذهنی از دنیای بیرون ایجاد کند.

از سوی دیگر گاهی انسان‌ها سعی می‌کنند بخشی از مدل‌های ذهنی خویش را از دنیای واقعی از طریق تفسیر داده‌ها یعنی داده‌های نمادین نشان دهند. این داده‌ها قسمتی از ادراک و فهم انسان‌ها را از دنیای واقعی نشان می‌دهند. تفسیر برخی از داده‌های نمادین (تصاویر گرافیکی نمادین^۳، صداها^۴ شبیه‌سازی شده^۵) خیلی راحت است. دیگر داده‌های نمادین عبارت‌اند از: یک واژه در یک زبان یا نوشتاری از یک الفبای معین که مبتنی بر برخی از انواع قرارداد کدگذاری^۶ است و اگر کسی بخواهد این قبیل از داده‌ها را تفسیر نماید باید از این قراردادهای کدگذاری آگاه باشد.

از آن‌جایی که داده‌های ادراکی به‌طور مستقیم دنیای واقعی را منعکس می‌کنند ممکن است که این داده‌ها، داده‌های مستقیم^۷ نیز نامیده شوند. داده‌های نمادین داده‌های غیر مستقیم^۸ هستند به این مفهوم که این داده‌ها حاصل یک عمل ابداعی توسط انسان می‌باشند که توسط وی تفسیر می‌شوند یا احتمالاً توسط شخص دیگری دوباره مورد تفسیر قرار می‌گیرند.

هر شخص از طریق حواس خود به داده‌های ادراکی دست می‌یابد و پس از آن با استفاده از مفاهیم و چارچوب‌هایی که در آن زمان در ذهن یا در چارچوب منابع^۹ وی موجودند داده‌های دریافت شده را به یک ادراک از وضعیت موجود تفسیر می‌کند. مفاهیم و چارچوب‌های منابع حاصل یک فرایند یادگیری بلندمدت هستند. بخش عمده‌ای از این یادگیری تحت تأثیر محیط اجتماعی‌ای است که در آن زندگی می‌کنیم، از این‌رو چارچوب منابع و تفسیرهای ما مبتنی بر اجتماع هستند.

