

ارزیابی کیفیت پرتال‌های داده باز در سطح ملی

مرکز توانمندسازی حاکمیت و جامعه جهاد دانشگاهی
آذر ۱۳۹۹

۱۳۷



بنام خداوند جان و آفریننده



ارزیابی کیفیت پرتال‌های داده باز در سطح ملی

نویسنده: رناتا ماچووا و مارتین لنیکا

مترجم: سیاوش امجدی

مجموعه داده‌باز و شفافیت - ۱

مرکز توانمندسازی حاکمیت و جامعه جهاد دانشگاهی

آذر ۱۳۹۹

فهرست مطالب

۲	درآمد مرکز توانمندسازی حاکمیت و جامعه.....
۵	چکیده.....
۵	۱. مقدمه.....
۸	۲. متدولوژی و روش تحقیق.....
۹	۳. مرور و پیشینه ادبیات.....
۹	۳- ۱. داده‌های باز و برنامه‌های مرتبط.....
۱۱	۳- ۲. مزایا، خطرات و تأثیرات بازکردن داده‌های دولتی.....
۱۳	۳- ۳. پرتال داده‌های باز، شرح و طبقه‌بندی.....
۱۶	۳- ۴. الزامات ارزیابی و کیفیت پرتال‌های داده باز.....
۱۸	۳- ۵. محک‌زن‌ها و مدل‌ها برای ارزیابی پیشرفت پرتال‌های داده باز.....
۲۰	۴. نتایج ارزیابی کیفیت پرتال‌های داده باز.....
۲۱	۴- ۱. ارزیابی کیفیت پرتال‌های داده باز از طریق رابط برنامه‌نویسی کاربردی (API).....
۲۶	۴- ۲. ارزیابی کیفیت پرتال‌های داده باز با استفاده از چارچوب محک‌زنی.....
۳۲	۵. یافته‌ها و بحث.....
۳۴	۶. نتیجه‌گیری.....
۳۶	منابع.....

درآمد مرکز توانمندسازی حاکمیت و جامعه

داده ماده خامی است که از آن می‌توان اطلاعات و دانش بدست آورد، بدون داده نمی‌توانیم اطلاعات بسازیم و بدون اطلاعات دانش جدیدی نیز وجود ندارد. یکی از انواع داده‌ها، داده باز است. داده باز داده‌هایی است که هرکس می‌تواند به آن‌ها دسترسی داشته باشد، از آن‌ها استفاده کرده و آن‌ها را به اشتراک بگذارد. فلسفه ایجاد داده‌های باز مدت‌هاست که تثبیت شده اما اصطلاح "داده‌های باز" اخیراً با ظهور اینترنت و شبکه جهانی وب و بخصوص با ابتکار داده‌های دولتی باز محبوبیت بیشتری یافت.

یکی از مهمترین انواع داده‌های باز، داده‌های باز دولتی (OGD) است که توسط نهادهای حاکمیتی و دولتی ایجاد می‌شود. این داده‌ها می‌تواند شامل اسناد، گزارش‌ها، پایگاه‌های اطلاعاتی، نقشه‌ها یا اطلاعات مکانی و غیره باشد. اهمیت داده‌های باز دولتی از آنجایی سرچشمه می‌گیرد که این داده‌ها بخشی از زندگی روزمره شهروندان است. داده‌های باز دولتی، می‌توانند به شفاف‌سازی اقدامات دولت‌ها کمک کنند، شواهدی در اختیار عموم مردم قرار دهند که نشان دهد منابع عمومی به درستی هزینه شده و سیاست‌ها در حال اجراست. گروه‌ها و سازمان‌های مختلفی از جمله خود دولت می‌توانند از در دسترس بودن داده‌های باز بهره‌مند شوند.

در سال ۲۰۰۹ و با گسترش تحولات تکنولوژیکی، اجتماعی جنبش جدیدی در حاکمیت با هدف بهبود رابطه مردم و دولت، افزایش اعتماد شهروندان به دولت و نیز پاسخگو کردن آن تحت عنوان "دولت باز" شکل گرفت که مبتنی بر مفهوم "داده باز" بود. پس از آن در سطح جهانی، داده‌های دولتی باز یک بخش مهم از «مشارکت دولتی باز» شد و منجر به پایه‌گذاری یک طرح بین‌المللی در سپتامبر ۲۰۱۱ توسط هشت دولت شد که در حال حاضر شامل ۵۵ کشور است. دولت‌های شرکت‌کننده در این طرح متعهد شده‌اند که دسترسی به اطلاعات درباره فعالیت‌های دولتی، حمایت از مشارکت مدنی و اجرای بالاترین استانداردهای یکپارچگی حرفه‌ای در سطح اداراتشان را افزایش دهند. داده‌های باز دولتی مزایای زیادی دارند که از جمله آن‌ها می‌توان به موارد زیر اشاره کرد:

- شفافیت و کنترل دموکراتیک
- توانمندسازی حاکمیت
- بهبود کارایی خدمات دولتی
- بهبود اثربخشی خدمات دولتی
- اندازه‌گیری تأثیر سیاست‌ها

به عنوان مثالی از شفافیت می‌توان به پروژه‌هایی مانند "درخت مالیات"^۱ در فنلاند و "پول من کجا می‌رود"^۲ در انگلستان اشاره کرد که نشان می‌دهد چگونه پول مالیات توسط دولت هزینه می‌شود. یا در مثال دیگر داده‌های باز از کلاهبرداری ۳/۲ میلیارد دلاری خیریه‌ای در کانادا جلوگیری کرد. در دانمارک وب سایت folketsting.dk فعالیت نمایندگان پارلمان در روند تصویب یک قانون را ردیابی کرده بنابراین شهروندان دانمارک می‌توانند رصد کنند دقیقاً چه اتفاقی در پارلمان رخ می‌دهد و هر نماینده پارلمان درگیر کدام قانون است.

داده‌های باز نظارت بر فعالیت‌های دولت از جمله ردیابی هزینه‌های بودجه عمومی و تأثیرات آن‌را آسان‌تر می‌کند. فرایندی که به افزایش نظارت بر عملکرد دستگاه‌های دولتی بنجامد مطمئناً سبب بهبود عملکرد سازمان‌های دولتی در ارائه خدمات به شهروندان نیز خواهد شد. از سوی دیگر اطلاعات به شهروندان آگاهی می‌دهد و یک شهروند آگاه بهتر می‌تواند در فرآیندهای دولتی مشارکت داشته باشد و در شکل‌گیری سیاست‌های دولتی مشارکت کند.

استدلال طرفداران سرسخت داده‌های دولتی باز آن است که این داده‌ها یک عامل کلیدی برای افزایش شفافیت و پاسخگویی عمومی است. در واقع این حق مردم است که بدانند دولت‌هایشان چگونه عمل می‌کنند و دولت‌ها در برابر اقداماتشان مسئول باشند. در واقع شفافیت واقعی یعنی نه تنها مردم بتوانند به اطلاعات به سادگی دست یابند بلکه امکان استفاده بیشتر، ترکیب و انتشار آن‌ها را بدون هیچ محدودیتی داشته باشند.

یک تصور عمومی آن است که عدم وجود این داده‌ها در حوزه عمومی، این اجازه را به کارمندان دولتی می‌دهد تا با مصونیت از مجازات رفتارشان، فاسدانه عمل کنند. داده‌های دولتی باز تجزیه و تحلیل، پردازش و ترکیب اطلاعات موجود را ساده‌تر می‌کند ضمن آن که سطح پاسخگویی دولت عمومی را افزایش داده تا سطح جدیدی از انتظارات عمومی و واکنش‌پذیری اقدامات دولت را بسازند. البته باید به این نکته توجه کرد که شفافیت از اجزای حیاتی دولت باز است که به دو شکل شفافیت در عملکرد و شفافیت دولت تقسیم می‌کنند. منظور از شفافیت دولت، شفافیت در عملیات، رویه‌ها و وظایف دولت است که ابزاری برای پاسخگویی است اما منظور از شفافیت داده، شفافیت داده‌هایی است که دولت ارائه می‌کند. بنابراین دولت باز و داده‌های دولتی باز هر یک می‌توانند بدون دیگری وجود داشته باشند و افزایش شفافیت لزوماً عاملی برای پاسخگویی بیشتر نیست. یک دولت به معنای واقعی می‌تواند شفاف باشد حتی اگر از فن‌آوری‌های جدید استقبال نکند و یک دولت می‌تواند داده‌های باز ارائه دهد اما همچنان غیرپاسخگو باشد. بنابراین اینگونه می‌توان نتیجه گرفت که داده‌های دولتی باز هر چند شفافیت ایجاد می‌کند اما لزوماً دولت‌ها را در برابر نتیجه سیاست‌هایشان پاسخگو نمی‌کند. آنچه در این بین مهم است آن است که دولتی که اطلاعات را به اشتراک می‌گذارد خود به این باور رسیده است که باید شهروندان در خط مشی عمومی مشارکت داشته و به مردم قدرت لازم را برای پاسخگو کردن دولت را داد.

