

معرفی ابزارها و نماگرهای کلیدی ارزیابی کیفیت ثبت‌ها و داده‌های اداری برای استفاده در سرشماری نفوس و مسکن

پریا ترابی کهلان^۱، لیدا کلهری ندرآبادی^۲

تاریخ دریافت: ۱۴۰۲/۰۵/۳۱

تاریخ پذیرش: ۱۴۰۲/۰۸/۱۶

چکیده:

استفاده از منابع اداری در اجرای سرشماری‌ها، امکان کاهش هزینه‌ها، بهبود کیفیت داده‌ها و تولید اطلاعات با توالی زمانی کوتاه‌تر را فراهم می‌کند. موارد ذکر شده می‌تواند علاوه بر نظارت سالانه‌ی نماگرهای اهداف توسعه‌ی پایدار، در تأمین نیازهای رو به گسترش نظام برنامه‌ریزی و پژوهشی کشور نیز نقش بسزایی داشته باشد، اما چالش‌های زیادی نیز در این رابطه وجود دارد که یکی از مهم‌ترین آن‌ها ارزیابی کیفیت ثبت‌های اداری است. ارزیابی کیفیت اصلی‌ترین جنبه‌ی به‌کارگیری داده‌های ثبتی و اداری در سرشماری و از ضروریات اجرای سرشماری ثبتی‌مبنا است که در آن نمی‌توان از معیارهای سنتی ارزیابی کیفیت استفاده کرد. به‌عبارت‌دیگر، علی‌رغم مزایای به‌کارگیری داده‌های ثبتی و اداری در سرشماری، بسیاری از مخاطره‌های کلیدی کیفیت وجود دارند که باید قبل از استفاده از آن‌ها در سرشماری بررسی و ارزیابی شوند. در این مقاله روش‌هایی که اداره‌های آمار ملی چگونه می‌توانند کیفیت داده‌های به‌دست‌آمده از ثبت‌ها را باهدف تولید برون‌داده‌های آماری باکیفیت ارزیابی کنند، مرور شده است. از این‌رو، ابزارها و نماگرهای کلیدی مورد استفاده برای تعیین کمیت ارزیابی کیفیت در هر یک از چهار مرحله‌ی ارزیابی کیفیت شامل منبع، داده‌های ورودی، فرایند و برون داد در فرایند سرشماری معرفی شده است.

واژه‌های کلیدی: سرشماری سنتی، سرشماری ثبتی‌مبنا، ارزیابی کیفیت داده‌های اداری.

۱ مقدمه

هدف را به‌طور منظم به‌عنوان مثال سالانه یا در برخی موارد حتی با تواتر زمانی کوتاه‌تر، ارائه می‌دهند. این بدان معناست که نظام ثبتی‌مبنا برای تولید آمارها در مورد گروه‌های جمعیتی و مناطق جغرافیایی کوچک بسیار مناسب‌تر از سرشماری سنتی با الزامات معمول برای حفاظت از محرمانگی است. هرچند حرکت از سرشماری سنتی به سرشماری ثبتی‌مبنا یا رویکرد ترکیبی برای NSOها مطلوب است و کاهش هزینه و بار پاسخ‌گویی را به دنبال دارد اما با چالش‌ها و پرسش‌های جدی نیز روبرو است که یکی از مهم‌ترین آن‌ها «ارزیابی کیفیت» داده‌های ثبتی به‌منظور استفاده در سرشماری ثبتی‌مبنا یا ترکیبی است [۵]. در نتیجه‌ی انتقال از رویکرد سرشماری سنتی به ثبتی‌مبنا، ممکن است تغییراتی در تعاریف متغیرها، جمعیت پایه و رده‌بندی برون‌داده‌ها ایجاد شود. تأثیر این تغییرات بر کیفیت برون‌داده‌های آماری باید با استفاده از آزمون و برنامه‌ی ارزیابی کامل مورد بررسی قرار گیرد و نتایج برای کاربران توضیح داده شود. معمولاً ارزیابی کیفیت آمارهای تولیدشده

سرشماری‌های سنتی هر ده سال و یا هر پنج سال یک‌بار با صرف وقت و هزینه‌های زیادی از طریق مراجعه‌ی حضوری به خانوار و تکمیل اطلاعات پرسش‌نامه بر اساس اظهار پاسخ‌گو انجام می‌شود که اداره‌های آمار ملی^۳ (NSOs) را با چالش نیاز به آمارهای بهنگام و در نتیجه گذار از سرشماری سنتی به سرشماری ثبتی‌مبنا (رویکرد ترکیبی) روبرو کرده است. بر اساس این رویکرد، گردآوری مستقیم داده‌ها از جامعه وجود ندارد و با استفاده از داده‌های اداری نگهداری شده در ثبت‌های مختلف از طریق فرایند جورسازی داده‌های موجود در منابع اداری و به‌کارگیری شماره‌های شناسایی فردی، نسبت به تولید اطلاعات آماری مورد نیاز اقدام می‌شود. از این‌رو، هزینه‌های گردآوری و تولید داده‌های آماری سرشماری‌ها کاهش یافته و با نیروی انسانی نسبتاً محدودی، نظام جامع ثبت‌های آماری باکیفیت و پایدار ایجاد می‌شود [۲]. مزیت اصلی داده‌های ثبتی‌مبنا این است که اطلاعات مربوط به کل جامعه‌ی

^۱ پژوهشکده‌ی آمار (نویسنده مسئول: torabiparya@yahoo.com).

^۲ استادیار پژوهشکده‌ی آمار.

کیفیت آمارهای تولیدی منجر می‌شود. سوم این‌که عموماً داده‌های اداری برای اهداف آماری گردآوری نشده‌اند. در نتیجه، در منابع داده‌ها ممکن است مفاهیم، رده‌بندی‌ها و تعاریف متفاوتی نسبت به موارد مورد نیاز در سرشماری اتخاذ شده باشد، یا آن‌ها به دوره‌های مرجع متفاوتی اشاره داشته باشند، در به‌روزرسانی اطلاعات تأخیر داشته و یا پوشانند محدودی از جامعه‌ی سرشماری داشته باشند. علاوه بر این، درستی و کامل بودن^۷ داده‌ها به‌شدت به اهمیت داده‌ها در عملکرد متولیان داده‌ها بستگی دارد. منابع اداری نیز ممکن است در طول زمان در معرض تغییرات و ناسازگاری‌ها در روش گردآوری داده‌ها در بخش‌های مختلف جامعه قرار داشته باشند. علاوه بر این، منابع داده‌ها ممکن است متغیرها یا شناسانه‌های^۸ لازم برای اتصال رکورد‌های مورد نیاز در سرشماری را نداشته باشند. در نهایت، پیچیدگی داده‌های اداری و در دسترس بودن و کامل بودن فراداده‌های مرتبط، بر توانایی NSO در درک، دسترسی و استفاده از منبع اداری تأثیرگذار خواهد بود. برای مثال، داده‌های اداری را می‌توان در ساختارهای داده‌ای بزرگ و پیچیده نگهداری کرد که این موضوع می‌تواند چالش‌های فنی مهمی را برای NSO در ارزیابی و غلبه بر این مشکلات ایجاد کند. همچنین، پیچیدگی داده‌های اداری ممکن است بر دسترسی‌پذیری و وضوح بعد کیفیت برون‌داد از دیدگاه کاربران داده‌ها تأثیرگذار باشد. به‌عبارت‌دیگر، درک استفاده از داده‌های اداری در سرشماری و تأثیر استفاده از آن‌ها بر کیفیت برون‌داد‌های سرشماری، برای کاربران داده‌های سرشماری ممکن است دشوار باشد [۱۳، ۱۴]. به‌طور کلی، ارزیابی کیفیت ثبت‌های اداری و داده‌های سرشماری به‌دست‌آمده از آن‌ها در چهار مرحله‌ی گردآوری داده‌ها، پردازش داده‌ها، تولید برون‌داد و ارزیابی انجام می‌شود. چهار مرحله‌ی مذکور به موارد زیر اشاره دارد:

- **کیفیت منبع:** ارزیابی کیفیت فراداده-مبنای^۹ منابع داده‌های اداری جدید یا موجود که در سرشماری مورد استفاده قرار می‌گیرند،
- **کیفیت داده‌های ورودی:** ارزیابی داده‌های اداری خام که توسط سازمان‌های اداری به NSO ارائه می‌شوند،
- **کیفیت فرایند:** ارزیابی تغییرات در کیفیت داده‌هایی که از یکپارچه‌سازی و پردازش داده‌های اداری در NSO حاصل

با استفاده از منابع اداری به دلیل پیچیدگی و چندبعدی بودن داده‌های مورد استفاده چالش خاصی را به همراه دارد. بسیاری از عوامل مؤثر بر کیفیت به‌طور کمی قابل اندازه‌گیری نیستند. علاوه بر این، آمارهای باکیفیت از نظر کاربران متفاوت هستند. به‌طور مثال، برای برخی از کاربران درستی از اهمیت بالایی برخوردار است اما برخی از کاربران دیگر ممکن است روزآمدی^۴ و دسترسی‌پذیری^۵ را در اولویت قرار دهند.

علی‌رغم تمام مزایای استفاده از داده‌های اداری، بسیاری از مخاطره‌های کلیدی کیفیت وجود دارند که باید قبل از استفاده از آن‌ها در سرشماری بررسی و ارزیابی شوند. اولاً، NSO تنها کنترل محدودی بر روش گردآوری و پردازش داده‌ها دارد و آن‌ها به‌طور قابل‌توجهی به متولیان داده‌های اداری وابسته هستند. به‌عنوان مثال، اگر متولیان داده‌ها نتوانند الزامات NSO برای ارائه‌ی داده‌های مناسب در زمان مناسب را برآورده کنند، روزآمدی نتایج سرشماری تحت تأثیر قرار خواهد گرفت. به‌طور مشابه، اگر متولیان داده‌ها به‌اندازه‌ی کافی با NSO در مورد هرگونه تغییرات احتمالی در منبع تعامل نداشته باشند، هم‌خوانی و مقایسه‌پذیری^۶ نتایج سرشماری تحت تأثیر قرار خواهد گرفت. دوم این‌که کاربرد داده‌های اداری توسط NSO برای اهدافی به‌غیر از اهداف اداری، منجر به افزایش دغدغه‌های قانونی، امنیت و حریم خصوصی می‌شود. NSO باید پذیرش عمومی داده‌ها را ارزیابی کند و نسبت به استقرار تعهدات مورد نیاز و اطلاع‌رسانی آن‌ها به عموم مردم و متولیان داده‌ها اطمینان حاصل کند. همچنین، کاربرد داده‌های اداری باید قانونی باشد. بدون پذیرش عموم مردم و متولیان داده‌ها یا مبنای قانونی معتبر برای استفاده از منابع اداری، خطر قابل‌توجهی برای اعتبار NSO و توانایی آن برای ارائه‌ی سرشماری باکیفیت وجود خواهد داشت [۱۶]. این مسئله می‌تواند ناشی از تغییرات رفتاری عموم مردم در نحوه‌ی تعامل با متولیان داده‌ها یا NSO به دلیل نگرانی در مورد نحوه‌ی استفاده NSO از داده‌های آن‌ها باشد. با این وجود، در سرشماری ثبتی مبنای به دلیل اظهاری نبودن اطلاعات مورد استفاده و به‌کارگیری داده‌های اداری مستند، متغیرهایی مانند سن، جنس، سواد، تحصیل، ازدواج، ویژگی‌های مسکن و ... به‌مراتب از دقت و کیفیت بالاتری برخوردار خواهند بود. همچنین با استفاده از داده‌های اداری، امکان اعتبارسنجی داده‌ها با منابع اطلاعاتی مختلف فراهم می‌گردد که در نهایت به افزایش

⁴Timeliness

⁵Accessibility

⁶Coherence and Comparability

⁷Accuracy and Completeness

⁸Identifiers

⁹Metadata-Based Quality Assessment

می‌شوند و

• **کیفیت برونداد:** ارزیابی کیفیت کلی نتایج آماری که برای کاربران منتشر می‌شوند [۱۶].