Perception data -۱

Symbolic data -۲

Iconographic pictures -۳

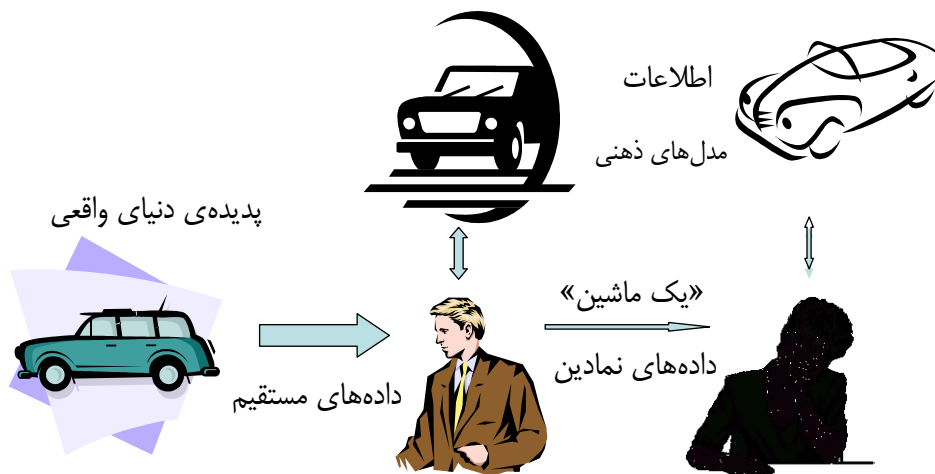
Onomatopoeic sounds -۴

Coding conversion -۵

Direct data -۶

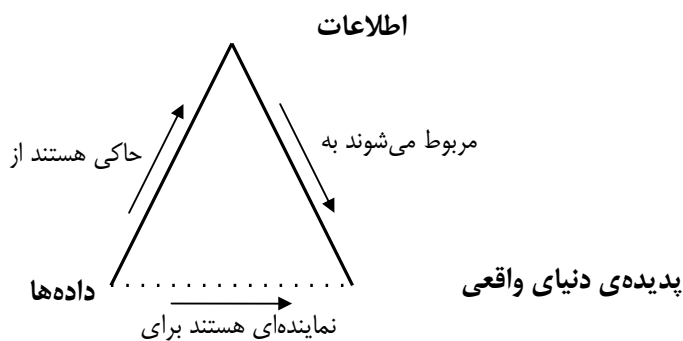
Indirect data -۷

Frame of reference -۸



شکل ۲: داده‌ها و اطلاعات

فلاسفه همواره در این که ما چگونه می‌توانیم اطلاعات و دانش را به‌دست آوریم در حال بحث هستند. در واقع در خصوص چگونگی عملکرد فرایند شکل‌گیری دانش، دیدگاه‌ها یا الگوهای متفاوتی وجود دارد ولی به نظر می‌رسد اغلب متفکران با این که ما برخی از انواع شکل‌های ذهنی یا مدل‌های مفهومی را از دنیای خارج درون ذهن خود شکل می‌دهیم و این مفهوم‌سازی پایه‌ای را برای درک واقعیت و برخورد ما با این واقعیت بنا می‌کند، موافق هستند. همان‌طور که در شکل ۳ به‌وسیله‌ی مثلث اوگدن^۱ نشان داده شده است داده‌های (نمادین) می‌توانند یک پدیده‌ی دنیای واقعی را نمایش دهند یا نماینده‌ای برای آن باشند. اما از آنجایی که این موضوع به پدیده‌های جهان واقعی‌ای بستگی دارد که در ابتدا توسط یک انسان به‌صورت ذهنی مفهوم‌سازی شده است. این رابطه فقط یک رابطه‌ی غیر مستقیم (تلویحی) است.



شکل ۳: مثلث اوگدن

از آن جایی که داده‌های نمادین خودشان بخشی از واقعیت هستند، ممکن است دوباره توسط انسان ادراک شده و به صورت مفاهیم و اطلاعات تفسیر شوند. تفسیرکننده ممکن است در اصل شخصی باشد که داده‌های نمادین را ذخیره کرده است ولی این تفسیرکننده ممکن است شخص دیگری نیز باشد. در مورد اول داده‌ها ممکن است برخی از چیزهایی را که یک شخص فراموش کرده است برای وی یادآوری نماید و در مورد آخر داده‌های نمادین ممکن است بخشی از یک فرایند ارتباطی میان دو شخص مورد بحث باشد.

ج) معادله‌ی منطق اطلاعات

لانگ‌فرس^۱ فرایند ذهنی تفسیر داده‌ها به اطلاعات را به وسیله‌ی فرمول منطق اطلاعات به صورت زیر شرح می‌دهد:

- $I = i(D, S, t)$

که در آن:

- I: عبارت است از محتوای اطلاعات به دست آمده توسط انسان
- i: عبارت است از فرایند تفسیر و ایجاد معنی یا مفهوم
- D: عبارت است از داده‌های دریافت شده
- S: عبارت است از چارچوب منابع، یا آگاهی انباشته شده‌ای که توسط مفسر استفاده می‌شود.
- t: عبارت است از زمان استفاده شده برای تفسیر