قانون انتشار و دسترسی آزاد به اطلاعات مصوب ۱۳۸۸ تمهیدی در راستای تقویت داده‌باز در نظام حکمرانی ایران بود. این قانون همچنان تحقق مطلوبی در نظام حکمرانی نیافته است و دسترسی شهروندان به داده‌های دستگاه‌های

1 Tax tree

2 Where does my money go?

حاکمیتی با موانع جدی روبرو است. اما در سال‌های اخیر، در نتیجه‌ی تسری گفتمان شفافیت و پیگیری‌های جامعه‌ی مدنی و اراده‌ی سیاسی برخی مسئولان، شماری از دستگاه‌های دولتی اقدام به اشتراک‌گذاری داده‌های خود در فضای عمومی کرده‌اند. طراحی و تدوین سامانه‌های قراردادها، حقوق و دستمزد و پرتال‌های شفافیت از سوی برخی دستگاه‌های دولتی نمودی از این روند است.

مقاله‌ی پیش رو که توسط مرکز توانمندسازی حاکمیت و جامعه جهاد دانشگاهی ترجمه و منتشر شده است چارچوبی برای سنجش و ارزیابی کیفیت پرتال‌های داده باز در سطح ملی ارائه داده و سپس بر اساس این چارچوب ۶۷ کشور به لحاظ راه‌اندازی پورتال‌های داده باز ارزیابی و رتبه‌بندی شده‌اند. یافته‌های این پژوهش نشان می‌دهد که انگلستان، هند و ایالات متحده، بسیاری از مجموعه داده‌ها را منتشر کرده و پرتال‌های پیشرفته‌ای را راه‌اندازی کرده‌اند.

امید است که انتشار این پژوهش گام کوچکی در راستای ارزیابی سامانه‌های داده‌باز در ایران بوده و به تقویت این گفتمان و عملکرد در نظام حکمرانی یاری رساند.

مرکز توانمندسازی حاکمیت و جامعه جهاد دانشگاهی - آذر ۱۳۹۹

ارزیابی کیفیت پرتال‌های داده‌باز در سطح ملی^۱

چکیده

طی چندسال گذشته، دولت‌های سراسر جهان شروع به توسعه و اجرای ابتکار عمل‌های مرتبط با داده‌های باز کرده‌اند، تا بدین واسطه بتوانند امکان انتشار داده‌های دولتی را در قالب‌های باز و مجدداً قابل استفاده فراهم آورند، بی‌آن‌که استفاده از آن‌ها برای جامعه هزینه یا محدودیتی داشته باشد. در نتیجه، تعداد زیادی مخازن، کاتالوگ‌ها و پرتال‌های داده‌باز در جهان، در حال ظهور هستند. توسعه کارآمد پرتال‌های داده‌باز، ارزیابی سیستماتیک کیفیت آن‌ها را به منظور درک بهتر و تشخیص انواع مختلفی از ارزش‌هایی که تولید می‌کنند، ضروری می‌سازد.

هم‌چنین، شهروندان انتظار دارند که داده‌های منتشر شده به واسطه مقامات رسمی کیفیت داده‌های رسمی را داشته باشند و بنابراین بایستی دقیق و معتبر باشند. هدف این مقاله بررسی و مقایسه کیفیت این پرتال‌ها است. برای این منظور، چارچوبی برای سنجش و ارزیابی کیفیت پرتال‌های داده‌باز در سطح ملی مطرح و اعتبارسنجی می‌شود. نتایج به دست آمده نشان می‌دهد که تعداد مجموعه داده‌های آنلاین و پیچیدگی پرتال‌های داده‌باز و کارکرد آن‌ها متفاوت است، این امر نشان‌دهنده فقدان هماهنگی و ضرورت وجود استانداردهای کیفیت است. به‌طور خاص، انگلستان، هند و ایالات متحده، بسیاری از مجموعه داده‌ها را منتشر کرده و پرتال‌های پیشرفته‌ای را راه‌اندازی کرده‌اند.

کلیدواژه‌ها: داده‌باز، دولت‌باز، پرتال‌های داده‌های باز ملی، چارچوب سنجش، تحلیل محتوا، ارزیابی، کیفیت

۱. مقدمه

اقتصاد جهانی به یک اقتصاد داده‌محور تبدیل شده است و در نتیجه، کسانی که توانایی استخراج حداکثر سود از داده‌های خود را دارند از قدرت سیاسی، اجتماعی، فرهنگی و به‌ویژه اقتصادی برخوردار خواهند بود [۵]. بنابراین، طی چندسال گذشته، تعداد دولت‌هایی که اقدام به ارائه داده‌های باز می‌کنند، رو به افزایش است. این جنبش به اصطلاح دولت‌باز منجر به راه‌اندازی پرتال‌ها و زیرساخت‌های متعددی برای داده‌های باز شده است، که هدف آن‌ها فراهم کردن دسترسی به داده‌های دولت و بررسی تأثیرات آن‌ها است [۲۴]، [۳۰]، [۵۳]، [۵۸]. با انتشار این داده‌ها در پرتال‌های داده‌باز، دولت‌ها داده‌هایی را که هزینه ایجاد آن‌ها در وهله نخست به‌طور غیرمستقیم از طریق مالیات شهروندان پرداخت شده است، به شهروندان ارائه می‌دهد [۳۲]. ظهور این پرتال‌ها نحوه جست‌وجوی شهروندان و محققان را برای دستیابی به داده‌ها تغییر می‌دهد. نکته اصلی ایجاد تغییری سریع در جهت به اشتراک‌گذاری داده‌ها و شفافیت

۱. این مقاله ترجمه‌ای است از:

است. این تغییر می‌تواند برای دستیابی به مزایای حاصل از ایجاد مخازن بیش‌تری از داده‌های باز در بخش دولتی مفید باشد [۵۵].

اگرچه منابع مختلفی از داده‌ها وجود دارند، داده‌های دولتی به‌ویژه به‌دلیل مقیاس، گستردگی و جایگاه آن‌ها به‌عنوان منبع استاندارد اطلاعات طیف گسترده‌ای از موضوعات، از اهمیت ویژه‌ای برخوردار هستند [۵۰]. این داده‌ها، داده‌های باز دولتی (OGD) نامیده می‌شوند. یادآوری صریح این نکته مهم است که داده‌های باز دولتی (OGD) نه معادل داده باز، بلکه زیرشاخه یا زیرمجموعه‌ای از آن هستند، که ممکن است به‌همان‌اندازه در بخش‌های تجاری، دانشگاهی یا بخش‌های غیردولتی ریشه داشته باشند [۱۸]. به‌دلیل تعداد زیاد داده‌های تولیدشده توسط بخش دولتی، مدل داده‌های باز به مدل داده‌های بزرگ باز تبدیل شده است [۱۱]، [۳۱]، [۳۸]. تقاطع داده‌های باز و بزرگ بیش‌تر مربوط به ادغام چندین منبع است؛ برای مثال، منابعی در سطوح بین‌المللی، ملی، منطقه‌ای و محلی بخش دولتی و سازمان‌های بین‌المللی [۳۱]. مارتون و همکاران [۳۸] اظهار می‌دارند که مفاهیم اساسی داده‌های باز و بزرگ ماهیتی فنی دارند، زیرا در زمینه‌های علوم کامپیوتر و مهندسی توسعه یافته‌اند. در هر دو مورد داده‌ها برای یک هدف جمع‌آوری می‌شوند و سپس به‌طور معمول برای هدفی متفاوت مورد استفاده قرار می‌گیرند [۱۱]. درحالی‌که داده‌های بزرگ اغلب با حجم زیادشان مشخص می‌شوند، با سرعت زیادی گردآوری می‌شوند و ممکن است بدون ساختار بوده و از منابع بسیاری به‌دست آیند، [منتهی]، داده‌های باز در خصوص چگونگی ساخت داده‌های استاندارد است که برای کامپیوتر خوانا بوده و بنابراین پیوندپذیر هستند [۳۸].

دولت‌ها و سایر مقامات بخش دولتی از خلال فعالیت‌های روزمره خود، مانند مدیریت حقوق بازنشستگی و پرداخت مستمری، جمع‌آوری مالیات، ثبت داده‌های ترافیک و صدور اسناد رسمی، مقادیر زیادی داده را تولید و جمع‌آوری می‌کنند [۵]، [۵۵]. برای مثال، این داده‌ها بزرگ‌ترین منبع اطلاعاتی در اروپا هستند که ارزش بازاری آن‌ها ۳۲ میلیاردیورو تخمین زده می‌شود. براساس نظر، بوخولتز و همکاران [۵] مجموع تأثیرات اقتصادی مستقیم و غیرمستقیم استفاده از داده‌های باز بزرگ در کل اقتصاد اتحادیه اروپا (EU) سالانه میلیاردها یورو می‌شود. از این‌رو، این داده‌ها، در ترکیب خلاقانه با سایر منابع داده‌ای، پتانسیل قابل‌توجهی در راستای استفاده مجدد برای تولید محصولات و خدمات جدید دارند [۸]. زیلنر و همکاران [۵۵] تخمین زده‌اند که روی‌هم‌رفته سود اقتصادی سالانه این داده‌ها در اتحادیه اروپا ۴۰ میلیارد یورو است؛ اتحادیه اروپا، علاوه بر مزایای اقتصادی فی‌نفسه، سایر ارزش‌های این داده‌ها را به‌رسمیت می‌شناسد؛ این ارزش‌ها شامل این موارد می‌شوند: کمک به فرآیند رسیدگی به چالش‌های اجتماعی و صنعتی، دستیابی به کارایی بیش‌تر از خلال به‌اشتراک گذاشتن داده‌ها در داخل و بین ادارات امور دولتی، و تقویت مشارکت شهروندان در زندگی سیاسی و اجتماعی و افزایش شفافیت دولت. با استفاده از تسهیلات فنی و پیشرفت در ترکیب و تحلیل مجموعه داده‌های مختلف، امکان استفاده بهتر از این داده‌های موجود در حال افزایش است [۲۱]. زمینه‌های کاربرد این فناوری‌ها، تحلیل داده‌ها برای تحلیل پیشرفته مجموعه داده‌های بزرگ (مزایایی برای کشف کلابرداری و امنیت سایبری)، افزایش کارایی برای تقویت شفافیت داخلی و بهبود کارایی برای ارائه خدمات بهتر به شهروندان را شامل می‌شود [۳۰]، [۵۴]، [۵۵]. طی سال‌های گذشته، تعداد زیادی اجتماعات داده باز که ایده‌ها و