این مقاله به این موضوع می‌پردازد که NSOها چگونه می‌توانند کیفیت داده‌های به‌دست‌آمده از ثبت‌ها را باهدف تولید بروندادهای آماری باکیفیت ارزیابی کنند. از این رو، ابزارها و نماگرهای کلیدی مورد استفاده برای تعیین کمیت ارزیابی کیفیت به ترتیب در هر یک از چهار مرحله‌ی مذکور در ارزیابی کیفیت در فرایند سرشماری مورد بحث قرار می‌گیرد.

اختلاف در تعاریف و تعارض منافع پیش روی سازمان‌ها در این گذار را بیان کردند.

در سال‌های اخیر سازمان‌های بین‌المللی و سازمان‌های آماری کشورهای مختلف نیز گام‌های مهمی در ارائه‌ی توصیه‌ها و تجارب اجرای سرشماری ثبتی مبنا برداشته‌اند که از آن میان می‌توان به توصیه‌های کمیسیون اقتصادی سازمان ملل برای اروپا در ارزیابی کیفیت داده‌های اداری مورد استفاده در سرشماری و کتاب راهنمای سازمان ملل برای سرشماری ثبتی مبنا اشاره داشت [۱۳، ۱۴، ۱۶].

۲ پیشینه‌ی پژوهش

اگرچه نتایج حاصل از سرشماری‌های سنتی با ارزش هستند، اما با افزایش جمعیت کشورها و افزایش روزافزون تقاضاهای آماری کاربران و سیاست‌گذاران در خصوص اطلاعات آماری بیشتر و دقیق‌تر و در فواصل زمانی کمتر به‌ویژه در چارچوب نظام‌های مدرن برنامه‌ریزی از یک سو و از سوی دیگر سنگین بودن و مشکلات اجرایی سرشماری‌ها به روش سنتی، همواره کشورها را بر آن داشته تا به سمت روش‌های بهینه‌تر انجام سرشماری پیش بروند. در همین راستا، شورای عالی آمار با تصویب سومین برنامه‌ی ملی توسعه‌ی آماری کشور ۱۳۹۶-۱۴۰۰، «تولید و انتشار آمارهای رسمی مورد نیاز ذینفعان داخلی و بین‌المللی با رعایت اصول بنیادین آمارهای رسمی و به‌کارگیری فناوری‌ها و روش‌های مدرن به‌ویژه ثبتی مبنا» را به‌عنوان مأموریت نظام آماری کشور و «کاهش تدریجی آمارگیری‌های سنتی و جایگزینی آن با آمارهای ثبتی مبنا» را به‌عنوان راهبردهای کلان نظام آماری تعیین کرد. مرکز آمار ایران در راستای مدرن‌سازی نظام آماری بهبود روش‌های سرشماری و طرح‌های آماری، ثبتی مبنا کردن سرشماری عمومی نفوس و مسکن را در دستور کار خود قرار داده است [۳]. لذا باهدف حرکت به سوی رویکرد ثبتی مبنا در سرشماری عمومی نفوس و مسکن، مطالعات گسترده و بهره‌گیری از تجربیات کشورهای پیشرو در این زمینه انجام شده است که از آن جمله می‌توان به [۲] و ارائه‌ی تجارب سایر کشورها [۱، ۱۲] اشاره داشت. [۵] نیز به ارزیابی کیفیت داده‌های ثبتی و اداری مورد استفاده در سرشماری پرداختند. آن‌ها توانستند چارچوب ارزیابی کیفیت را به‌گونه‌ای طراحی کنند تا ابعاد مختلف کیفیت داده‌های ثبتی مورد توجه قرار گیرد. علاوه بر آن، [۴] ضمن بررسی نقدانه‌ی تجربیات جهانی در زمینه‌ی سرشماری ثبتی مبنا، زیرساخت‌های کشور برای گذار از سرشماری سنتی به ثبتی مبنا را بررسی کردند. آن‌ها چالش‌های موجود،

۳ چارچوب کیفیت

کیفیت در آمار به این موضوع بستگی دارد که آیا برونداد آماری، کاربرد مورد نظر را تحقق می‌بخشد یا نه. به‌عنوان مثال، اداره‌ی آمار اروپا کیفیت در آمار را مطابق با مجموعه استانداردهای ISO-۹۰۰۰ این‌گونه تعریف می‌کند: «میزانی که مجموعه‌ای از ویژگی‌های ذاتی یک شیء را تحقق می‌بخشد» [۱۰]. در آمار رسمی، شیء می‌تواند «تولید، خدمات، فرایند، نظام، روش‌شناسی، سازمان، منبع، یا داده‌های آماری» باشد [۸]. بنابراین، کیفیت داده‌های اداری مورد استفاده در سرشماری باید با توجه به روش‌های گردآوری و پردازش داده‌ها توسط متولیان و NSOها در کل فرایند تا مرحله‌ی ارائه‌ی بروندادهای نهایی سرشماری در نظر گرفته شود. باین وجود، در طول مراحل فوق، احتمال بروز خطاهایی وجود دارد که می‌تواند منجر به کاهش کیفیت شود. به عبارتی، خطا به‌عنوان تفاوت بین برآورد نهایی و پارامتر واقعی جامعه در نظر گرفته می‌شود. این موضوع در مدل فرایند کسب‌وکار آماری عمومی^{۱۰} (GSBPM) شناخته شده است و ساختار استاندارد را برای تشریح اکثر فرایندهای آماری ارائه می‌کند و بیان‌گر «کیفیت» به‌عنوان جنبه‌ای است که تمام مراحل آن را در بر دارد [۱۵]. بهبود کیفیت محصولات آماری نیازمند بهبود کیفیت فرایندهای آماری است. GSBPM مجموعه‌ای از فرایندهای کسب‌وکار را معرفی می‌کند که برای تولید آمارهای رسمی لازم است و به این دلیل چارچوبی برای مستندسازی، ارزیابی و بهبود کیفیت فرایند فراهم می‌کند. مدیریت کیفیت در GSBPM به‌عنوان فرایندی فراگیر تعریف می‌شود که شامل سازوکارهای ارزیابی کیفیت و کنترل است. این مدل بر اهمیت ارزیابی و بازخورد ارزیابی کل فرایند کسب‌وکار آماری تأکید دارد. GS-BPM به‌طور خاص، برای تمام فعالیت‌های تولیدی در نظر گرفته شده است که توسط تولیدکنندگان آمارهای رسمی انجام می‌شود و منجر

¹⁰General Statistical Business Process Model

روش‌های ارزیابی کیفیت منابع داده‌ها را می‌توان در چهار گروه اصلی رده‌بندی کرد:

- **مدیریت و برنامه‌ریزی:** این رده شامل فعالیت‌هایی است که به رابطه‌ی بین NSO و متولیان داده‌ها مربوط می‌شود. این موضوع مستلزم ایجاد حاکمیت و قاعده‌ها، حفظ ارتباطات خوب و مدیریت توافق‌نامه‌ها، سیاست‌ها و برنامه‌های تحویل داده‌ها است. هرگونه قرارداد تبادل داده نه‌تنها باید قوانین مربوط و محدودیت‌های محرمانه را در نظر بگیرد، بلکه باید سطح تأیید عمومی استفاده از یک منبع داده‌ی خاص را نیز در نظر داشته باشد،
- **سازمان‌دهی و مسئولیت‌ها:** این رده شامل جنبه‌های مرتبط با تعیین مسئولیت‌ها درباره‌ی منابع داده‌ها است. در این زمینه، ایجاد هماهنگی بین متولیان داده‌ها برای تأمین داده‌ها و حصول اطمینان از تحویل به‌موقع آن‌ها امری ضروری است،
- **فرایندها:** این رده عمدتاً با سامانه‌هایی سروکار دارد که انتقال حجم زیادی از داده‌ها را فراهم می‌کنند. انتقال داده‌ها باید در یک محیط امن مدیریت شود که سطوح موردنیاز حریم خصوصی و محرمانگی را میسر سازد،
- **ابزارها:** این رده شامل به‌کارگیری و توسعه‌ی ابزارهایی برای پشتیبانی از کیفیت منبع داده‌ها است. به‌طور مثال می‌توان به استانداردهایی برای قالب داده‌ها، مدل داده‌ها، پروتکل انتقال داده‌ها، مستندسازی منابع اداری (فرا داده‌ها) که اشیاء^{۱۱} (واحدها، وقایع) و متغیرها را با رده‌بندی آن‌ها توصیف می‌کنند، اشاره کرد. به‌عنوان مثالی دیگر، ارزیابی تأثیر حریم خصوصی ابزاری است که مرتبط با ارزیابی پذیرش عمومی است و یک چارچوب قانونی برای حصول اطمینان از اختیارات NSO به‌منظور دریافت و استفاده از داده‌ها نیز ضروری است.

علاوه بر بعد محیط سازمانی، تمام ابعاد کیفیت (شامل مرتبط بودن، روزآمدی و به‌موقع بودن، هم‌خوانی و مقایسه پذیری، دسترسی‌پذیری و تفسیرپذیری) به‌جز مؤلفه‌ی درستی، در این مرحله از ارزیابی صادق هستند که در ادامه موردبررسی قرار می‌گیرند. در واقع، بعد کمی درستی بیشتر به خود اقلام داده‌ای اشاره دارد تا به کیفیت ثبت.