د) پردازش اطلاعات و پردازش داده‌ها

آدمی قادر است اطلاعات را بدون کمک هیچ یک از ابزارهای خارجی پردازش نماید. قبلاً در مورد این که ذهن چگونه دریافت‌های دنیای خارج را به منظور ایجاد منابع و اطلاعات تفسیر می‌کند، بحث شده است. فرایند تفسیر تا اندازه‌ای توسط یک تفسیر اولیه که در ذهن انسان قرار دارد کنترل می‌شود. حتی اگر یک شخص با یک ذهن خالی یا تقریباً خالی متولد شود چارچوب منابع وی همواره از طریق دریافت‌ها و تفسیرهای جدید بهنگام^۲ می‌شود. هم‌چنین انسان روی اطلاعاتی که از قبل در ذهنش وجود دارد تأمل می‌کند و این فرایند دوباره به صورت مفاهیم جدید یا اصلاح شده یا دانش جدید یا اصلاح شده از وی و دنیای بیرون حاصل خواهد شد. ما به طور دقیق نمی‌دانیم انسان چگونه اطلاعات را خلاصه^۳ می‌کند، اما قطعاً برای این کار هم عناصر استقرایی^۴ و هم عناصر

1- Langefors

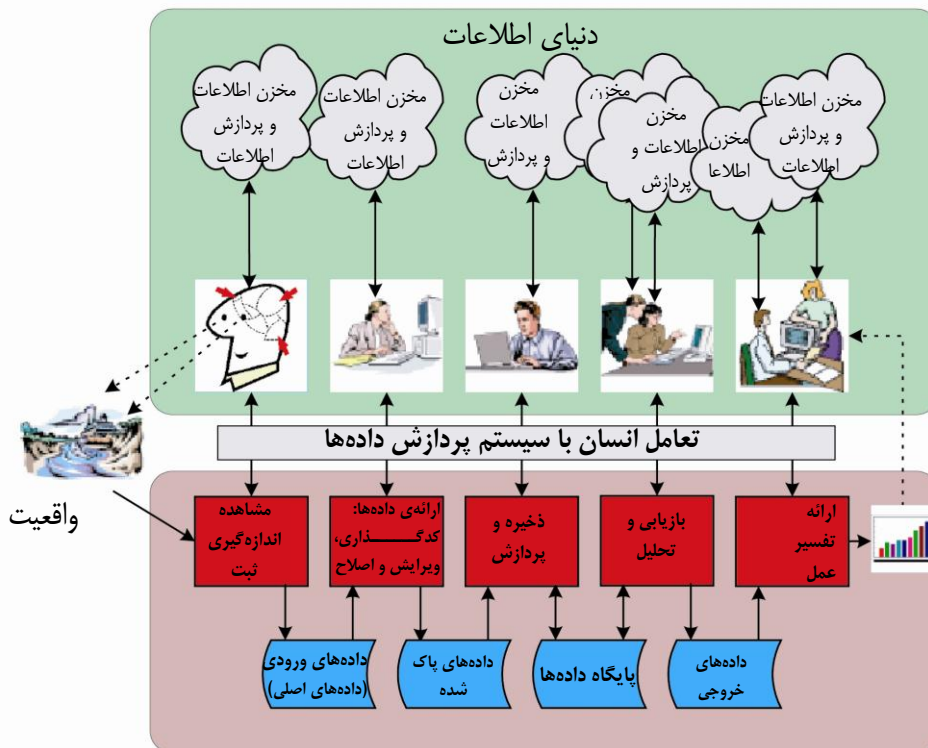
2- Update

3- Digests

4- Induction

استنباطی^۱ وجود دارند. برای شرح یک نوع معین از فرایندهای ذهنی که ما نمی‌توانیم واقعاً آن را مورد تحلیل قرار دهیم، واژه‌ای به‌عنوان «بینش»^۲ به‌کار می‌رود.

اطلاعات عمومی اغلب به‌صورت قوانین یا قواعد فرمول‌بندی می‌شوند. در چنین مواردی استفاده از واژه‌ی «دانش» به جای اطلاعات عادی است. هم‌چنین اطلاعات خاص گاهی «دانش» نامیده می‌شوند. به‌ویژه وقتی که اکتساب این اطلاعات به‌جای تفسیر یک دریافت ساده حاصل یک سری عملیات ذهنی پیچیده‌تر باشد.