برنامه‌های جدیدی را توسعه می‌دهند در پیرامون پرتال‌های داده‌باز دولت (OGD) ظهور یافته‌اند. بنابراین، هر جا که ممکن است، هدف باید ایجاد داده‌های باز بزرگ باشند [۴۲]. تنها درخواست حداقلی و غیررادیکیال این است: گشودگی داده‌ها به صورت پیش فرض. دولت‌ها هرچه زودتر داده‌های خود را افشا کنند، بازده بیش‌تری خواهند داشت [۵۵].

با این حال، افشای این حجم عظیم از داده‌ها لزوماً معادل شفافیت بیش‌تر نیست و ضرورتاً پاسخ‌گویی را تسهیل نمی‌کند [۳۲]. به جز اهمیت اقتصادی، موارد دیگری وجود دارند که به تنظیم داده‌های دولت مربوط‌اند. از جمله قابلیت یافتن داده، برداشت داده (داده‌کاوی)، مشارکت اجتماعی و هم‌کنش‌پذیری یا تعامل‌پذیری [۲۰]، [۵۰]. هم‌چنین، کیفیت این داده‌ها و پرتال‌های مربوطه ممکن است از کشوری به کشوری دیگر متفاوت باشد، و این می‌تواند ارزش آن‌ها را برای کاربران تحت‌تأثیر قرار دهد.

قابلیت یافتن داده‌های مرتبط، پیش‌شرط استفاده از پتانسیل داده‌های باز است. ایجاد پرتال از مجموعه داده‌های موجود یکی از راه‌هایی است که به ما می‌گوید که چگونه این مجموعه داده‌ها را دسترس‌پذیرتر کرده و یافتن آن‌ها را آسان‌تر سازیم [۲۷]. از سوی دیگر، علاوه بر وجود سیستم‌های مدیریتی متفاوت برای پرتال‌های داده باز و برنامه‌های مربوط به داده‌های باز، این پرتال‌ها و برنامه‌ها از حیث محتوا، عملکرد و استانداردهای فناوری تنوع زیادی دارند [۴]، [۳۰]. اما مهم‌تر از همه، آن‌ها از نظر سودمندی و متناسب‌بودن با وظیفه‌شان، از یکدیگر تمایز دارند [۴]. استخراج اطلاعات ارزشمندی که از این منابع مختلف داده به دست می‌آید، مستلزم ارزیابی کیفی است [۶]. کارلو و همکاران [۷] کیفیت یک پرتال وب را به واسطه تعریف یک مدل کیفیت داده ارزیابی کرده‌اند که شامل ۳۳ مشخصه کیفیت داده است. آن‌ها این مشخصات را در چهار دسته کیفیت داده گروه‌بندی کرده‌اند. در حالی که برخی از جنبه‌های کیفیت داده‌های باز با جنبه‌های پرتال وب هم‌راستا هستند، اما در مورد داده‌های باز، چشم‌اندازهای کیفیت پرتال‌های داده‌های باز در یک حوزه خاص (برای مثال، سیستم مدیریت داده، باز بودن داده‌های ارائه شده بر مبنای پروانه یا فرمت، فراداده) نیازمند شناسایی و ارزیابی است. در زمینه استفاده از پرتال‌های داده‌های باز، کیفیت داده‌ها نقشی اساسی بازی می‌کنند و سطح مشخصی از کیفیت داده‌ها برای استفاده از داده‌های باز دولتی (OGD) بسیار مهم است [۵۷].

با این همه، آمبریج و همکاران [۴۸] استدلال می‌کنند که علی‌رغم اشتیاق موجود نسبت به در دسترس قرار دادن مقادیر زیادی از داده‌های باز دسترس‌پذیر، نخستین انتقادات از کیفیت پایین فراداده‌ها و منبع داده پرتال‌های داده باز حکایت می‌کند و این یک خطر جدی است که می‌تواند پروژه داده‌های باز را مختل کند. کوچرا و همکاران [۲۷]، آمبریج و همکاران [۴۸]، و هم‌چنین زویدروایک و جانسن [۵۷] ادعا کرده‌اند که برای درک بهتر مسئله کیفیت پرتال‌های داده باز و بررسی تأثیر روش‌های بهسازی در طول زمان، نیاز به چارچوبی برای محک‌زدن و ارزیابی کیفیت وجود دارد.

بنابراین، این مقاله تلاش می‌کند با مرور ادبیات، و با تمرکز بر ارزیابی کیفیت ابعاد مختلف این پرتال‌ها و ارائه یک تحلیل در سطح ملی از پرتال‌هایی که در حال حاضر در دسترس هستند، چارچوب سنجش جدیدی را مطرح کند و بدین واسطه شکاف‌های تحقیق را پر سازد؛ این چارچوب جدید سنجش برای درک این که آیا این پرتال‌ها داده‌هایی را

عرضه می‌دارند که در واقع استفاده، استفاده مجدد و بازتوزیع آن‌ها پاسخ‌گویی عمومی را تسهیل می‌نماید یا نه، بکار می‌آید.

شاکله مقاله به شرح زیر است. در بخش دوم، متدولوژی تدوین شده و روش تحقیق انتخاب و شرح داده شده است. در بخش سوم، مروری بر ادبیات و سوابق داده‌های باز و برنامه‌های مربوطه، مزایا، خطرات و تأثیرات ارائه داده‌های باز دولتی، پرتال‌های داده باز، کیفیت و الزامات ارزیابی آن‌ها و غیره عرضه می‌شود. بخش چهارم، نتایج ارزیابی کیفیت پرتال داده باز را ارائه می‌دهد. در بخش پنجم، این یافته‌ها مورد بحث قرار گرفته است. نهایتاً ملاحظات مربوط به نتیجه‌گیری در بخش ششم ارائه شده است.

۲. متدولوژی و روش تحقیق

در این پژوهش روشی انتخاب شده که بتوان براساس آن یک تحلیل چند بعدی را اجرا کرد [۳۹]. که به‌وضوح بین بعدهای مختلف و امتیاز آن‌ها و ارزیابی کیفیت پرتال‌های داده‌های باز ملی در سطح جهان، تمایز می‌گذارد. این روش تحقیق شامل چهار مرحله متمایز است:

۱. شناسایی منابع داده‌های باز دولتی (OGD) به‌عنوان پرتال‌های داده باز در سطح ملی، با توجه به تعریف [۳۰] و رده‌بندی ارائه شده توسط شاخص جهانی داده‌های باز (GODI) [۴۰] و هم‌چنین شاخص و گزارش جهانی بارومتر داده‌های باز (ODB) [۵۲]؛

۲. پیشنهاد یک چارچوب سنجش (تعریف ابعاد کیفیت، و معیارهای خاص برای هر یک از آن‌ها) مبتنی بر مرور نظام‌مند ادبیات، همان‌طور که توسط پتیکروو و رابرتز [۴۱] تعریف شده است، یعنی: مروری که با هدف شناسایی جامع همه مطالعات مربوطه برای پاسخ‌گویی به یک پرسش خاص و ارزیابی اعتبار (یا درستی) هر مطالعه صورت می‌گیرد، و هنگام نتیجه‌گیری این مرور مد نظر قرار می‌گیرد. بنابراین، مراحل زیر تعریف شده‌اند:

- تعریف اصطلاحات، کلیدواژه‌ها و استراتژی‌های تحقیق برحسب پرسش تحقیق؛
- انتخاب منابع (کتاب‌خانه‌های دیجیتالی) جست‌وجو؛
- کاربرد اصطلاحات و کلیدواژه‌های جست‌وجو در منابع؛
- ارزیابی اعتبار مطالعات شناسایی شده در تحقیق؛ و
- انتخاب مطالعات نخستین به‌واسطه بکارگیری معیارهای شمول و حذف بر روی نتایج جست‌وجو.

۳. جمع‌آوری داده‌ها با استفاده از روش پرسشنامه‌ای، به‌عنوان روش نمونه‌گیری داده‌های تحقیق؛ و

۴. پردازش داده‌ها و ارائه نتایج، که مواردی چون محاسبه آمارهای توصیفی مختلف، مانند فراوانی و فراوانی نسبی تمام مقادیر برای هر یک از معیارها، و ساخت نمودارهای مختلف، با استفاده از نرم‌افزار اکسل را شامل می‌شود.