به برون‌داده‌های آماری می‌شود. این مدل به‌گونه‌ای طراحی شده است که مستقل از منبع داده باشد، بنابراین می‌توان از آن برای توصیف و ارزیابی کیفیت فرایندهای مبتنی بر آمارگیری‌ها، سرشماری‌ها، رکوردهای اداری و سایر منابع استفاده کرد. لوتیان و همکاران نیز نیاز به درک تمام مراحل فرایند تولید آماری را هنگام استفاده از منابع داده‌های جایگزین مانند داده‌های اداری را مورد تأکید قرار داده‌اند [۱۱]. به‌این‌ترتیب، ارزیابی کیفیت منابع اداری مستلزم بررسی خطاهایی است که ممکن است قبل و بعد از ارائه‌ی داده‌ها به NSOها و تعیین چگونگی کاهش چنین خطاهایی رخ داده‌اند (مانند تغییرات در گردآوری، پردازش و/یا یکپارچه‌سازی با منابع دیگر)؛ به‌عبارت‌دیگر، در این مطالعه علاوه بر اقتباس از GSBPM، از مدل ارائه‌شده در [۹] نیز استفاده‌شده است که در این چارچوب سه دیدگاه مختلف در مورد کیفیت، یعنی دیدگاه منبع، فراداده‌ها و داده‌ها در نظر گرفته‌شده‌اند. دیدگاه منبع بر جنبه‌های ضروری کیفیت برای تحویل منبع داده‌ها تمرکز دارد. درحالی‌که دیدگاه فراداده‌ها بر جنبه‌های فراداده‌های منبع داده‌ها تمرکز دارد. در دیدگاه داده‌ها، جنبه‌های فنی و درستی مرتبط با کیفیت داده‌ها موردبررسی قرار می‌گیرد. هر یک از این دیدگاه‌ها شامل چند بعد کیفیت داده‌ها است که هر یک از طریق نماگرهای کیفیت ارزیابی می‌شوند. مطالب ارائه‌شده در این پژوهش مطابق با رویکردهای ارائه‌شده، به معرفی ابعاد کیفیت، نماگرها و روش‌های مورداستفاده در ارزیابی داده‌های اداری با تمرکز ویژه بر سرشماری می‌پردازد. در واقع، در این مطالعه چهار مرحله‌ی تولید سرشماری معرفی شده‌اند: منبع، داده‌ها، فرایند و برون‌داد. سپس با شناسایی ابعاد کلیدی کیفیت در هر مرحله و ابزارها و نماگرهای مربوط برای ارزیابی کیفیت، چگونگی ارزیابی کیفیت داده‌های اداری تشریح می‌شود.

۴ کیفیت منابع

در انتقال از رویکرد سنتی به روش‌شناسی سرشماری ثبتی مبنا (ترکیبی)، لازم است کیفیت ثبت‌های اداری در اولین مرحله‌ی ممکن ارزیابی شوند. یکی از پیش‌نیازهایی که NSO باید قبل از تکمیل طرح سرشماری در نظر داشته باشد، تحقیق درباره‌ی امکان استفاده از هر ثبتی است؛ اما حتی در مواردی که داده‌های یک ثبت خاص قبلاً استفاده‌شده باشند، ارزیابی مجدد ثبت باید قبل از هر سرشماری انجام شود زیرا ممکن است داده‌ها از آن زمان تاکنون به‌روز شده باشند و از نظر مفاهیم، رده‌بندی‌ها، پوشانش و روش‌های گردآوری دچار تغییراتی شده باشند.

¹¹ Objects

۱۰۴ مرتبط بودن

تحت تأثیر هرگونه تأخیر بین تاریخ وقوع یک رویداد و تاریخ ذخیره‌ی آن در ثبت قرار گیرد. در صورت امکان، NSOها باید متولیان ثبت را برای ثبت هر دو تاریخ ترغیب کنند.

به موقع بودن یا هماهنگی بین تاریخ مورد انتظار و تاریخ واقعی تحویل داده‌ها نیز موضوعی است که باید مورد توجه قرار گیرد زیرا NSO معمولاً مسئولیت تولید بروندهای سرشماری را بر اساس برنامه‌ی زمانی توافق شده بر عهده دارد و نمی‌خواهد هیچ تأخیری در تولید داده‌های سرشماری بر این امر تأثیر بگذارد. بنابراین، هنگام ارزیابی روزآمدی و به موقع بودن منبع داده‌ها باید عوامل کلیدی زیر را در نظر گرفت:

- **تأخیر زمانی در منبع داده‌ها:** علاوه بر تلاش برای تحویل سریع، فراوانی نگهداری رکورد و بهنگام سازی داده‌ها توسط متولیان منبع می‌تواند یک عامل کلیدی در دستیابی به روزآمدی باشد. عنصر دیگری که مطرح است، تعهدات و انگیزه‌های فردی برای ثبت یا لغو ثبت در سازمان‌ها است (مانند مواردی که تغییر اقامتگاه معمولی وجود دارد).
- **تطابق دسترسی به داده‌ها با الزامات سرشماری:** هنگام بررسی سرشماری‌های مبتنی بر داده‌های اداری، اگر زمان عرضه با تاریخ انتشار موردنظر مطابقت نداشته باشد، مشکلات اساسی در کیفیت داده‌ها به وجود می‌آید و
- **تاریخ احتمالی دریافت داده‌ها:** تعیین تاریخ‌های دسترسی به داده‌های منابع اداری به پردازش و قالب داده‌های ارائه‌شده توسط سازمان‌ها بستگی دارد.

۳۰۴ هم‌خوانی و مقایسه پذیری

ارزیابی میزان موفقیت پیوند یک منبع اداری و ترکیب آن با سایر منابع داده‌ها برای استفاده در سرشماری حائز اهمیت است. اطلاعات ارزیابی مرتبط بودن از جمله اطلاعات در مورد تفاوت‌ها در مفاهیم اساسی، تعاریف، رده‌بندی‌ها و روش‌های گردآوری نیز می‌تواند در ارزیابی هم‌خوانی استفاده شود. همان‌طور که پیش‌ازاین نیز ذکر شد، داده‌هایی که در ثبت‌های اداری نگهداری می‌شوند به دلیل تغییر سیاست، قانون و دستورالعمل‌ها در طول زمان و مناطق مختلف تغییر می‌کنند. چنین تغییراتی می‌تواند بر مفاهیم، تعاریف، رده‌بندی‌ها، پوشش یک منبع و تمام نماگرهای اندازه‌گیری و نمایندگی در ذیل بعد مرتبط بودن تأثیرگذار باشد. این موضوع برای سرشماری اهمیت ویژه‌ای دارد، زیرا

درک تفاوت‌های بین جامعه‌ی اداری و جامعه‌ی سرشماری موردنظر و معیارها یا متغیرهای موجود در منبع اداری و مشخصه‌های موردنظر سرشماری برای ارزیابی مرتبط بودن^{۱۲} ضروری هستند. خطاهای ناشی از این تفاوت‌ها، به ترتیب به‌عنوان خطای نمایندگی^{۱۳} و خطای اندازه‌گیری نامیده می‌شوند.

خطاهای نمایندگی: در منبع داده‌ی اداری، «شیء» عنصر پایه از جامعه‌ای است که اطلاعات آن گردآوری می‌شود. شیء می‌تواند یک فرد، خانوار یا سکونت‌گاه باشد. خطاهای نمایندگی، خطاهایی هستند که به واحد جامعه‌ی موردنظر که گاهی اوقات به‌عنوان جامعه‌ی هدف نامیده می‌شود، مربوط می‌شوند. NSO در ارزیابی ثبت، باید مطلع شود که آیا ثبت می‌تواند اطلاعات مربوط به جامعه‌ی هدف را فراهم کند یا نه. همچنین برای ارزیابی مرتبط بودن، NSO باید تعیین کند که آیا مجموعه‌ای از اشیاء در یک منبع داده‌ی اداری با جامعه‌ی سرشماری موردنظر (جامعه‌ی هدف) مطابقت دارند یا نه. دو نماگر کلیدی نمایندگی برای برقراری بعد مرتبط بودن یک منبع اداری برای کمک به ارزیابی آن به‌صورت زیر پیشنهاد می‌شود:

- تطابق اشیاء در ثبت با واحدهای هدف در سرشماری و
- پوشش مجموعه‌ی اشیاء در مقایسه با جامعه‌ی هدف سرشماری،

خطاهای اندازه‌گیری: به میزان عدم مطابقت اقلام داده‌ای در ثبت با مفاهیم و تعاریف موضوعات سرشماری اشاره دارد. بنابراین، NSO باید تعیین کند که آیا اطلاعات گردآوری‌شده در مورد اشیاء نیازهای سرشماری را با توجه به مفاهیم هدف (مانند تابعیت، وضعیت شغلی، اندازه‌ی خانوار و غیره) برآورده می‌کند یا نه.

۲۰۴ روزآمدی و به‌موقع بودن

حتی اگر یک ثبت اداری بتواند دوره‌ی زمانی مربوط به سرشماری را پوشش دهد، برای قابل‌استفاده بودن داده‌ها در برنامه‌ی زمانی سرشماری، باید آن‌ها در دسترس بوده و به‌موقع به NSO تحویل داده شوند. لازم است که تفاوت بین تاریخ مرجعی که داده‌ها به آن اشاره دارند و تاریخی که داده‌ها به NSO تحویل داده می‌شوند، به حداقل برسد زیرا هر چه این تأخیر طولانی‌تر باشد، حتی در صورت درستی داده‌ها، آن‌ها کمتر مرتبط خواهند بود. با این وجود، درستی ممکن است

¹²Relevance

¹³Representation Error

- محدودیت در نحوه‌ی استفاده از داده‌ها،
- محدودیت در نگهداشت داده‌ها^{۱۵} و قاعده‌های حذف یا امحاء داده‌ها و
- قاعده‌های مربوط به روش‌های افشای اعمال‌شده توسط NSO که بر بروندهای سرشماری تأثیرگذار است.

سطح پذیرش عمومی: تصمیم NSO برای دسترسی و استفاده از یک منبع داده‌ی خاص در سرشماری، می‌تواند به پذیرش عمومی بستگی داشته باشد. بنابراین، NSO باید در استفاده از منابع اداری در سرشماری شفاف باشد، مزایای آن را برای عموم مردم برجسته کند و در عین حال حریم خصوصی و امنیت را تضمین کند.

سهولت انتقال داده‌ها به NSO: متولیان داده‌ها ممکن است مدل‌ها، قالب‌ها، طرح‌ها، نرم‌افزارها و سخت‌افزارهای بسیار متفاوتی را با آنچه NSO با آن آشنایی دارد، اتخاذ کنند. این موارد شامل نحوه‌ی نگهداری و انتقال داده‌ها، از جمله اقدامات امنیتی برای انتقال است. ساختار داده‌ها نیز می‌تواند پیچیده و اندازه‌ی فایل‌ها بسیار بزرگ باشد. آگاهی NSO از چنین تفاوت‌ها و پیچیدگی‌هایی برای ارزیابی امکان دریافت و انتقال مجموعه داده‌ها به سامانه‌های NSO، امری ضروری است.