شکل ۴: سیستم حمایت‌شده‌ی رایانه‌ای یا سیستم اطلاعات رایانه‌ای شده یعنی سیستم پردازش داده‌ها و سیستم‌های اطلاعاتی که با هم در تعامل‌ند.

مهم است که بدانیم، اطلاعات (الف) حداقل به این معنی که در بیرون از ذهن انسان جایی که به‌وجود آمده‌اند، نمی‌توانند ذخیره شوند؛ (ب) نمی‌توانند حداقل بدون انحرافات مهم و ناشناخته‌ی کم یا زیاد، به دیگر مردم انتقال داده شوند. ما می‌خواهیم اطلاعات را بیرون از ذهن مان با استفاده از وسایل ذخیره‌سازی (اطلاعات) خارجی به‌صورت یکی از توانایی‌های به‌خاطر سپردن و به‌خاطر آوردن اضافی و تقویتی خودمان ذخیره نموده و آن‌ها را هم به دلایل شخصی و هم اجتماعی با دیگر مردم مبادله کنیم، هنوز هم همه‌ی انسان‌ها همواره می‌خواهند تا این امور را با یکدیگر انجام دهند. قطعاً برای انجام این امور ما فرایندهایی را در دسترس خود داریم که توسط خداوند

deduction -۱

intuition -۲

به ما داده شده‌اند، به وسیله‌ی شانس و مکانیسم‌های ژنتیکی توسعه یافته‌اند یا توسط بشر اختراع شده‌اند. برای مثال ما در زبان‌های مختلف از پیام‌های گفتاری و نوشتاری استفاده کرده و رایانه‌ها را برای پردازش داده‌های نمادین به کار می‌بریم.

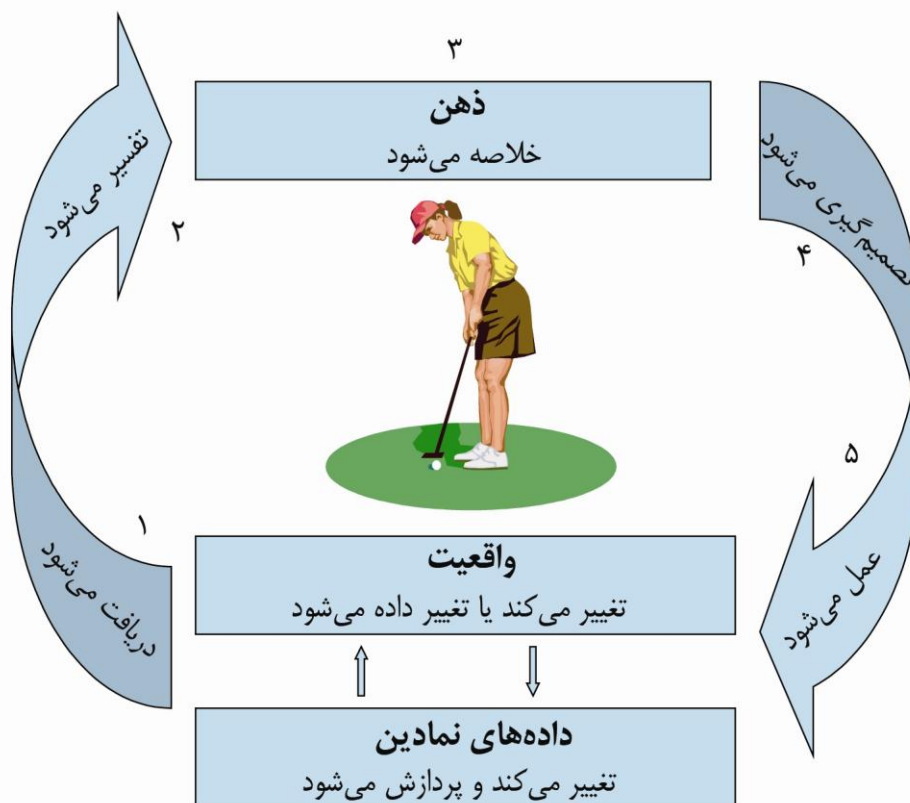
شکل ۵ یک نظر اجمالی را از برخی تعامل‌های اولیه میان (موارد زیر) ارائه می‌کند.