این مطالعه براساس تکنیک‌های کمی و با عطف به ادبیات شکل گرفته توسط مایلور و بلکمن انجام شده است [۳۹]. به عبارت دقیق‌تر، این پژوهش بر پایه رویکرد تحلیلی توصیه شده توسط آن‌هاست. این رویکرد مبتنی بر روش

جمع‌آوری داده‌ها و ابعاد مورد استفاده است، پیروی می‌شود. این پژوهش با پرسش‌هایی آغاز می‌شود که با استفاده از داده‌های جمع‌آوری شده می‌توان به این پرسش‌ها پاسخ داد. مرحله بعدی، تعریف فرآیند جمع‌آوری و تحلیل داده‌ها است. مرحله آخر هم، به جست‌وجوی روندها و الگوها، یعنی تفسیر داده‌ها مربوط است.

۳. مرور و پیشینه ادبیات

در پرداختن به مرور و پیشینه ادبیات، جریان‌های تحقیقاتی مختلفی که با این مطالعه مرتبط هستند، مورد بررسی قرار می‌گیرند. بخش نخست، داده‌های باز و برنامه‌های مربوطه را تعریف می‌کند. مزایا، خطرات و تأثیرات ارائه داده‌های باز دولتی به دنبال آن خواهد آمد. پس از آن، موضوع پرتال‌های داده باز، شرح و طبقه‌بندی آن‌ها ارائه و مورد بحث قرار خواهد گرفت. در ادامه، کیفیت و الزامات ارزیابی پرتال‌های داده باز دنبال می‌شود. سرانجام، ادبیات مربوط به محک‌زن‌ها و مدل‌های ارزیابی پیشرفت پرتال‌های داده باز مورد بررسی قرار می‌گیرد.

۳ - ۱. داده‌های باز و برنامه‌های مرتبط

متخصصان در بخش دولتی و خصوصی [۲۰] نسبت به موضوع داده‌های باز علاقه‌مند شده‌اند و این علاقه بخشی از نیروی محرکه خود را از فشارهای موجود برای افزایش شفافیت بخش دولتی و بخشی دیگر را نیز به دلیل تمایلی می‌گیرد که در حال حاضر نسبت به ایجاد داده‌های بزرگ و تحلیل داده‌ها وجود دارد [۹]. دولت باز به‌عنوان چتری برای بسیاری از ایده‌ها و مفاهیم عمل می‌کند. اکوسیستم داده‌های باز اغلب اکوسیستم‌های دولتی هستند، زیرا داده‌های باز توسط دولت‌ها منتشر می‌شوند، اگرچه عناصر این اکوسیستم‌ها را بخش خصوصی نیز می‌تواند تهیه کند. اکوسیستم‌های دولت باز بر روابط متقابل متعدد و متنوع میان داده‌ها، ارائه‌دهندگان داده‌های باز، کاربران داده‌های باز، زیرساخت‌های مادی و مؤسسات تأکید می‌کنند [۵۸]. تمامی این‌ها عمدتاً، معادل با برخورداری از یک دولت الکترونیک و استفاده از فناوری‌های ارتباطی و اطلاعاتی مدرن (ICT) است [۵]، [۱۵]، [۴۳]، [۴۷].

با این همه، موضوع دولت باز، موضوع کاملاً جدیدی نیست، زیرا استفاده مجدد از اطلاعات بخش دولتی (PSI) موضوع دستورالعمل‌های اتحادیه اروپا بوده است [۱۶]. اولین دستورالعمل اطلاعات بخش دولتی (PSI) در اواخر سال ۲۰۰۳ اتخاذ شد. این دستورالعمل یک چارچوب قانون‌گذاری مشترک را معرفی می‌کند که نشان می‌دهد نهادهای بخش دولتی چگونه باید اطلاعات خود را برای استفاده مجدد در دسترس قرار دهند و با هم‌آهنگ‌سازی نظام استفاده مجدد از اطلاعات بخش دولتی (PSI)، موانعی مانند اقدامات تبعیض‌آمیز و بازارهای انحصاری را از بین ببرند [۱۶]، [۴۳]. سپس، این دستورالعمل در سال ۲۰۰۹ و مجدداً در سال ۲۰۱۳ مورد بازنگری قرار گرفت، و بدین‌سان نهادهای دولتی بیش‌تری را در قلمرو خود قرار داد و آن‌ها را تشویق کرد داده‌های خود را با هزینه کم‌تر یا به‌صورت رایگان ارائه دهند: بازتابی از آن‌چه که تا آن زمان در بسیاری از کشورهای اتحادیه اروپا انجام می‌شد. گشودگی داده‌ها به‌طور پیش‌فرض در حال تبدیل شدن به یکی از اصول اساسی مقررات داده‌های باز در دستورالعمل اتحادیه اروپا درباره اطلاعات بخش دولتی (PSI) است [۵۰]. برنامه‌های داده باز یک نقطه‌ی شروع برای تقویت بازار داده است که می‌تواند از اطلاعات باز (محتوا) و فناوری‌های داده‌های بزرگ استفاده کند. بنابراین، سیاست‌های فعال در زمینه داده‌های باز می‌تواند

به سود بخش خصوصی باشد و در عوض رشد این صنعت را تسهیل کند. در پایان، این امر با افزایش درآمدهای مالیاتی از صنعت رشد یابنده داده، به بودجه دولتی سود می‌رساند [۵۵]. تعداد برنامه‌های داده‌های باز در دوره ۲۰۰۹ تا ۲۰۱۴ [۲۱]، [۴۳] از دو درصد به بیش از سیصد درصد افزایش یافته است و عضویت در برنامه شراکت در دولت باز (OGP) از هشت کشور در سال ۲۰۱۱ به شصت و نه کشور شرکت‌کننده در سال ۲۰۱۶ رسیده است.

در حالی که، اجرای برنامه‌های داده باز در بین دولت‌ها مشابه به نظر می‌رسد (برای مثال، بیش‌تر از طریق یک پرتال وب متمرکز، که افراد بتوانند در آن مجموعه داده‌ها را بارگیری کنند)، دولت‌ها ممکن است انگیزه‌های مختلفی را برای شروع برنامه‌های داده باز داشته باشند [۲۰]، [۲۴]. همچنین مطالعات اخیر نشان داده است که برنامه‌های فعلی رویکردهای مختلفی را برای فراهم‌سازی داده‌ها به کار می‌گیرند و محدودیت‌های مهمی را مانند تکثیر داده‌ها به نمایش می‌گذارند [۲۳]، [۲۴]. بنابراین، سایاگو و همکاران [۴۴] برای توجیه به کار بستن چارچوب پیشنهادی و برای درک وضعیت برنامه‌های داده‌های باز دولتی (OGD)، ارزیابی جامع و کاملی از موارد منتخب انجام دادند. رگی و ریگی [۴۳]، درباره استراتژی‌های مبتنی بر اطلاعات که مناطق اتحادیه اروپا و کشورهای عضو هنگام انتشار داده‌های دولت در شبکه به اجرا درمی‌آورند، کاوش کردند. سیاست انسجام و بودجه‌های ساختاری آن، که شامل تمام مناطق اتحادیه اروپا و کشورهای عضو است، بستری ایده‌آل برای محقق کردن حضور رویکردهای مختلف برای انتشار داده‌های دولت است. آن‌ها سه رویکرد را شناسایی کردند: رویکرد کاربرمحور، که نشانگر تلاش‌هایی است در جهت قابل فهم کردن داده‌ها برای شهروندان غیرمتخصص، ارائه شفاف داده‌ها و دسترس‌پذیری داده‌ها برای کاربران؛ رویکرد نظارتی، که با مشخصه‌هایی تعریف می‌شود که هدف آن‌ها اطمینان از صحت، اعتبار، امنیت، مدیریت و محافظت از ذخایر اطلاعاتی است؛ و رویکردی با محوریت کاربری مجدد، که بر اهمیت بارگیری داده‌ها در قالب قابل‌خوانش توسط رایانه و سایر ویژگی‌های مربوط به کیفیت داده‌ها تأکید می‌کند. لی و همکاران [۲۹]، سپس برنامه‌های داده باز را در برخی از خلاق‌ترین کشورهای جهان بررسی می‌کنند و خاطر نشان می‌سازند که این برنامه‌ها می‌توانند توسط دولت یا توسط جامعه اداره شوند.

داده‌های باز، محتواها یا داده‌هایی هستند که استفاده، استفاده مجدد و همچنین بازتوزیع آن‌ها برای همگان آزاد است. بیشتر داده‌های باز در واقع داده‌های خام هستند. با این حال، بازنشر آن‌ها به معنای استناد به منبع اصلی است، نه تنها برای اعتبار دادن به داده‌ها بلکه برای اطمینان از این که این داده‌ها تغییر نیافته‌اند یا به طرز نادرستی نمایش داده نشده‌اند [۱۹]، [۲۰]، [۲۶]. سپس داده‌های پیوند شده^۱ روشی را برای انتشار داده‌های ساخت‌یافته توصیف می‌کنند تا بتوانند از طریق فناوری‌های وب معنایی، مانند شناسه‌های منابع یکسان (URI)، چارچوب توصیف منابع (RDF)، واژگان و هستی‌شناسی‌ها، به یکدیگر پیوند خورده و کاربری بیش‌تری داشته باشند. داده‌های پیوند شده

۱ Linked Data: داده پیوندشده، در علم رایانش، نوعی داده ساخت‌یافته است که به شیوه‌ای پیوند درونی با داده‌های دیگر دارد که برای پرسمان‌های معنایی مناسب می‌باشند. داده پیوند شده بر اساس فن‌آوری‌های استاندارد وب مثل HTTP، RDF، و URI ساخته شده است، اما به جای آن که از آن‌ها برای خدمت به صفحات وب و فقط برای انسان‌ها استفاده کند، این داده‌ها را به شیوه‌ای که به صورت خودکار توسط رایانه‌ها قابل خواندن باشند، گسترش می‌دهد. اطلاعات را به اشتراک بگذارد. م.