وجود فراداده‌های واضح و جامع: ارزیابی تفسیرپذیری به وجود دسترسی به مستندات و فراداده‌های واضح و جامع در مورد منبع اداری مربوط می‌شود. بدون فراداده‌ها، درک و ارزیابی منبع اداری در ارتباط با کاربرد مدنظر مقدور نیست. فراداده‌ها باید حاوی جزئیاتی درباره‌ی مفاهیم زیر باشند:

- سازمان اداری و هدف گردآوری،
- مفاهیم، تعاریف، رده‌بندی‌ها و پروتکل‌های مورداستفاده،
- فرایندها و روش‌های گردآوری، پردازش، اعتبارسنجی و تضمین کیفیت و
- متغیرها و واحدهای گزارش‌دهی از جمله واژه‌نامه‌های داده‌ها، ساختار فایل‌ها، قالب‌ها و روابط درون‌داده‌ای.

اغلب اوقات، فراداده‌های واضح و کامل برای تمام جنبه‌های منبع اداری در مرحله‌ی اولیه‌ی بررسی منبع برای استفاده توسط NSO موجود نیستند و همکاری با متولیان داده‌ها به‌منظور ایجاد فراداده‌های مرتبط ضروری خواهد بود. این موضوع به ارتباط مناسب با متولیان داده‌ها و تمایل آن‌ها برای همکاری با NSO بستگی دارد که در بخش بعدی بیشتر موردبررسی قرار خواهد گرفت.

ثبات در طول زمان یک مسئله‌ی کلیدی است. در ارزیابی فقدان هم‌خوانی یا مقایسه پذیری، NSO باید بداند چه تغییراتی در ثبت در طول زمان یا چه تفاوت‌هایی در دامنه‌ها (مانند مناطق جغرافیایی یا زیرگروه‌های جمعیتی) ایجادشده است که می‌تواند بر موارد زیر تأثیرگذار باشد:

- تعریف و پوشانش واحد موردنظر در سرشماری،
- مفاهیم، تعاریف و رده‌بندی‌های مرتبط با متغیرهای منبع اداری موردنظر در سرشماری و
- گردآوری داده‌ها، پردازش و روش‌های تضمین کیفیت ثبت برای به‌کارگیری در اهداف سرشماری.

یکی دیگر از ملاحظات کلیدی در ارزیابی بعد هم‌خوانی و مقایسه پذیری منبع داده‌ها، سهولت اتصال داده‌ها در هر ثبت با سایر مجموعه داده‌های مرتبط با سرشماری، به‌طور مثال، از طریق استفاده از شماره‌های شناسایی یکتا یا ترکیب سایر متغیرهای شناساننده مانند نام، جنس و تاریخ تولد است.

۴.۴ دسترسی پذیری و تفسیرپذیری

دسترسی‌پذیری معمولاً به‌عنوان سهولت دسترسی کاربران به داده‌های سرشماری تعریف می‌شود و شامل مواردی مانند سهولت تشخیص ماهیت داده‌ها و فراداده‌ها توسط کاربر و همچنین مناسب بودن قالب ارائه‌ی این اطلاعات به کاربر است. دسترسی‌پذیری در سرشماری ثبتی مبنای سهولت NSO در دریافت کامل داده‌های ورودی از متولیان داده‌ها اشاره دارد. تفسیرپذیری به دسترسی هرگونه اطلاعات تکمیلی یا فراداده‌ها مربوط می‌شود که برای کمک به کاربران در تفسیر، درک و استفاده از داده‌های منتشرشده ضروری است. نماگرهایی که در ادامه می‌آید برای ارزیابی بعد دسترسی‌پذیری و تفسیرپذیری در مرحله‌ی ارزیابی کیفیت منبع پیشنهاد می‌شود:

محدودیت‌های ارائه و استفاده از داده‌ها برای NSO: شناسایی هرگونه محدودیتی که ممکن است بر توانایی NSO در دریافت و استفاده از یک منبع اداری تأثیرگذار باشد، مهم است. این محدودیت‌ها می‌تواند شامل موارد زیر باشد:

- محرمانه‌سازی رکوردها یا متغیرها،
- کنترل افشا (پیش از تحویل)، مانند کدگذاری شناسانه‌ها، روش پریشیدگی^{۱۴}، کدگذاری داده‌های ارائه‌شده،

¹⁴Perturbation Method

¹⁵Retention of Data

۵.۴ محیط سازمانی

منع نیز باید در نظر گرفته شود. استانداردهای کیفیت متولیان: NSO باید ارزیابی کند که آیا متولیان می‌توانند انتظارات کیفیت داده‌های ارائه شده به NSO را برآورده کنند یا نه. این ارزیابی شامل اطلاعات مربوط به اصول، استانداردها و دستورالعمل‌های اتخاذ شده توسط متولیان داده‌ها برای تضمین کیفیت، از جمله روش گردآوری، پردازش و ارائه داده‌ها به NSO است.

در حرکت به سوی روش مبتنی بر ثبت، NSO برای گردآوری، پردازش و تحویل داده‌های اداری باکیفیت مورد انتظار و بر اساس برنامه‌ی زمانی توافق شده، وابسته به متولیان داده‌ها است. از این رو، نماگرهای زیر برای بعد محیط سازمانی در مرحله‌ی ارزیابی کیفیت منبع پیشنهاد می‌شود:

قوت رابطه‌ی متقابل: باید فرایندهایی را برای مدیریت ارتباط با متولیان داده‌ها در نظر گرفت و نسبت به استمرار تعامل اطمینان حاصل کرد. این فرایندها باید شامل سازوکارهایی در موارد زیر باشند:

- اطلاع‌رسانی الزامات NSO به متولیان،

- اطلاع‌رسانی بهنگام هرگونه تغییرات اثرگذار بر منبع داده‌ها توسط متولیان (مانند مبنای قانونی درباره‌ی داده‌ها، مفاهیم، رده‌بندی‌ها و فرایندها و روش‌های گردآوری، مدیریت و ارائه‌ی داده‌ها)،

- طرح هرگونه پرسش با متولیان درباره‌ی منبع داده‌ها و

- ارائه‌ی بازخورد درباره‌ی یافته‌ها به متولیان که می‌تواند منجر به بهبود منبع شود.

تجربیات پیشین: NSO باید میزان پاسخ‌گویی متولیان به درخواست‌های پیشین در ارتباط با داده‌ها را ارزیابی کند، خواه مشکلاتی (مانند تحویل دیر هنگام، کیفیت پایین داده‌ها) در ارائه‌ی پیشین داده‌ها وجود داشته باشد یا نه. **توافقات رسمی:** NSO باید به دنبال ایجاد یا توسعه‌ی توافق‌نامه‌های مکتوب در موارد باشد:

- نقش‌ها و مسئولیت‌های NSO و متولیان، از جمله این‌که آیا NSO نقشی در تأیید هرگونه تغییر در منبع اداری مورد استفاده در سرشماری دارد یا خیر،

- مبنای قانونی برای ارائه‌ی داده‌ها و هرگونه الزامات امنیتی یا محرمانگی و

- مشخص‌سازی الزامات، مطابق با تفاهم‌نامه‌ی تحویل یا اشتراک‌گذاری داده‌ها.

وضعیت متولیان: مخاطره‌ی مرتبط با وضعیت متولی داده‌ها باید، با توجه به این‌که آیا آن‌ها سازمانی شناخته شده، باثبات و معتبر هستند یا نه توسط NSO ارزیابی شود. NSO باید در نظر داشته باشد که آیا مبنای قانونی یا نظارتی برای عملکرد اداری متولی وجود دارد که موجب حصول اطمینان از پایداری و کیفیت منبع شود یا نه. مخاطرات مرتبط با پیچیدگی سازمان (های) متولی مؤثر در گردآوری، پردازش و تحویل

۵ کیفیت داده‌های ورودی

کیفیت داده‌های ثبتی مورداستفاده در سرشماری می‌تواند تحت تأثیر عوامل متعددی قرار گیرد. علاوه بر عواملی که ممکن است به‌طور کلی بر کیفیت ثبت تأثیرگذار باشند، عوامل دیگری نیز بر کیفیت متغیرها و واحدهای داده‌های ارائه شده تأثیرگذار خواهند بود که ارزیابی آن‌ها در این بخش خلاصه می‌شود.

۱.۵ درستی و قابلیت اعتماد

همان‌طور که پیش‌از این نیز ذکر شد، در ارزیابی درستی داده‌های ورودی، NSO باید بین خطاهای نمایندگی (خطاهای مربوط به پوشانندگی جامعه‌ی هدف) و خطاهای اندازه‌گیری (خطاهای مربوط به متغیر خاص موردنظر) تمایز قائل شود. خطاهای نمایندگی زمانی رخ می‌دهند که واحدها به درستی یا به سرعت (به‌عنوان مثال، به دلیل تأخیر در واردسازی یا حذف از ثبت) به سازمان اداری گزارش نشوند یا برخی از رکوردهای داده‌ها به دلیل مشکلات فنی به NSO ارسال نشوند. میزان چنین خطاهایی را می‌توان با برآورد سطح بیش‌پوشانندگی (موارد تکراری) یا کم‌پوشانندگی (موارد گمشده) در مقایسه با سایر منابع داده‌ها ارزیابی کرد. به‌عنوان مثال، ممکن است مقایسه‌ی ثبت جمعیت و تأمین اجتماعی نشان دهد که گرچه افرادی در ثبت تأمین اجتماعی به‌عنوان متوفی شناسایی شده‌اند، اما هنوز در ثبت جمعیت کشور ارائه می‌شوند. نماگرهای پایه که می‌تواند برای ارزیابی خطاهای نمایندگی به کار رود، عبارت‌اند از:

- تعداد کل واحدهای دریافتی (برای مقایسه با تعداد مورد انتظار)

و

- درصد واحدهای تکراری.

ارزیابی جامعی از کم‌پوشانندگی و بیش‌پوشانندگی داده‌ها را می‌توان با محاسبه و مقایسه‌ی تعداد کل واحدها و همچنین جدول‌بندی‌های تقاطعی فراوانی یا درصد مشخصه‌های کلیدی (مانند جنس، سن و

را می‌توان به‌سادگی با هر تأخیری که در دریافت داده‌ها توسط NSO از آنچه در ابتدا با متولی ثبت توافق شده بود، ارزیابی کرد.

جغرافیا) در سطح کل، بین منبع اداری و سایر منابع (که به‌عنوان مرجع یا استاندارد طلایی تطبیقی در نظر گرفته شده‌اند) انجام داد. نماگر کلیدی در ارزیابی کم‌پوشانسی عبارت‌اند از:

- درصد واحدهای منبع (ثبت) مرجع که در منبع اداری ارائه شده گمشده هستند.
- همچنین، بیش‌پوشانسی را می‌توان با نماگرهای زیر ارزیابی کرد:
- درصد واحدهای منبع که در جامعه‌ی ثبت لحاظ نشده است و
- درصد واحدهای منبع که متعلق به جامعه‌ی هدف ساکن موردنظر NSO نیستند.