- اطلاعات و پردازش‌های آن در ذهن آدمی.
- داده‌ها اطلاعات بیرون ذهن آدمی را نمایش می‌دهند.
- دنیای واقعی، واقعیت.

این شکل یک نمای پویا را از دنیا ارائه می‌کند. در جایی که:

۱. انسان واقعیت را با استفاده از حواس خویش و احتمالاً برخی ابزارهای ساخت دست بشر دریافت می‌نماید.
۲. ذهن داده‌های دریافتی را با استفاده از چارچوب منابع موجود تفسیر می‌کند.
۳. ذهن مفاهیم و اطلاعات موجود را خلاصه می‌نماید.
۴. انسان برای انجام برخی از امور تصمیم می‌گیرد. به‌عنوان مثال واقعیت را تغییر می‌دهد یا داده‌هایی را به‌وجود آورده و بر طبق آن عمل می‌کند.
۵. واقعیت مورد تغییر قرار می‌گیرد یا خودش تغییر می‌کند.

یک «چرخه‌ی واقعیت»^۱ ابتدایی وجود دارد.



شکل ۵: تعامل‌های میان انسان، داده‌ها و واقعیت

۱-۲- داده‌های آماری [۶]

در یک نگاه کلی می‌توان داده‌های آماری را به سه گروه تقسیم کرد:

داده‌های خرد^۱: داده‌هایی در ارتباط با ویژگی‌های واحدهای آماری از قبیل افراد، خانوارها و کارگاه‌ها هستند که از طریق سرشماری یا آمارگیری به‌دست می‌آیند.

داده‌های کلان^۲: داده‌های مشتق‌شده (به‌دست‌آمده) از داده‌های خرد که توسط آماره‌هایی چون مجموع، میانگین، فراوانی و ... بیان می‌شود.

فرا داده‌ها^۳: داده‌هایی هستند که داده‌های خرد، کلان و سایر داده‌ها را توصیف می‌نمایند. در ادامه هر یک از این انواع داده‌ها به‌تفصیل مورد بررسی قرار گرفته است. داده‌های آماری موضوع اصلی هستند که در فراداده‌های

۱- Micro data

۲- Macro data

۳- Metadata

آماري فراهم و توضيح داده شده‌اند. بنابر اين به منظور فهم معنا و محتوای فراداده‌های آماری، باید فهمی از این که داده‌های آماری چه هستند و این که چه چیزهایی در مورد آن‌ها می‌تواند توضیح داده شود؛ پیدا کرد.

۱-۲-۱- داده‌های خرد

داده‌های خرد که گاهی داده‌های مشاهده‌شده یا داده‌های اندازه‌گیری‌شده نامیده می‌شوند، نتیجه‌ی مشاهده یا اندازه‌گیری مجموعه‌ای از مشخصات شیئی (حالات^۱ و رخدادها^۲) هستند. مشخصه‌ی شیئی می‌تواند به صورت یک زوج مرتب فرمول‌بندی شود:

$$C_o = (O(t), v(t)) \quad \text{یا} \quad C_o = (t).v(t)$$

که در آن :

(۱) نوع شیء

(۲) متغیر V

(۳) پارامتر زمان t

است. در برخی موارد O برداری از انواع اشیا است؛ در چنین مواردی V رابط یا متغیری است که براساس رابطه شکل گرفته است. $O(t)$ جمعیت اشیا موجود در زمان t و $v(t)$ وضعیت متغیر v در زمان t است که به شیئی‌های $O(t)$ مربوطند. در اغلب موارد، پارامتر زمان t در $v(t)$ و $O(t)$ زمان یکسانی نیست.