روشی برای انتشار داده‌ها است، به‌گونه‌ای که می‌تواند تعامل بین منابع مختلف داده یا توسعه خدمات الکترونیک پیشرفته با ارزش افزوده را به‌واسطه ترکیب مجموعه داده‌های مختلف از چندین منبع داده‌باز دولتی تسهیل کند؛ همچنین، ارزش هر نوع داده هر بار که مورد استفاده مجدد قرار می‌گیرد و به منبع دیگری مرتبط می‌شود، افزایش می‌یابد و این امر می‌تواند با فراهم‌آوردن داده‌های آگاهی‌دهنده و روشن‌گر درباره هر مجموعه داده موجود، برای مثال ارائه فراداده‌ها، تسهیل و تحریک شود [۱۴]، [۱۵]، [۱۷]. سپس، این ایده‌ها به پیشرفت داده‌های باز پیوند شده (LOD)، که ترکیبی از هر دو هستند، منجر شد. این پیشرفت، ایجاد داده‌های ساخت‌یافته و در دسترس قراردادن بدون محدودیت آن‌ها برای استفاده مجدد دیگران را میسر ساخت. عمل به‌هم‌پیوستن داده‌ها برای کاهش موانع تکنولوژیکی و کاهش هزینه فرآیندهای انباشتگی داده‌ها، بسیار توصیه می‌شود [۱۵]، [۱۷]، [۵۰].

معنای استفاده و استفاده مجدد از این داده‌ها عبارت است از استفاده از آن‌ها به روش‌های جدید به‌واسطه افزودن ارزش به آن‌ها، ترکیب اطلاعات از منابع مختلف داده، ایجاد تلفیقی از عناصر نامتجانس و برنامه‌های جدید، برای اهداف تجاری و غیرتجاری [۱۳]، [۱۵]، [۳۰]. ایده اصلی داده‌های باز دولتی ساده است: داده‌های دولت بایستی [به‌سان] یک منبع مشترک باشند. ارائه داده‌های باز دولتی، نه تنها برای ادارات دولتی که این داده‌ها را جمع‌آوری و منتشر می‌کنند، بلکه برای شهروندان، کسب و کارها و سایر بخش‌های دولتی نیز واجد ارزش است؛ زیرا اگر داده‌های باز دولتی صرفاً منتشر شده و مورد استفاده قرار نگیرند، ارزش چندانی نخواهند داشت، و استفاده از آن‌ها به معنای درگیر شدن ذی‌نفعان و تمرکز بر توسعه اکوسیستم‌های پایدار کاربران است [۱۷]، [۵۲]. اساساً همکاری ذی‌نفعان مختلف در یک شبکه داده‌باز می‌تواند برای تسهیل استفاده از داده‌های باز دولتی (OGD) صورت گیرد. همچنین، ممکن است بین کسب و کارها برای استفاده از داده‌های باز رقابت وجود داشته باشد، برای مثال، رقابت بر سر به دست آوردن کاربران نهایی داده‌های باز به‌عنوان مشتریان (پرداخت‌کنندگان) برای خدماتی که این کسب و کارها براساس داده‌های باز دولتی (OGD) توسعه داده‌اند. همچنین، ممکن است بین ارائه‌دهندگان داده‌های باز رقابت وجود داشته باشد، زیرا آن‌ها ممکن است بخواهند با بیان اینکه مقادیر بیشتری از داده‌های باز یا مجموعه داده‌های بیشتری نسبت به سایر نهادها بخش دولتی در اختیار دارند، به تبلیغ سازمان خود بپردازند [۲۹]، [۵۴]، [۵۸].

۳ - ۲. مزایا، خطرات و تأثیرات بازکردن داده‌های دولتی

جانسن و همکاران [۲۰] و اوبالدی [۴۷] بحث جامعی را درباره چالش‌های حوزه داده‌های باز دولتی (OGD) ارائه داده‌اند. کوچرا و چلاپک [۲۶] مجموعه‌ای از مزایایی که می‌تواند به‌واسطه انتشار داده‌های باز دولتی (OGD) به دست آید، نمایش داده‌اند و همچنین مجموعه‌ای از خطرات را که باید هنگام ارائه مجموعه داده‌های باز در نظر گرفت، ذکر کرده‌اند. همچنین، کووان و همکاران [۱۱] با استفاده از چندین مثال عملی سعی دارند بسیاری از موضوعات مرتبط و فرصت‌های مشابه با داده‌های باز را به‌نمایش گذارند. علاوه بر این، نویسندگان مختلف تصدیق کرده‌اند که انتشار داده‌های دولتی در قالب‌های باز مزایای قابل توجهی برای شهروندان، کسب و کارها، محققان و سایر ذی‌نفعان ایجاد می‌کند تا از خلال تحلیل داده‌های پیشرفته، مسائل بخش دولتی و خصوصی را به روش‌های جدیدی درک کنند [۲۰]، [۵۶]. داده‌های باز به کاربران نهایی و به‌طورکلی جامعه کمک می‌کنند که به اطلاعات موردنیاز خود دست

یافته و آن‌ها را تلفیق کنند، و در نتیجه در زمینه گذار به سوی اقتصاد دانش‌بنیان و جامعه اطلاعاتی، کارایی بیشتر و مدیریت موفقیت‌آمیزتری داشته باشند [۵۴]، [۵۴]. با این حال، یانگ و کان‌کان‌هالی [۵۴] اظهار داشتند که علی‌رغم این که نهادهای دولتی به‌طور فعالانه با سازمان‌دهی داده‌هایی مانند مسابقات چالشی مختلف، استفاده از داده‌های خود را تبلیغ می‌کنند، هنوز پاسخی از سوی ذی‌نفعان خارجی در زمینه استفاده از داده‌های باز دولتی (OGD) به‌عنوان اهمی برای فعالیت‌های خلاقانه دریافت نشده است. همچنین، یافته‌های زویدروایک و جانسن [۵۶] با این دعوی موافق است که نتایج استفاده مجدد از داده‌ها مورد بحث قرار نگرفته است و ارائه‌دهندگان این داده‌ها (بخش دولتی) بازخورد اندکی دریافت کرده‌اند. بنابراین، نهادهای بخش دولتی بایستی از فرآیندهایی برخوردار باشند که به‌وضوح مشخص سازد کدام داده‌ها را در چه قالب‌هایی، در کدام بازه‌های زمانی و تحت کدام پروانه‌ها با کاربران به‌اشتراک بگذارند، تا بتوانند اطمینان حاصل کنند که محدودیتی در استفاده مجدد از این داده‌ها وجود ندارد [۴۹].

در حالت ایده‌آل، در دسترس قرار دادن این داده‌ها بر روی وب منجر به شفافیت، مشارکت و خلاقیت بیشتر در سراسر جامعه می‌شود. با این حال، فقط انتشار داده‌ها در وب کافی نیست. برای پیشرفت صادقانه جامعه باز ضرورت دارد که پلتفرم‌های انتشار، برخی از الزامات قانونی، اداری و همچنین فنی را برآورده سازند [۴]. در عمل، دستیابی به داده‌های خام، قرار دادن آن‌ها در یک بستر معنادار، پردازش و استخراج اطلاعات ارزشمند از آن‌ها اغلب بسیار دشوار است. در نتیجه، طی دو سال گذشته راه‌حل‌های مختلفی برای پشتیبانی از کل چرخه حیات استفاده مجدد از داده‌ها، برای مثال یافتن داده‌ها، تمیز کردن، تلفیق، پردازش و بصری‌سازی آن‌ها، عرضه شده است [۱۷]، [۲۳]. بر این اساس [۵۶]، فرآیند داده‌های باز شامل کلیه فعالیت‌هایی می‌شود که بین لحظه شروع ایجاد داده‌ها و لحظه‌ای که داده‌ها مورد بررسی قرار می‌گیرند، صورت می‌گیرد: از جمله فعالیت‌های مربوط به انتشار، یافتن و استفاده مجدد از داده‌ها. در این فرآیند نه تنها نشردهندگان و کاربران داده‌های باز، بلکه غالباً تعداد بیشتری از ذی‌نفعان نیز درگیر می‌شوند؛ از جمله تسهیل‌گران داده‌های باز، واسطه‌ها (سازمان‌هایی که کاربران و تولیدکنندگان داده‌های باز را گرد هم می‌آورند)، شهروندان، کسب‌وکارها یا قانون‌گذاران داده‌های باز (اتحادیه اروپا و احزاب سیاسی ملی) [۲۱]، [۵۶].