۶ کیفیت فرایند

فرایندهای تبدیل داده‌های اداری برای کاربرد در سرشماری معمولاً شامل موارد زیر است:

- اتصال داده‌های منابع اداری مختلف،
- یکپارچه‌سازی در یک فایل یا ثبت آماری،
- موارد تکراری و رفع مغایرت، بهنگام سازی ثبت‌ها با استفاده از رویکرد نشانه‌های زندگی^{۱۶} (SOL)،
- پاک‌سازی و ویرایش (برای اصلاح خطاهای موجود در داده‌ها) و
- جانمایی (برای رفع ناسازگاری‌ها و مقادیر گمشده).

هر یک فرایندهای فوق، از طریق اطلاعات به‌دست‌آمده از مراحل کیفیت منبع و داده‌ها انجام می‌شوند. به‌عنوان مثال، اگر در مرحله‌ی کیفیت داده‌ها مشخص شود که یک منبع داده دارای مقادیر نادرست یا گمشده است، ممکن است جانمایی یا ویرایش داده‌ها برای استفاده از آن‌ها در سرشماری لازم باشد. البته، این فرایندها برای بهبود کیفیت داده‌ها و مناسب بودن برون‌داده‌ها برای هدف موردنظر طراحی شده‌اند. با این وجود، پردازش داده‌ها به‌خودی‌خود ممکن است میزانی از عدم قطعیت در داده‌ها را ایجاد کند که عدم کنترل این موضوع می‌تواند منجر به کاهش کیفیت برون‌داد شود. تغییر کیفیت ناشی از اجرای چنین فرایندهایی را «کیفیت فرایند» می‌نامند. این بخش بر برخی از ملاحظات کلیدی پیرامون این فرایندها تمرکز دارد که ممکن است بر درستی و کامل بودن داده‌ها تأثیرگذار باشد.

۱۰۶ درستی و اتصال رکورد

اگر NSO برای مجموعه داده‌های سرشماری به بیش از یک منبع اداری متکی باشد، لازم است که داده‌های رکورد واحد از منابع مختلف به یکدیگر متصل شوند. میزان پیوندپذیری بر درستی داده‌های ورودی تأثیرگذار است. کیفیت اتصال داده‌های رکورد واحد از منابع مختلف به دسترسی، یکتایی و وجود شناسانه‌های مشترک بستگی دارد. برای

نماگرهای فوق منوط به دو فرض کلیدی هستند. اولاً، یک ثبت پایه‌ی کامل (ثبتي) که فرض شده است بیشترین پوشانسی مناسب را از جامعه‌ی هدف ساکن دارد) برای محاسبه‌ی بیش‌پوشانسی در دسترس باشد. ثانیاً، باید مشخص شود که کدام واحدهای ثبت پایه‌ی کامل در داده‌های ورودی باید برای محاسبه‌ی کم‌پوشانسی در نظر گرفته شوند. رویکرد دیگر در ارزیابی کم‌پوشانسی و بیش‌پوشانسی، اجرای یک آمارگیری نمونه‌ای است که دو روش ممکن را برای مقایسه ارائه می‌دهد: (الف) اجرای یک آمارگیری پوشانسی مستقل برای اندازه‌گیری کم‌پوشانسی و (ب) واری نمونه‌ای از منبع ارائه‌شده در میدان برای اندازه‌گیری بیش‌پوشانسی. برای ارزیابی خطاهای اندازه‌گیری، NSO باید برای هر متغیر خاص قادر به برآورد تعداد و درصد مقادیر گمشده و مقادیر خارج از محدوده یا نامعتبر (مانند سن ثبت‌شده‌ی ۱۲۰ سال) باشد. فنون و معیارهای آماری، مانند توزیع‌های فراوانی، می‌توانند الگوهای غیرمنتظره و نقاط دورافتاده را که نشان‌دهنده‌ی خطاهای اندازه‌گیری سیستماتیک هستند، آشکار سازند.

۲۰۵ روزآمدی و به‌موقع بودن

با مقایسه‌ی تاریخ مرجع، تاریخ تحویل تعیین شده و تاریخ تحویل واقعی داده‌ها می‌توان معیارهای روزآمدی را در ارزیابی کیفیت داده‌های ورودی تعیین کرد. بنابراین، دو جنبه از روزآمدی را می‌توان از طریق موارد زیر ارزیابی کرد:

- تفاوت بین ثبت هرگونه تغییر در داده‌های منبع توسط متولی ثبت و تاریخ رخداد واقعی تغییر در جامعه و
- تفاوت بین تاریخ دریافت داده‌ها توسط NSO و تاریخ دوره‌ی مرجعی که داده‌ها به آن اشاره دارند، درحالی‌که به‌موقع بودن

¹⁶Signs Of Life

متداولترین روشها برای ارزیابی کیفیت اتصال عبارتاند از:

- تعیین نسبت رکوردهای پیوندی و غیر پیوندی،
- برآورد نرخ مثبت نادرست و نرخ منفی نادرست با استفاده از بازبینی دفتری نمونه‌ای از رکوردهای پیوندی. باید توجه داشت که بازبینی دفتری تنها در شرایطی که داده‌ها با استفاده از الگوریتم چکیده‌سازی به دست نیامده‌اند، قابل اجرا است. در صورت چکیده‌سازی داده‌ها، NSO باید در جهت دسترسی به نمونه‌ای از رکوردهای پیوندی با تمام اطلاعات اصلی داده‌ها به منظور ارزیابی اتصال، تلاش کند و
- مقایسه‌ی توزیع ویژگی‌های رکوردهای پیوندی و غیر پیوندی، برای مثال با متغیرهایی مانند سن، جنس و قومیت. تفاوت در ویژگی‌ها به برخی از اریبی‌های ناشی از خطای اتصال اشاره دارند که این مسئله به معنای عدم اتصال برخی از انواع خاص رکوردهای واحد به علت دشوار بودن اتصال آن‌ها است. صرف نظر از این که یک کلید یا متغیر شناسانه‌ی یکتا در دسترس است یا این که چندین متغیر به صورت ترکیبی برای شناسایی جورها در فرایند اتصال استفاده می‌شود، باید نماگرهایی برای ارزیابی کیفیت متغیرهای اتصال وجود داشته باشد، مانند:
- تعداد و درصد کلیدهای اتصال تکراری که می‌تواند برای محاسبه‌ی یک متغیر شناسایی تکی (مانند شماره شناسایی فردی) یا ترکیب متغیرهایی که کلید اتصال را ارائه می‌دهند (مانند سن، تاریخ تولد و نشانی)، به کار رود و
- نماگرهای درستی (از جمله عدم وجود کلیدهای اتصال یا تعداد مقادیر گمشده یا غیرقابل قبول) ممکن است میزان تصادفی نبودن خطاهای اندازه‌گیری را به عنوان مثال در مواردی که شواهدی از هرگونه اریبی سیستماتیک وجود دارد (، نشان دهد. بنابراین، محاسبه‌ی این نماگرها بر اساس مناطق جغرافیایی و زیرگروه‌های جمعیتی حائز اهمیت است.
- در نهایت، در مواردی که متغیرهای اتصال به صورت رمزگذاری شده یا چکیده به NSO ارائه شده باشند، باید تأیید شود که الگوریتم چکیده‌سازی انجام شده توسط متولی داده‌ها با الگوریتم چکیده‌سازی مورد استفاده در NSO مطابقت دارد، در غیر

کشورهایی که یک شناسانه‌ی یکتای مشترک را در ثبت‌های اداری خود اتخاذ می‌کنند، این فرایند نسبتاً ساده است. چنین شناسانه‌ای، با تسهیل فرایند ارزیابی کامل بودن و درستی، تلاش لازم برای اتصال داده‌ها را کاهش می‌دهد. برای ارزیابی هم‌خوانی داده‌های حاصل از منابع مختلف، NSO باید تعداد و درصد رکوردهای ارائه شده را در مواردی که چنین کلیدی وجود ندارد، تعیین کرده و اطمینان حاصل کند که خطای سیستماتیک وجود ندارد. در مواردی که چنین شناسانه‌ای وجود ندارد، کیفیت اتصال می‌تواند یک نگرانی بزرگ باشد. در چنین مواردی، اتصال رکورد با استفاده از کلیدهای جور (متغیرهایی که در واحدهای هر منبع داده مشترک هستند مانند نام، تاریخ تولد، جنس و نشانی) ممکن است امکان‌پذیر باشد، اما معمولاً زمان‌بر است و نتایجی با عدم قطعیت بیشتری به همراه دارد. در این مورد، NSO باید اطمینان حاصل کند که متغیرهای اتصال در تمام منابع از کیفیت عالی برخوردارند، در غیر این صورت کیفیت اتصال‌های بعدی و در نتیجه قابلیت اعتماد داده‌ها دچار مشکل خواهند شد. مخاطره‌ی جورهای نادرست^{۱۷} (و برعکس، ناجورهای نادرست^{۱۸}) باید ارزیابی شود (برای جزئیات بیشتر به [۷] مراجعه شود). کلیدهای جور می‌توانند با استفاده از روش قطعیتی^{۱۹} (در مواردی که مجموعه‌ی یکسانی از کلیدهای جور برای هر مجموعه داده‌ی پیوند داده شده وجود دارد) یا با روش‌های احتمالاتی (در مواردی که جورهای دقیق امکان‌پذیر نیست و اتصال‌ها بر اساس مقایسه و واجد شرایط بودن شباهت نسبی رکوردها در بین متغیرهای شناسایی ایجاد می‌شوند) به یکدیگر متصل شوند. لازم است که این متغیرها درست و کامل باشند تا از کیفیت بالای اتصال رکورد اطمینان حاصل شود. درستی اتصال را می‌توان از طریق تفاوت بین ویژگی‌های رکوردهای پیوندی و غیرپیوندی ارزیابی کرد. یکی از روش‌های دیگر اتصال که می‌تواند پس از اعمال روش‌های قطعیتی و احتمالاتی برای رکوردهای غیر پیوندی اعمال شود، اتصال دفتری^{۲۰} است که شامل بازرسی بصری رکوردهای غیر پیوندی است. انجام این کار می‌تواند پرزحمت و زمان‌بر باشد و علاوه بر این، در شرایطی که داده‌ها با استفاده از الگوریتم چکیده‌سازی^{۲۱} به دست می‌آیند، امکان‌پذیر نیست. خطای اتصال می‌تواند از طریق رکوردهای غیر پیوندی که باید پیوند داده می‌شدند (اصطلاحاً منفی‌های نادرست) و رکوردهای پیوندی که نباید پیوند داده می‌شدند (اصطلاحاً مثبت‌های نادرست)، رخ می‌دهد.