عنصر اصلی پایه‌ای اطلاعات در مورد مشاهدات و اندازه‌های مشخصه‌ی شیئی، در اصطلاح پیغام‌های الکترونیکی^۳ نامیده می‌شوند که دارای ساختار معنایی زیر هستند:

$$m_o = \langle o_i, p, t \rangle$$

$$\text{یا } m_o = [o_i.v(t) = v_j]$$

که در آن:

(۱) O_i نمونه‌ای شیئی متعلق به نوع شیئی O است.

(۲) p خاصیت است، نوعاً به عنوان مقدار O_j از بردار v بیان می‌شود.

(۳) t نقطه یا بازه‌ی زمانی است که در طول آن زمان گمان می‌رود که شیء دارای خاصیت p است یا آن را داشته است.

O_i می‌تواند برداری از اشیا، p رابطه (مانند «ازدواج») یا زوج $\langle v_i, v_j \rangle$ باشد که v بر اساس رابطه شکل گرفته است.

^۱ States

^۲ Events

^۳ E-messages

در نوعی تعامل میان سیستم اطلاع‌رسانی آماری و فراهم‌کننده‌ی ورودی برای سیستم، فراهم‌کننده مجموعه‌ای از سؤال‌ها را دریافت می‌کند که اغلب به‌صورت سلسله‌مراتبی بر حسب پاسخگو سازمان‌دهی شده‌اند. گاهی اوقات پاسخگو کاملاً مانند یکی از اشیاء تعیین شده است.

پرسش‌نامه‌ها با فراداده‌هایی به شکل توضیحات، دستورالعمل‌ها و غیره همراه هستند. در بعضی از سیستم‌ها در صورت نیاز ممکن است پاسخ‌گو متقابلاً فراداده‌های بیشتری درخواست نماید. زمانی که پیغام‌های پاسخگویان به سیستم اطلاع‌رسانی آماری ارسال می‌شود، انواع دیگری از فراداده‌ها مانند شرایط استثنایی مربوط به فرایند مشاهده به همراه آن خواهد آمد.

زمانی که مجموعه‌های ساخت‌یافته‌ی سلسله‌مراتبی پیغام‌های پاسخگویان به سیستم اطلاع‌رسانی آماری وارد می‌شوند اغلب دیر یا زود به فایل‌های مسطح^۱ یا جداول رابطه‌ای^۲ مطابق با روش‌های نسبتاً خوب استاندارد شده‌ای که توسط بسیاری از نرم‌افزارهای تجاری حمایت می‌شوند، تبدیل می‌شوند (قس. ابزارهای نقل و انتقال در سیستم‌های مدیریت پایگاه داده‌های رابطه‌ای). در حالت ایدئال باید در فرا اطلاعات همراه، به‌طور سیستماتیک از پردازش‌های موازی اجتناب شود. اما این پردازش‌ها هنوز به درجه‌ی خاصی از استانداردها نائل نشده‌اند.

ثبت مشاهدات: شامل داده‌های خرد مشاهده‌شده یا به‌دست آمده است که مهم‌ترین نوع محصول‌های داده‌ای حاصل از آمارگیری است. کاربران ماهر بیش از پیش تقاضای دسترسی به داده‌های خرد، برای انجام تحلیل‌های مورد نظرشان در محیط‌های خاص کامپیوتری دارند. خدمات آماری می‌توانند از طریق تهیه‌ی فایل از داده‌های خرد بدون ذکر نام^۳ مانند فایل‌هایی که به فایل‌های عمومی موسوم است، به این درخواست‌ها و تقاضاها پاسخ دهند. کاربر خارجی ثبت مشاهدات نباید در وضعیتی قرار گیرد که به کارکنانی یکبار داده‌ها را تولید کرده‌اند (شاید در سال‌های قبل) دسترسی داشته باشد. بنابراین ثبت مشاهدات باید با مجموعه‌ی مناسبی از فراداده‌ها همراه باشد.

^۱ Flat files

^۲ Relational tables

^۳ Anonymized