جتزک و همکاران [۲۱] یک مدل مفهومی را توسعه داده‌اند که نشان می‌دهد چگونه داده‌های باز به‌عنوان یک منبع می‌توانند به ارزش تبدیل شوند. آن‌ها روابط علی بین چهار عامل زمینه‌ای و توانمندساز (باز بودن داده‌ها، مدیریت داده‌ها، قابلیت‌ها و قابلیت اتصال فنی) و چهار نوع سازوکار تولید ارزش (کارایی، نوآوری، شفافیت و مشارکت) و ارزش را نشان داده‌اند. آن‌ها نتیجه گرفته‌اند که سازوکارهای تولید ارزش به عوامل توانمندساز وابسته هستند. اگر این رابطه به‌خوبی درک شود، انتخاب مجموعه داده‌های درست، پلتفرم‌های داده و رویه‌های مدیریت داده آسان‌تر است. در سال ۲۰۱۴، آن‌ها این مدل را با تمرکز بر تولید ارزش پایدار از خلال داده‌های باز در یک جامعه تسهیمی^۱ به‌روزرسانی کردند. آن‌ها این سازوکارهای تولیدکننده ارزش را معرفی کردند [۲۲]: شفافیت اطلاعات، تأثیرگذاری جمعی، کارایی داده محور و نوآوری داده محور. همه سازوکارها به بخش خصوصی و دولتی وابسته هستند و در کنار هم انگیزه، فرصت و توانایی تولید ارزش از داده‌ها را فراهم می‌کنند. آن‌ها همچنین مدعی هستند که انگیزه، فرصت و توانایی

۱ Sharing society: جامعه‌ای که در آن فعالیت‌های اجتماعی و اقتصادی حاوی تبادلات آنلاین است. م.

افراد در استفاده از داده‌ها برای تولید ارزش تحت‌تأثیر این عوامل قرار می‌گیرد [۲۲]: محرک‌های ارائه‌شده، سطح بازبودن فنی و قانونی داده‌ها، پختگی مدیریت داده‌ها، قابلیت‌های کلی مربوط به داده‌ها در جامعه، بلوغ و تفوق تکنولوژیکی. این مدل‌ها بعداً توسط بیلکووا و دیگران [۳] استفاده شده و بهبود یافتند. و ماچووا و لنیکا [۳۴] برای ارزیابی تأثیرات داده‌های باز در اقتصاد، آموزش، محیط‌زیست، بهداشت، سیاست و قانون‌گذاری، توسعه اجتماعی، بازرگانی و تجاری، این مدل را به‌کار گرفتند. مایر-رابلر و هوبر [۳۶] در مورد تأثیر داده‌های باز بر روابط بین شهروندان، بخش دولتی و مقامات سیاسی در راستای به‌کارگماشتن آن‌ها در همکاری، تصمیم‌گیری مشترک، توسعه مشترک و مسئولیت‌های مشترک بحث کردند؛ درحالی‌که، گایگر و فون لوكه [۱۵] ارزش افزوده آن دسته از داده‌های دولتی را که آزادانه قابل دسترسی هستند تحلیل و در مورد چالش‌های داده‌های باز دولتی (OGD) در سطوح مختلف مدیریتی بخش دولتی بحث کردند.

طی ۱۰ تا ۱۵ سال گذشته، چارچوب‌ها و شاخص‌های مختلف توسعه دولت الکترونیک برای کمک به ارزیابی فرصت‌ها و چالش‌های دولت الکترونیک و برنامه‌های وابسته به دولت باز، معرفی شده است. اوایل سال ۲۰۱۰ شاخص‌های جدیدی را به تحقیقات دولت الکترونیک اضافه کرده‌اند که بر ارزیابی تأثیرات داده‌های باز تمرکز دارند. برای مثال، این شاخص‌ها شامل موارد روبه‌رو می‌شود: (۱) فهرست شبکه و بارومتر داده‌های باز (ODB) تولیدشده توسط بنیاد وب جهان‌گستر (W3F)؛ (۲) بنیاد اطلاعات باز (OKF) شاخص جهانی داده‌های باز (GODI)، شاخص داده‌های دولتی باز، مفید و قابل استفاده مجدد (OURdata) معرفی شده توسط سازمان همکاری اقتصادی و توسعه (OECD)؛ و (۴) شاخص امتیازبندی اطلاعات بخش دولتی (PSIS) که توسط اتحادیه اروپا (EU) معرفی شده است [۳۵]. همچنین، این چارچوب‌ها در مطرح‌ساختن یک چارچوب محک‌زن جدید تأثیر دارند.

۳ - ۳. پرتال داده‌های باز، شرح و طبقه‌بندی

یکی از موضوعات اصلی در مورد راه‌اندازی یک دولت باز، دسترسی به داده‌های باز است که به‌طورکلی در پرتال‌های داده‌های باز گردآوری می‌شوند [۵۳]. بنابراین، یکی از نخستین مشکلاتی که هنگام کار با هر داده‌ای باید حل شود این است که در چه جایی می‌توان آن‌ها را پیدا کرد. آنچه کاربران در استفاده از داده‌ها دقیقاً به آن نیاز دارند، عبارت است از مجموعه داده‌های درستی که حاوی متغیرهای درست هستند، سال درست ثبت آن‌ها، طبقه‌بندی درست آن‌ها و غیره [۵۰]. با توجه به تعریف داده‌های باز، این داده‌ها بایستی به‌خوبی شرح داده شده و از کیفیت خوبی برخوردار باشند تا افراد بتوانند آن‌ها را به دانش تبدیل کرده و استفاده مفیدی از آن‌ها به‌عمل آورند. در چندسال گذشته [۱۰]، تعداد فزاینده‌ای از دولت‌ها پرتال‌های داده‌های باز و وب‌سایت‌های تخصصی را راه‌اندازی کرده‌اند که در آن‌ها یک رابط انتشار، امکان بارگذاری مجموعه داده‌ها و مجهز شدن به فراداده‌هایی با کیفیت بالا را فراهم می‌کند [۴۷]، [۵۰]. پرتال داده‌های باز یک سیستم مبتنی بر وب است که مجموعه داده‌های موجود را از چندین منبع که ممکن است فرمت‌های مختلفی داشته باشد، گردآوری می‌کنند و بر روی داشبوردهای کاربرپسند منتشر می‌کنند تا کاربران بتوانند این مجموعه داده‌ها از طریق رابط برنامه‌نویسی کاربردی (API) مشاهده و بارگیری کرده، به آن‌ها دسترسی پیدا کنند، و آن‌ها را منتشر کنند. این مجموعه داده‌ها، با برچسب‌های تعریف‌شده توسط کاربر، در یک کاتالوگ قابل جست‌وجو

سازمان‌دهی می‌شوند [۲۷]، [۵۰]. این مجموعه داده‌ها می‌توانند به‌واسطه یک اپراتور کاتالوگ که ممکن است ابتکار عمل یک آژانس دولتی، یک شهروند یا غیره باشد، اداره شود. هر پرتال مجموعه داده‌های مختلفی را عرضه می‌دارد که به‌طور مستقیم در دسترس بودن داده‌ها برای افشاکری عمومی را بازتاب می‌دهد [۲۷]. مجموعه داده واقعی را نمی‌توان بخشی از پیشینه کاتالوگ در نظر گرفت، اما پیشینه کاتالوگ معمولاً حاوی یک لینک بارگیری یا پیوندی به صفحه وب است که از آن‌جا می‌توان مجموعه داده واقعی را به‌دست آورد [۱۲]. هر مجموعه داده می‌تواند چندین منبع داده را در بر داشته باشد [۵۰].

پرتال‌های داده باز برای دسترسی بهتر به داده‌ها، مجموعه داده‌های باز را با توجه به حوزه، ارائه‌دهندگان، فرمت و سایر خصوصیات آن‌ها دسته‌بندی می‌کنند [۵۳]. پرتال‌های داده باز همچنین معمولاً دارای کلیدواژه‌های جست‌وجو و رابط‌های مرورگر مختلفی هستند تا به کاربران در یافتن مجموعه داده‌های مرتبط و بازیابی فراداده‌های متناظر برای ارائه شرحی از نهاد منتشرکننده مجموعه داده‌ها و همچنین محتوای مجموعه داده و حوزه جغرافیایی، حوزه قضایی و دوره زمانی آن، کمک کنند [۵۱]. فرمت مجموعه داده‌ها نیز نیاز به توجه دارد، زیرا ممکن است این مسئله به بسیاری از مشکلات مربوط به هم‌کنش‌پذیری (تعامل‌پذیری) و تلفیق داده‌ها منجر شود [۵۱]. سایر مسائل مربوط به بستر مجموعه داده‌ها، به کامل بودن، نمایش داده‌های باز، اعتبار، قابلیت اطمینان، صراحت و جامعیت داده‌ها و عرضه‌داشت گزارشاتی درباره تحلیل این داده‌ها ربط پیدا می‌کند. در کنار مسائلی که به محتوای داده‌ها مربوط‌اند، کیفیت کلی داده‌ها را نیز بایستی در نظر گرفت [۱۳]، [۱۶]، [۱۹]. همچنین، استانداردهای داده‌ها، گدها، واژگان و الگوها از جنبه‌های مهم مجموعه داده‌ها هستند [۵۱]. پرتال‌های متعددی وجود دارند که به‌جای در دسترس قرار دادن مستقیم داده‌های خام برای بارگیری، API‌هایی را برای داده‌های مبتنی بر وب ارائه می‌دهند که توسعه‌دهندگان را قادر می‌سازد به داده‌های برنامه‌های خود دسترسی پیدا کنند [۲۸]. یک شرح بسنده از پرتال باید قادر باشد تا مضامین را از کلیدواژه‌ها به‌وضوح متمایز کند. مضامین همواره از میان واژه‌های کنترل‌شده انتخاب می‌شوند، منتهی برچسب‌ها این‌گونه نیستند [۳۳]. علاوه بر این مهم است که داده‌ها از یک فرمت خوانا برای رایانه برخوردار باشند تا بتوانند از ابزارهای خودکار پشتیبانی کنند. ویژگی‌های مربوط به اعتبار، کیفیت و دانه‌بندی^۱ از آن‌رو ضروری هستند که بتوانند از طیف گسترده‌ای از موارد استفاده و نیز از امکان تحلیل نتایج با کیفیت بالا، پشتیبانی کنند [۴].