¹⁷False Matches

¹⁸False Non-Matches

¹⁹Deterministic Method

²⁰Clerical Linkage

²¹Hashing Algorithm

۳.۶ ارزیابی پوشانش در ثبت‌های آماری جمعیت با استفاده از آمارگیری‌های مستقل

علاوه بر رویکرد نشانه‌های زندگی، روش‌های دیگری نیز برای برآورد پوشانش (و همچنین خطای محتوای) سرشماری‌ها وجود دارد. این روش‌ها عبارت‌اند از: فنون تضمین کیفیت مانند واری‌های سازگاری درونی، تحلیل‌های جمعیت شناختی، مقایسه با داده‌های سایر منابع از جمله سرشماری‌های پیشین، آمارگیری‌های خانواری جاری و آمارگیری‌های پس‌سرشماری. چنین روش‌هایی، صرف‌نظر از روش‌شناسی سرشماری اتخاذشده، موردنیاز هستند اما باید بر نقشی که یک آمارگیری پس‌سرشماری می‌تواند در ارزیابی کیفیت ثبت آماری ایفا کند، تأکید کرد. آمارگیری پس‌سرشماری، بازسازی کامل نمونه‌ی نمایانگر جامعه‌ی سرشماری است که پس‌از آن جورسازی افراد شمارش‌شده در آمارگیری پس‌سرشماری با داده‌های گردآوری‌شده‌ی سرشماری انجام می‌شود. با استفاده از فرایندهای برآورد نظام دوگان^{۲۲} (DSE)، نتایج مقایسه عمدتاً برای اندازه‌گیری خطای پوشانش (و محتوا) در سرشماری استفاده می‌شود. آمارگیری باید به‌گونه‌ای طراحی شود که برآوردهای قابل‌اعتماد کم‌شماری یا بیش‌شماری برای مناطق جغرافیایی زیرملی و همچنین کل کشور به دست آید. علاوه بر این، برخی برآوردها را می‌توان از طریق پساتبقة‌بندی گروه‌های زیرجمعیتی برحسب جنس و گروه سنی انجام داد. باین‌وجود، لازم است که برای جلوگیری از خطاهای سیستماتیک مشابهی که ممکن است در داده‌های سرشماری وجود داشته باشد، آمارگیری پس‌سرشماری مستقل از مدیریت ثبت آماری اجرا شود. به‌طورکلی، مراحل اصلی اندازه‌گیری و تعدیل تعداد افراد و خانوارهایی که در سرشماری شمارش نشده‌اند یا در مکان اشتباهی شمارش شده و در ثبت آماری ثبت‌شده‌اند، عبارت‌اند از:

- جورسازی رکوردهای آمارگیری پس‌سرشماری با رکوردهای ثبت‌شده در ثبت آماری با به‌کارگیری ترکیبی از جورسازی خودکار و جورسازی دفتری،
- استفاده از رکوردهای جورشده و رویکرد برآورد نظام دوگان برای برآورد تعداد افراد و خانوارهای گمشده در سرشماری و آمارگیری پس‌سرشماری،
- جستجوی موارد تکراری در ثبت و آمارگیری پس‌سرشماری برای برآورد سطح بیش‌شماری (مواردی که بیش از یک‌بار شمارش شده‌اند)،

این صورت، اتصال داده‌های ارائه‌شده به سایر منابع داده‌ها امکان‌پذیر نخواهد بود و مرتبط بودن داده‌ها تضعیف می‌شود.

۲.۶ ارزیابی بیش‌پوشانشی در ثبت‌های آماری جمعیت با استفاده از نشانه‌های زندگی

در مرحله‌ی کیفیت فرایند، NSOها باید توجه ویژه‌ای به ارزیابی (و کاهش) بیش‌پوشانشی داشته باشند زیرا این امر به‌شدت بر بعد درستی تأثیرگذار است. روش نشانه‌های زندگی به‌عنوان ابزاری برای به حداقل رساندن بیش‌پوشانشی مورد استفاده قرار می‌گیرد. در این روش تعدادی از قاعده‌های فعالیت اتخاذ می‌شود تا اطمینان حاصل شود که تنها افرادی که زنده هستند و مجموعه‌ای از معیارهای اقامت از پیش تعریف‌شده را برآورده می‌کنند، در سرشماری گنجانده شوند. موفقیت روش نشانه‌های زندگی به در دسترس بودن نماگرهای خوبی از نشانه‌های فعالیت در داده‌های اداری فردی یا ترکیبی متکی است. کاربرد این روش معمولاً شامل ایجاد برخی مفروضات است که مشخص می‌کند چه کسی فعال و چه کسی غیرفعال است. انتخاب نشانه‌های نماگرهای فعالیت (قاعده‌های تصمیم‌گیری) باید با ارزیابی کیفیت در مرحله‌ی منبع و داده‌های ورودی، از جمله مشاوره با متولیان داده‌ها، اعتبارسنجی متقابل بین منابع و در طول زمان و نظر کارشناسی انجام شود. برای مثال، در اسپانیا ([۱۶]) نشانه‌های زندگی با استفاده از داده‌های چندین منابع اداری از جمله ثبت مالیات، پرونده‌های تأمین اجتماعی و پایگاه داده‌ی ملی بیکاری و جابه‌جایی‌های شناسایی‌شده در ثبت جمعیت محلی تعیین می‌شوند. این منابع اداری در سطح فردی و خانواری به هم مرتبط هستند و افرادی که سطح آستانه‌ی نشانه‌های زندگی خاص که نشان‌دهنده‌ی حضور فعال آن‌ها در کشور است را به دست آوردند، در شمارش جمعیت گنجانده می‌شوند و بقیه‌ی افراد که غیرفعال شناخته‌شده‌اند، در شمارش در نظر گرفته نمی‌شوند. اگرچه استفاده از نشانگرهای نشانه‌های زندگی موجود در یک منبع داده ممکن است استفاده‌ی بهینه از داده‌هایی که قبلاً نگهداری شده‌اند را امکان‌پذیر کند، اما NSOها باید محتوای واقعی و پوشانش نشانگرها را از جنبه‌ی ملی به‌دقت بررسی کنند تا از سازگاری و هماهنگ‌سازی، به‌ویژه در مورد تواتر، روزآمدی و روش‌شناسی گردآوری داده‌ها اطمینان حاصل کنند.

²²Dual System Estimation

می‌شوند و سپس برای حل چنین مشکلاتی جانمایی می‌شوند و در نهایت برای اطمینان از کیفیت جانمایی اعتبارسنجی می‌شوند. پس از جانمایی کردن داده‌ها، ویرایش داده‌ها باید تکرار شود و اگر متعاقباً داده‌ها در طول هر فرایندی تغییر دیگری داشته باشند، به‌منظور حفاظت اطلاعات فردی واپایی افشای آماری دوباره تکرار شود [۱۲]. ویرایش و جانمایی ممکن است هم برای تک منبع و هم داده‌های یکپارچه موردنیاز باشد. ثبت‌های پایه عمدتاً برای ارائه‌ی اطلاعات در مورد شمارش جمعیت و اطلاعات پایه در مورد موضوعات سرشماری مربوط استفاده می‌شوند. سایر ثبت‌های تخصصی یا تکمیلی برای ارائه‌ی اطلاعات مربوط به متغیرهای سرشماری خاص و گاهی اوقات برای ارائه‌ی داده‌هایی که به‌طور کامل یا جزئی در ثبت‌های پایه گمشده هستند، استفاده می‌شوند. از ثبت‌های تخصصی برای اهداف اعتبارسنجی نیز استفاده می‌شود که از طریق آن مجموعه داده‌هایی که از ثبت پایه و سایر ثبت‌ها ترکیب شده‌اند، با جانمایی کردن بی‌پاسخی پرسش و مقادیر غیرقابل قبول بهبود می‌یابند، بدین ترتیب مجموعه داده‌های نهایی برای تولید بروندهای نهایی سرشماری ایجاد می‌شوند. کیفیت فرایندهای ویرایش و جانمایی باید به‌طور سیستماتیک پایش و ارزیابی شود. پیشنهاد می‌شود که مقدار اصلی، مقدار ویرایش شده، نوع خطا و داده‌های مربوط به واری‌های ناموفق برای تحلیل تغییرات ویرایش پایش شوند. تحلیل تغییرات برای هر مرحله‌ی ویرایش و جانمایی باید بر اساس اطلاعات زیر انجام شود:

- فراوانی و نوع ویرایش،
- میزان تغییرات و
- میزان فزونگی^{۲۳} در واری‌های سازگاری.

برای پایش روش‌های جانمایی و کاربرد قاعده‌های ویرایش می‌توان از نماگرهای زیر استفاده کرد:

۱. نرخ شکست ویرایش: این نرخ از تقسیم تعداد خطاهای برطرف‌شده بر تعداد کل خطاها (خطاهای برطرف‌شده و برطرف‌نشده) در فایل داده‌ها به دست می‌آید. این نماگر بین مقادیر ۰ و ۱ متغیر است. مقدار بالا در این نماگر به این معنی است که قاعده‌ی ویرایش به‌درستی به‌عنوان معیار خطایابی تعریف‌شده است و مقدار کم این نماگر به این معنی است که قاعده‌ی ویرایش با شکست مواجه شده و باید بهبود یابد،

۲. نرخ تعدیل: این نرخ از تقسیم تعداد خانوارها یا افراد ایجادشده تحت عنوان جانمایی خانوار غایب (در صورت استفاده) بر تعداد

- برآورد جمعیت‌های مناطق محلی برحسب سن و جنس با متعادل کردن بیش‌برآوردی و کم‌برآوردی با استفاده از ترکیب رگرسیون آماری و فنون برآورد کوچک‌ناحیه‌ای و
- جانمایی کردن افراد و خانوارهای گمشده در پایگاه داده‌ی سرشماری یا ثبت آماری.

۴.۶ روش‌های تصمیم‌گیری در مورد متغیرهای گزارش‌شده در منابع مختلف و تصحیح مغایرت

هنگام استفاده از داده‌های منابع مختلف، روش‌هایی برای ارزیابی کیفیت و در نتیجه اطلاع‌رسانی انتخاب متغیرها در مواردی که صفت مشابهی در ثبت‌های مختلف گزارش می‌شود، موردنیاز است. به‌عنوان مثال، در ثبت محلی اسپانیا وضع زناشویی قانونی افراد ثبت نمی‌شود. بنابراین برای تعیین این صفت، از داده‌های چندین ثبت مختلف استفاده می‌شود. از آنجاکه یک فرد ممکن است در منابع داده‌های مختلف با اطلاعات مغایر ظاهر شود، قاعده‌هایی مانند روش نشانه‌های زندگی لازم است تا مشخص شود کدام مقدار درست است. همچنین، ممکن است نیاز به رفع مشخصات متناقض نشانی‌های ثبت‌شده در ثبت‌های مختلف و شناسایی درست‌ترین اطلاعات وجود داشته باشد. چنین تفاوت‌هایی می‌تواند به‌عنوان مثال از تأخیر افراد در گزارش‌دهی تغییر نشانی یا گزارش‌دهی نشانی منزل دوم ناشی شود. اطلاعات نشانی متناقض (متعدد) و هر تصمیم مربوط می‌تواند منجر به کم‌پوشانشی در برخی مناطق و بیش‌پوشانشی در برخی مناطق دیگر شود. بنابراین NSO باید تصمیم بگیرد که اطلاعات سرشماری به کدام نشانی منتسب شود (به‌طور مثال نشانی که اخیراً گزارش شده باشد، می‌تواند در نظر گرفته شود).

۵.۶ ارزیابی کیفیت فرایند ویرایش و جانمایی

ویرایش و جانمایی یک فرایند تکراری برای بازبینی داده‌ها به‌منظور تصحیح هرگونه خطای ناشی از واردسازی داده‌های نامعتبر، ناسازگار یا غیرقابل قبول و جایگزینی این موارد و داده‌های گمشده با مقادیر قابل قبول است. در نتیجه انجام این فرایند، درستی و هم‌خوانی داده‌ها را بهبود می‌بخشد و آن‌ها را برای اهدافی که گردآوری شده‌اند، مناسب می‌سازد. در این فرایند، ابتدا مقادیر گمشده، نامعتبر یا ناسازگار ویرایش

²³Extent of Redundancy

کل خانوارها یا افراد به دست می‌آید،

۲۰۷ درستی

اندازه‌گیری درستی کل شمارش نهایی سرشماری به‌طور سنتی بر اساس برآورد نظام دوگان انجام می‌شود که معمولاً شامل اجرای آمارگیری پس‌سرشماری و سپس اتخاذ روش‌های گروباژگیر برای برآورد کم‌پوشانشی و بیش‌پوشانشی است. برآوردهای حاصل را می‌توان بر اساس داده‌های اداری در مورد تولدها، فوت‌ها و جریان‌های مهاجرت برای هرسال بین سرشماری‌های ده سالانه اصلاح کرد. علاوه بر این، برخی از NSOها به‌طور سنتی آمارگیری‌های نمونه‌ای با حجم کمتر برای واریسی کیفیت پس‌سرشماری را انجام می‌دهند که در آن داده‌ها در مورد تمام سؤالات سرشماری گردآوری می‌شوند و سپس با مجموعه داده‌های آماری جور می‌شوند تا خطای پاسخ‌گو اندازه‌گیری شود. در روش‌شناسی سرشماری ثبتی مبنای روش‌های سنتی برای تعیین کیفیت و پوشانش کل تا حدودی هنوز قابل اجرا هستند اما روش‌های جدیدی با زبینی‌شده نیز ضروری هستند. به‌طور ویژه، روش سنتی برآورد نظام دوگان به‌خوبی با بیش‌پوشانشی مقابله نمی‌کند، زیرا آریبی را نشان می‌دهد. علاوه بر این، پس از فرایندهای مختلف برای ارزیابی کیفیت در مراحل منبع، داده‌ها و فرایند، برخی از NSOها ممکن است راهبردی را اتخاذ کنند که نیازمند اتکای کمتری به رویکرد برآورد نظام دوگان برای ارزیابی پوشانش است و بیشتر بر کیفیت نگهداری و بهنگام سازی ثبت‌های جمعیت متکی است. به‌عنوان مثال، بسیاری از کشورها تلاش کرده‌اند تا روش‌هایی را توسعه دهند که هدف آن‌ها کاهش بیش‌پوشانشی از طریق روش‌های نشانه‌های زندگی است. فرایند قابل اندازه‌گیری دیگری که برای برون‌دادهای سرشماری قبل از انتشار عمومی آن‌ها اعمال می‌شود و ممکن است بر درستی داده‌ها نیز تأثیر بگذارد، شامل معیارهای اتخاذشده برای اطمینان از محرمانگی با محافظت کردن در برابر افشای آماری است. تضمین کیفیت داده‌ها و حفاظت از محرمانگی اطلاعات فردی از برخی جهات اهداف متضادی هستند، اما هر دو جنبه‌های ضروری برون‌داد سرشماری به حساب می‌آیند. شایان ذکر است که فارغ از روش‌های گردآوری داده‌ها، انتشار برون‌دادهای کاملاً درست (به‌ویژه برای کوچک ناحیه‌ها) و درعین حال در نظر گرفتن سطوح کافی مهار افشا برای اطمینان از حفاظت محرمانگی گرچه غیرممکن نیست اما دشوار است.

۳۰۷ روزآمدی و به‌موقع بودن

همان‌طور که پیش‌ازاین نیز ذکر شد، روزآمدی به فاصله‌ی زمانی بین تاریخ داده‌های سرشماری و تاریخ انتشار داده‌ها اشاره دارد. یکی از

۰۳. نرخ جانهدی: این نرخ نشان‌دهنده‌ی درصد رکوردهای جانهدی شده در کل رکوردها است. این نماگر باید برای هر متغیر محاسبه شود. نرخ‌های جانهدی، نماگرهای کلیدی کیفیت برای سرشماری هستند. جانهدی واحد باید از یک فرایند آماری استوار و شفاف پیروی کند و به‌خوبی مستندسازی شود و

۰۴. شاخص عدم تشابه: این شاخص نشان‌دهنده‌ی میزان تغییر دو توزیع (مشاهده‌شده و کل به انضمام مقادیر جانهدی‌شده) در سطح متغیر است. این شاخص به‌صورت زیر محاسبه می‌شود:

$$ID = \frac{1}{\sqrt{2}} \sum_{k=1}^K |f_{yk} - f_{yk}^*|,$$

که در آن k رسته‌های متغیر، f توزیع درصد متغیر قبل از جانهدی و f^* توزیع درصد متغیر بعد از جانهدی است. محدوده‌ی این شاخص از ۰ تا ۱۰۰ متغیر است که در آن ۱۰۰ به معنای حداکثر عدم تشابه است [۱۶].

۷ کیفیت برون‌داد

برای آن دسته از کشورهایی که به روش‌شناسی گردآوری داده‌های سرشماری ترکیبی یا ثبتی مبنای روی آورده‌اند، ارزیابی کیفیت برون‌داد برای تعیین این‌که آیا این گذار بر کیفیت کلی برون‌دادهای تأثیر گذاشته است یا نه بسیار مهم است. ارزیابی کیفیت برون‌داد نباید محدود به برآورد عدم حتمیت کل (بعد درستی) شود، بلکه باید شامل ارزیابی در تمام ابعاد دیگر کیفیت نیز باشد.

۱۰۷ مرتبط بودن

همان‌طور که پیش‌ازاین نیز ذکر شد، مرتبط بودن به میزان برآورد کردن نیازهای کاربران هم از نظر پوشانش و هم از نظر محتوا توسط برون‌دادهای سرشماری اشاره دارد. سه رویکرد رایج برای ارزیابی مرتبط بودن برون‌داد عبارت‌اند از:

- پرسش مستقیم از کاربران از طریق اجرای آمارگیری نیازهای کاربران (قبل از سرشماری) یا آمارگیری رضایت‌مندی کاربران (پس از سرشماری)،
- ایجاد سازوکارهای بازخورد کاربران در فرایند سرشماری و
- تحلیل کاربرد داده‌های سرشماری (به‌عنوان مثال از طریق بررسی گزارش‌ها و مقالات منتشرشده).

داده‌ها. از این رو، چنین مقایسه‌هایی باید با توجه به نتایج ارزیابی در مراحل منبع و داده‌های ورودی انجام شود.

۵.۷ دسترسی پذیری و تفسیر پذیری

دسترسی‌پذیری هم به سهولت فهم و تشخیص ماهیت داده‌ها توسط کاربر و هم به مناسب بودن شیوه‌ها و قالب‌هایی که داده‌ها را می‌توان در دسترس کاربر قرار داد، اشاره دارد. تفسیرپذیری یا وضوح به در دسترس بودن هرگونه اطلاعات تکمیلی یا فراداده‌های ضروری برای کمک به کاربر در تفسیر و درک برون داده‌های آماری اشاره دارد. این ابعاد را می‌توان از طریق مقایسه‌ی شیوه‌های انتشار با مجموعه‌ی اولیه‌ی نیازهای کاربران که از طریق مشاوره از قبل تعیین شده است، به‌آسانی ارزیابی کرد. تدوین آمارگیری رضایت‌مندی کاربر توسط NSO و یا تشکیل کارگروه بازبینی کارشناسی پس از تکمیل برنامه‌ی برون داد برای ارزیابی میزان دسترسی‌پذیری و وضوح (و همچنین مرتبط بودن) از دیدگاه کاربران مفید خواهد بود.

۸ بحث و نتیجه‌گیری

فارغ از روش گردآوری داده‌ها، ارزیابی کیفیت برون داد سرشماری همواره یک کار مهم و ضروری بوده است. چندین بعد بین‌المللی پذیرفته شده وجود دارد که می‌تواند برای کمک به NSOها در ارزیابی کیفیت فرایندها و داده‌های سرشماری به کار گرفته شود تا هم تولیدکنندگان و هم کاربران بتوانند کیفیت برون داد آماری را درک کنند. این ابعاد کلیدی عبارت از مرتبط بودن درستی و قابلیت اعتماد، روزآمدی و به موقع بودن، هم‌خوانی و مقایسه پذیری و دسترسی‌پذیری و تفسیرپذیری هستند.