تا چندی پیش، فراداده‌ها تنها یک دغدغه مربوط به متخصصان اطلاعات بود که در زمینه فهرست‌نویسی، طبقه‌بندی، و شاخص‌سازی فعالیت می‌کردند. با این‌همه، امروزه تعداد بیشتری از تولیدکنندگان و مصرف‌کنندگان محتوای دیجیتال وجود دارد که به فهرست‌بندی نیاز دارند [۱۷]. بدون وجود فراداده‌های کافی، مانند شرح‌ها و برچسب‌ها، کاربران چه به‌واسطه جست‌وجوی دستی و چه از طریق جست‌وجوی خودکار، قادر به یافتن مجموعه داده‌ها نخواهند بود و این داده‌ها برای آن‌ها مفید واقع نمی‌شود [۴]. ساختار فراداده‌ای پرتال داده، خصوصیات رایجی را که برای توصیف هر مجموعه داده در یک پرتال معین به‌کار گرفته می‌شود، خلاصه می‌کند. این خصوصیات عمدتاً شامل

ویژگی‌هایی مانند نام و شرح مجموعه داده، و URL^۱ منابع واقعی، برای مثال فایل‌ها یا نقطه دسترسی کاربران به سرویس می‌شود. با استفاده از این فراداده‌ها، کاربران می‌توانند با جست‌وجو و فیلترکردن، به سرعت مجموعه داده‌های مورد نیاز را پیدا کنند [۵۰]. از منظر معناشناسی فراداده‌ها، مهم‌ترین ابتکاری که یک پرتال داده باید برای تسهیل هم‌کنش‌پذیری انجام دهد، ایجاد یک واژگان RDF^۲ به نام واژگان کاتالوگ داده‌ها (DCAT) است که توسط کنسرسیوم وب جهان‌گستر (W3C) معرفی شده است. نشردهندگان داده‌ها با استفاده از واژگان کاتالوگ داده‌ها (DCAT) برای توصیف مجموعه داده‌ها، قابلیت یافتن داده‌ها را افزایش می‌دهند و اپلیکیشن‌ها را به راحتی به استفاده از فراداده‌هایی از کاتالوگ‌های مختلف قادر می‌سازند [۱۲]، [۳۳]. هم‌چنین، برخی از نویسندگان، مانند مالی و همکاران [۳۳]، واژگان DCAT RDF خود را به عنوان یک فرمت تبادل داده ارائه داده‌اند تا بدین طریق بتوان یک شرح استاندارد از کاتالوگ داده‌ها ارائه داد.

پرتال‌های داده‌باز را می‌توان برحسب پوشش جغرافیایی (سطح اداری)، به گروه‌های زیر تقسیم کرد [۲۷]: محلی، که متعلق به شهرها / شهرستان‌ها است یا فقط به پوشش شهر / شهرستان تعلق دارد؛ منطقه‌ای، که متعلق به یک قدرت منطقه‌ای است (حاکمیت شهرستان یا دولت ایالتی فدرال) یا پوشش منطقه‌ای دارد؛ ملی، که متعلق به یک نهاد دولتی مرکزی یا با پوشش سراسر کشور است؛ و بین‌المللی، که متعلق به یک نهاد بین‌المللی یا با پوشش بین‌المللی است. سپس، لنیکا و ماچووا [۳۱] این طبقه‌بندی را با افزودن سطح جدیدی از گردآورندگان داده‌های باز به عنوان یک دسته اصلی و مهم‌ترین کاتالوگ داده گسترش دادند [۱۹]، [۳۱]. دسته‌بندی‌های دیگری از منابع داده‌های باز را می‌توان براساس پارادایم وب انجام داد، یعنی آن‌ها براساس پارادایم سنتی Web 1.0 یا پارادایم جدیدتر Web 2.0 دسته‌بندی می‌شوند [۱]. کولپرت و همکاران [۱۰]، برحسب قابلیت استفاده از پرتال‌های داده‌های باز، یک سیستم پنج‌مرحله‌ای را ارائه دادند که کاربرد آن نشان‌دادن کارکرد اصلی یا قابلیت است که پرتال داده برای آن ساخته شده یا مورد استفاده قرار می‌گیرد. این مراحل با توجه به زمانی که لازم است برای اجرای کامل هر مرحله صرف شود، تنظیم می‌شوند. این دسته‌بندی با پرتال‌هایی که به مجموعه داده‌های مختلف پیوند می‌خورند، شروع می‌شود و با پرتال‌های فراداده‌ای برای مجموعه داده‌ها و استفاده مجدد از مجموعه داده‌ها ادامه می‌یابد. دسته چهارم خود به انتشار داده اقدام می‌کند. سرانجام، یک مرکز داده^۳ که داده‌ها در آن به یک منبع مشترک تبدیل می‌شوند، راه‌اندازی می‌شود.

پرتال داده‌باز یکی از راه‌حل‌هایی است که کاربرد آن، قابلیت یافتن مجموعه داده‌هایی را که به طور رایگان در دسترس هستند، به طرز قابل توجهی تقویت می‌کند [۲۷]. با این حال، بسیاری از دولت‌ها بر توسعه یک پرتال ملی داده‌های باز دولتی (OGD) تمرکز می‌کنند، گویی که این کار بیش از توسعه زیرساخت‌های فنی برای ارائه داده‌های باز دولتی در راستای استفاده دیگران، اولویت دارد [۴۷]. زیرساخت‌ها می‌توانند با ارائه بینشی درباره چگونگی مشارکت افراد در

۱ Uniform Resource Locator (URL): منبع یاب یک‌نواخت. م.

۲ Resource Description Framework (RDF): چارچوب توصیف منبع. م.

استفاده مجدد از داده‌ها و درباره کیفیت داده‌های باز، استفاده از داده‌های باز دولتی (OGD) را بهبود بخشند [۵۷]. درک پیش‌شرط‌های داده‌های باز دولتی (OGD) کارآمد در یک زمینه خاص، برای راه‌اندازی وبسایت‌هایی که ایجاد ارزش را امکان‌پذیر می‌کنند، ضروری است و درک این پیش‌شرط‌ها هسته اصلی مسئولیت انتشار داده‌های دولتی را می‌سازد. بیشتر انتقادات موجود در مورد پرتال‌های ملی داده‌های باز دولتی (OGD) بر این واقعیت استوار است که به نظر می‌رسد دولت علاقه‌مند است که داده‌ها را با اسلوب خاصی ارائه دهد، و این امر به دورشدن از هدف ارائه داده‌های بیشتر به کاربرانی که واقعاً علاقه دارند تا از این داده‌ها برای اهداف خود استفاده کنند منجر شده، و ارائه این داده‌ها را محدود می‌کند [۴۷]. جانسن و همکاران [۲۰] چندین عامل از جمله فقدان توضیح برای معنای داده‌ها، و فقدان دانش برای درک داده‌ها را برمی‌شمرند که از استفاده عمومی از داده‌های باز ممانعت به عمل می‌آورند. مارتین و همکاران [۳۷] هفت دسته از خطرات مرتبط با پرتال‌های داده باز را معرفی می‌کنند: مدیریت داده‌ها، مسائل اقتصادی، پروانه‌ها و چارچوب‌های قانونی، مشخصه داده‌ها، فراداده‌ها، دسترسی و مهارت‌ها. همچنین، به نظر می‌رسد عوامل مختلفی از عوامل نهادی گرفته تا عوامل فنی بر توسعه و اجرای پرتال‌های داده باز دولتی (OGD) در سطح ملی تأثیر می‌گذارند [۲۰]. بنابراین، منطقی است که استدلال کنیم کشورهای مختلف در توسعه و اجرای سیستم داده‌های باز دولتی (OGD) خود، از توانایی‌های متفاوتی برخوردار هستند [۴۴].