در این مقاله ابزارها و نماگرهای کلیدی ابعاد کیفیت مذکور در چهار مرحله‌ی ارزیابی کیفیت منبع، داده‌های ورودی، فرایند و برون داد در فرایند سرشماری ارائه شدند. در این راستا برای گذار باکیفیت از سرشماری سنتی به سرشماری ثبتي‌مبنا و تضمین برون داد آماری باکیفیت، لازم است NSOها در مراحل اولیه‌ی برنامه‌ریزی، به دنبال شناسایی منابع اداری مرتبط با سرشماری باشند که بر اساس طیف وسیعی از نیازهای مختلف کاربران در نظر گرفته شده‌اند و مشخص شود که نتایج مورد انتظار یا مورد نیاز هر منبع چه خواهد بود تا بتوان مرتبط بودن را ارزیابی کرد. همچنین ارتباط بین NSO و متولیان داده‌های اداری باید توسط سازوکارهای استوار ارتباط، تفاهم‌نامه‌های کتبی و درک عالی از نیازهای هر دو طرف پشتیبانی شود و باید مبنای

مزایای اصلی سرشماری ثبتي‌مبنا یا ترکیبی، امکان تولید روزآمدتر و مکررتر برون داده‌ها نسبت به سرشماری سنتی ده سالانه است. با توجه به این موضوع، روزآمدی برآوردهای قابل تولید باید به‌عنوان یک عامل کلیدی کیفیت در نظر گرفته شوند. البته، روزآمدی برون داده‌ها تا حد زیادی با روزآمدی تحویل داده‌های ورودی و زمان صرف شده توسط NSO برای انجام پردازش کامل داده‌ها تعیین می‌شود. اغلب بین روزآمدی و درستی رابطه‌ای وجود دارد و به‌طور کلی هر چه NSO زمان بیشتری را برای اطمینان از درستی داده‌ها در طول پردازش اختصاص دهد، تحویل برون داده‌ها به کاربر بیشتر طول می‌کشد. شاخص‌های کمی را می‌توان برای اندازه‌گیری روزآمدی نتایج نهایی شامل هرگونه تأخیر بین گردآوری داده‌ها، دریافت داده‌ها، اتصال داده‌ها و انتشار آمارها به کار برد. به‌عنوان مثال، روزآمدی کل می‌تواند به‌صورت دوره‌ی زمانی از پایان دوره‌ی مرجع تا دریافت داده‌های اداری، تقسیم بر زمان از پایان دوره‌ی مرجع تا تاریخ انتشار، ضرب در ۱۰۰ درصد محاسبه شود [۱۴]. بعد به موقع بودن را نیز می‌توان به‌سادگی با تعیین تفاوت تاریخ از پیش تعیین شده‌ی تحویل برون داده‌ها و تاریخ واقعی تحویل ارزیابی کرد.

۴.۷ هم‌خوانی و مقایسه پذیری

برآوردهای تولید شده توسط داده‌های اداری باید با مشخصه‌های شناخته شده‌ی جمعیت هم از لحاظ طولی (از طریق مقایسه‌ی یک سرشماری با سرشماری دیگر) و هم از لحاظ جغرافیایی و همچنین با برون داده‌های حاصل از سایر عملیات داده‌های آماری NSO سازگار باشد. علاوه بر این، ارزیابی میزان مقایسه پذیری و سازگاری داده‌های برون داد با استانداردهای بین‌المللی بسیار مهم است. تحلیل‌های جمعیت شناختی یک رویکرد خاص است که می‌تواند برای ارزیابی هم‌خوانی و مقایسه پذیری (و همچنین درستی) برون داده‌های سرشماری به کار رود. شایان ذکر است که هنگام استفاده از تحلیل‌های جمعیت شناختی، برآوردها در دو منبع می‌توانند در گروه‌های سنی-جنسی مختلف یا توزیع‌های دیگر متفاوت باشند. این تفاوت‌ها می‌تواند ناشی از موارد زیر باشد:

- جامعه‌های هدف متفاوت،
- تاریخ‌های مرجع متفاوت یا تغییرات جمعیت (به‌ویژه هنگام مقایسه با داده‌های سرشماری پیشین)،
- تفاوت‌های مفهومی و تغییرات رده‌بندی بین متغیرهای مورد مقایسه در منابع مختلف و
- تفاوت در نمونه‌گیری، روش‌های گردآوری و رویکردهای پردازش

سرشماری استفاده می‌شود. انتقال یکپارچه‌ی منابع داده‌های اداری به فرایند سرشماری باید توسط تحقیقات امکان‌سنجی با منابع کافی انجام شود. توصیه می‌شود که NSOها قبل از سرشماری اصلی، تعدادی آزمایش را اجرا کنند تا از شناسایی هرگونه مشکل پیش‌بینی‌نشده اطمینان حاصل کنند و زمان کافی برای تصحیح یا تعدیل روش‌ها، فرایندها یا نظام‌ها فراهم شود. همچنین باید نتایج هرگونه ارزیابی کیفیت و حصول اطمینان از تمام مراحل تولید سرشماری ذخیره و منتشر شود. این امر به منظور اطمینان از درک کاربران از نقاط قوت و محدودیت داده‌ها حائز اهمیت است و می‌تواند در تعیین این‌که آیا تعادل مناسب در ابعاد کیفیت به منظور اطلاع‌رسانی بهبود سرشماری‌های آینده به دست آمده است یا خیر، کمک‌کننده باشد. بازبینی کارشناسی با همکاری متولیان داده‌ها و کارشناسان موضوعی و مقایسه‌ی منابع در طول زمان برای شناسایی دغدغه‌های مربوط به کیفیت یک منبع یا ثبت مهم است. همچنین، استفاده از آمارگیری‌های مناسب می‌تواند در شناسایی و تعدیل خطاهای پوشانش و اندازه‌گیری حائز اهمیت باشد.

قانونی برای ارائه و استفاده از داده‌ها وجود داشته باشد تا تعامل خوبی بین NSO و متولیان وجود داشته باشد تا درک عمیقی از منبع داده‌ها به دست آید. از آنجایی‌که داده‌های اداری برای نیازهای سرشماری گردآوری نمی‌شوند، درک و ارزیابی تفاوت‌های بین جامعه‌ها، مفاهیم و تعاریف موردنیاز برای NSOها ضروری است. لازمی این کار دسترسی به فراداده‌های شفاف و جامع در مورد منبع اداری است.

چالش‌ها و محدودیت‌هایی در دستیابی و یکپارچه‌سازی منابع اداری در سرشماری وجود دارد که می‌تواند شامل منابع و هزینه‌ها، مخاطره‌های مرتبط با توانایی متولیان داده‌ها برای تحویل به موقع و مطابق با استانداردهای کیفیت موردنظر و همچنین قابل قبول بودن کاربرد منبع از نظر عموم مردم و کاربران داده‌های سرشماری باشد و لازم است NSO نسبت به آن آگاهی یابد.

NSOها، باید با ارائه‌ی استدلالی واضح از مجموعه‌ی مزایا در برابر هرگونه مخاطره و هزینه و به‌ویژه اطمینان‌بخشی در مورد حفظ محرمانگی، در استفاده از داده‌های اداری شفاف باشند و عموم مردم و کاربران داده‌ها باید بدانند که چگونه و چرا از داده‌های اداری در

مراجع

- [۱] فتحی، ا. (۱۳۹۴). چشم‌اندازی به تلفیق داده‌های ثبتی با داده‌های سرشماری عمومی نفوس و مسکن، دو ماهنامه تحلیلی-پژوهشی، ۱۵، ۲۹-۳۴.
- [۲] دفتر جمعیت، نیروی کار و سرشماری. (۱۳۹۸). ضرورت‌ها و مزایای تغییر روش اجرای سرشماری به ثبتی‌مبنا، مرکز آمار ایران.
- [۳] دبیرخانه‌ی ستاد سرشماری ثبتی‌مبنا ۱۴۰۵. (۱۴۰۱). گزارش اقدامات سرشماری عمومی نفوس و مسکن ثبتی‌مبنا ۱۴۰۵، مرکز آمار ایران.
- [۴] عباسی، م. ب؛ و حسینی، س. م. (۱۳۹۸). تأملاتی درباره‌ی اجرای سرشماری ثبتی‌مبنا در ایران. فصلنامه جمعیت، ۲۵ (۱۰۴ و ۱۰۳)، ۱۰-۲۱.
- [۵] میرفلاح نصیری، س. ن؛ کیانی، ک؛ محزون، ع. ا؛ قصوری، ش؛ شیرینی، م؛ شباک، ا؛ پوررضانور، م؛ فتحی، ا؛ خدامرادی، ح؛ عباسی، آ؛ سرخیل، ب؛ و فیاض، س. (۱۳۹۸). ارزیابی کیفیت داده‌های ثبتی یا اداری برای استفاده در سرشماری ثبتی‌مبنا، پژوهش‌کده‌ی آمار، تهران.
- [۶] یکانی، م؛ و زبرجد، س. (۱۳۹۴). فایل آدرس زیربنای سرشماری ثبتی‌مبنا، دو ماهنامه تحلیلی-پژوهشی، ۱۲، ۳۸-۴۲.
- [7] European Statistical System (ESS). (2019). *The Second Revision of the Quality Assurance Framework of the European Statistical System*.
- [8] Eurostat (2020). *European Statistical System Handbook for Quality and Metadata Reports*, <https://ec.europa.eu/Eurostat/documents/3859598/10501168/KS-GQ-19-006-EN-N.pdf>
- [9] Daas, P., Ossen, S., Vis-Visschers, R., and Arends-Tóth, J. (2009). *Checklist for the Quality evaluation of Administrative Data Sources*, Statistics Netherlands, The Hague Heerlen.
- [10] International Organization for Standardization (2015). *Quality management systems — Fundamentals and vocabulary*. ISO 9000:2015(en) <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en>.

- [11] Lothian, J., A. Holmberg and A. Seyb (2019). An evolutionary schema for using “it-is-what-it-is” data in official statistics. *Journal of Official Statistics*, 35, 137-165.
- [12] United Nations Department of Economic and Social Affairs, Statistics Division. (UNSD). (2019). *Handbook on Population and Housing Census Editing*, Revision 2. United Nations. New York
- [13] United Nations Economic Commission for Europe (UNECE). (2018). *Guidelines in the Use of Registers and Administrative Data for Population and Housing Censuses*. United Nations, New York and Geneva.
- [14] United Nations Economic Commission for Europe (UNECE). (2021). *Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses*. United Nations, New York and Geneva.
- [15] United Nations Economic Commission for Europe (UNECE). (2017). *Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources*. Version 2.0, United Nations, New York and Geneva.
- [16] United Nations Statistics Division (UNSD). (2022). *Handbook on Registers-Based Population and Housing Censuses*. <https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3e-Handbook-E.pdf>.

Introduction of tools and key indicators to assess the quality of administrative data for use in the population and housing census

Paria Torabi Kahlan¹, Lida Kalhori Nadrabadi²

Abstract:

The utilization of administrative resources in censuses offers the possibility of reducing costs, enhancing data quality, and expediting information production. Alongside the annual monitoring of indicators for sustainable development goals, these initiatives can significantly contribute to meeting the growing demands of the country's planning and research system. However, numerous challenges exist in this regard, with one of the most crucial being the assessment of the quality of administrative data. Quality assessment stands as the paramount aspect of employing registration and administrative data in censuses, constituting a necessity for register-based population and housing censuses where traditional quality assessment criteria prove inadequate. Despite the advantages of utilizing register and administrative data in censuses, numerous critical quality risks must be addressed and assessed before their incorporation. This paper reviews the methods through which national statistics offices can evaluate the quality of data obtained from registers, with the objective of producing high-quality statistical outputs. Consequently, the tools and key indicators utilized to quantify the quality assessment at each of the four stages — source, input data, process, and output — in the census process are introduced.

Keywords: Traditional Census, registers-based census, Assessing the quality of administrative data.

¹Statistical Research and Training Center, Tehran, Iran. torabiparya@yahoo.com

²Statistical Research and Training Center, Tehran, Iran. kalhori@srct.ac.ir.