۳-۴. الزامات ارزیابی و کیفیت پرتال‌های داده باز

با افزایش تعداد و تنوع منابع داده، ضرورت ایجاد فراداده مناسب (شرح، پوشش جغرافیایی، محدودیت‌ها) بیشتر می‌شود تا ذی‌نفعانی که در این عرصه متخصص نیستند بتوانند به‌سادگی داده مورد نظرشان را بیابند و از آن استفاده کنند. این که داده‌های افشا شده بایستی کیفیت داشته و خوب باشند، ایده مبرهنی به نظر می‌رسد. اما تحقق آن در بستر دولت باز ساده نیست. علاوه بر آن داده باکیفیت می‌تواند ویژگی‌های مختلفی داشته باشد. ریسک (فرا) داده با کیفیت پایین بر کشف و مصرف یک مجموعه داده در یک پرتال و بین پرتال‌ها اثر می‌گذارد. فقدان فراداده از سویی مستقیماً بر جست‌وجو و یافتن مجموعه داده‌های مرتبطی که می‌تواند مورد نیاز یک کاربر باشد اثر می‌گذارد. از سوی دیگر، شرح نادرست بر روی مجموعه داده‌ها می‌تواند پردازش و تلفیق آن‌ها با مجموعه داده‌های دیگر را با چالش مواجه کند [۴۸].

ادبیات مربوط به کیفیت داده‌ها، ویژگی‌های مختلفی برای کیفیت داده مطرح کرده‌اند. این ویژگی‌ها اغلب وابسته به دیدگاه نویسندگان و بستر مورد نظر آن‌ها است. گرایش دیگری که در ارزیابی کیفیت داده می‌توان مشاهده کرد، تمایل به مورد توجه قراردادن دیدگاه کاربران است. در واقع، مرسوم‌ترین تعریف از کیفیت داده این است که داده‌ها بایستی متناسب با کاربرانشان باشند، یعنی داده‌های گردآوری شده با نیازهای کاربر متناسب داشته باشد [۷]. کیفیت داده در پژوهش‌های پیشین با چند بُعد تعریف شده است. این ابعاد نشانگر پایایی داده‌های ساخته شده هستند [۲]، [۶]. بتینی و همکاران [۲] مطالعه‌ای دقیق و نظام‌مند بر روی روش‌های ارزیابی و ارتقای کیفیت داده انجام داده‌اند. این روش‌ها از چند بُعد مختلف مقایسه شده‌اند، از جمله فازها و گام‌های روش، استراتژی‌ها و تکنیک‌ها، ابعاد کیفیت داده، نوع داده و نهایتاً نوع سامانه اطلاعاتی که در هر روش به آن پرداخته می‌شود. کابالیرو و همکاران [۶] به ارزیابی

کیفیت داده پرداختند. آن‌ها مدعی شدند که ضرورت ارزیابی کیفیت استفاده از مجموعه داده‌ها بیش از هر زمانی است، چراکه نقشی که یک مجموعه داده می‌تواند در ارزش‌آفرینی ایفا کند تنها در بستر استفاده از آن قابل‌تخمین است. مهم‌ترین ویژگی در ارزیابی کیفیت استفاده از مجموعه داده‌های غیرهمگون، پایایی است. پایایی دارای سه بخش است: زمینه‌ای، زمانی و عملکردی. شواهد بیش‌تر در تأیید اهمیت کیفیت داده را می‌توان در پژوهش‌تین مشاهده کرد [۴۶].

سنجه‌های کیفیت داده را می‌توان در مورد حوزه‌های باز نیز بکار بست. ضوابط و سنجه‌های کیفیت داده در مورد داده‌های باز اخیراً برای ارزیابی کیفیت مجموعه داده‌های باز و نیز پرتال‌ها شکل گرفته است [۵۳]. براساس پیشینه پژوهشی تحلیل‌شده، پرتال‌های داده‌باز برای این‌که بتوانند شرایط اطلاعات باکیفیت را تأمین کنند، بایستی براساس ساختار نهادی هر کشور، به تمامی نهادها و الزامات آن‌ها بپردازند [۳۲]. همچنین زبانی که هر پرتال داده‌باز به آن وابسته است، یکی از چالش‌های ذاتی کیفیت محسوب می‌شود [۴۲]. پژوهش‌ها در مورد کیفیت پرتال‌های داده‌های باز تأیید می‌کنند که موج اول پرتال‌ها تنها از امکانات اولیه‌ای همچون بارگذاری و بارگیری داده برخوردار بودند [۱]، [۸]. پرتال‌های امروزی و زیرساخت آن‌ها اغلب اجازه مشارکت در بهبود کیفیت مجموعه داده‌های منتشر شده را به کاربر نمی‌دهند [۱]، [۵۷]. لورنچیو [۳۲] این هشت مشخصه اساسی را برای پرتال‌های داده‌باز تعریف کرده است: کیفیت، کامل بودن، دسترس‌پذیری و رؤیت‌پذیری، کاربردپذیری و جامعیت، بروزبودگی، ارزشمندبودن و مفیدبودن، دانه‌بندی و مقایسه‌پذیری. اگرچه موفقیت داده‌های باز به کیفیت مجموعه داده‌های منتشرشده وابسته است، کیفیت مجموعه داده‌های منتشر شده بسیار با هم تفاوت دارد و کاربران نیز ممکن است در مورد کیفیت داده‌های باز نگرانی داشته باشند [۳۷]، [۴۲]، [۵۷].

تجربه‌های اخیر نیز نشان می‌دهد که کیفیت فهرست داده‌ها می‌تواند بر توانایی کاربران در یافتن جای داده مورد نظرشان کمک کند. به همین خاطر، کوچرا و همکاران [۲۷] در مورد محدودیت‌هایی که کیفیت این فهرست می‌تواند داشته باشد بحث می‌کنند. این محدودیت‌ها عبارت‌اند از: ناقص بودن شرحی که در فهرست نوشته شده و عدم بررسی صحت این شرح. آن‌ها تکنیک‌هایی برای رفع این محدودیت‌ها پیشنهاد می‌دهند. منتشرکنندگان داده‌های باز و کاربران آن اغلب از نیازها و فعالیت‌های یکدیگر اطلاع ندارند. برای مثال، بسیاری از ارائه‌دهندگان داده‌باز اغلب تنها دغدغه در اختیارگذاشتن داده را دارند و نمی‌دانند که کاربران چه فرمتی را ترجیح می‌دهند. همچنین نمی‌دانند که نحوه انتشار داده در تشویق به استفاده از داده‌باز اهمیت دارد [۵۶]. لورنچیو [۳۲] مجموعه‌ای از شرایط لازم در مورد فهرست داده‌ها را قید می‌کند که برای کمک به شفافیت سازمان‌های دولتی مورد نیاز است و به نهادهای بخش دولتی اجازه می‌دهد که پاسخگو شوند. این الزامات به نوع نهادهایی که مجموعه داده‌ها پوشش داده‌اند، نوع داده ارائه شده، استراتژی‌های جست‌وجوی اطلاعات مورد پشتیبانی و برخی جنبه‌های کیفی داده‌های باز ارائه شده بستگی دارد. کوچرا و همکاران چهار شرط برای کیفیت مجموعه داده تعریف می‌کنند [۲۷]: صحت، تمام اطلاعات آمده در فهرست باید مطابق با داده‌ای باشد که توصیف شده است و باید تمامی موارد ذکر شده در فهرست صحت داشته باشد؛ کامل بودن، تمامی ویژگی‌های اجباری فهرست باید پُر شده باشد و تمامی مجموعه داده‌های باز دولتی منتشر شده

باید در فهرست ثبت شده باشند و نباید موردی تکراری در فهرست بیاید؛ پایایی، بایستی از اصطلاحات یا مفاهیمی یکسان برای طبقه‌بندی داده‌هایی که به یک دسته یا نوع تعلق دارند استفاده شود، به داده‌های گم‌شده باید در کل فهرست به یک صورت پرداخته شود؛ و بروز بودن، تمامی اطلاعات آمده در فهرست داده‌ها باید بروز باشد.

برای این‌که کاربران بتوانند کیفیت داده را ارزیابی کنند، باید ماهیت داده را بشناسند، چراکه تولیدکنندگان داده‌ها نمی‌توانند تمامی کاربران را پیش‌بینی کنند. ارائه فراداده باکیفیت به همان اندازه کیفیت داده اهمیت دارد [۵۵].

باین‌حال، نقش نهادهای دولتی در استراتژی‌ها و برنامه‌های داده باز تنها به انتشار داده ختم نمی‌شود. به‌غیر از افزایش تنوع و ارتقای کیفیت داده‌ای که در دسترس عموم قرار می‌گیرد، نهادهای دولتی کوشیده‌اند ذی‌نفعان بیرونی را به استفاده از داده‌های باز برای فعالیت‌های نوآورانه تشویق کنند [۴۰]. کوچرا و همکاران [۲۷] دو نوع استراتژی برای ارتقای کیفیت داده شناسایی کرده‌اند: استراتژی مبتنی بر داده و استراتژی مبتنی بر فرآیند. اولی مبتنی بر اصلاح مستقیم داده است. کارهایی چون تصحیح مقادیر نادرست داده یا نرمالیزاسیون آن. دومی مبتنی بر طراحی مجدد فرآیند خلق داده و فرآیندهای اصلاح آن با هدف شناسایی و اصلاح مشکلات مربوط به کیفیت است. کارهایی چون افزودن یک گام تأیید اعتبار به فرآیند گردآوری داده. امروزه تلاش‌های بسیاری برای مطالعه جنبه‌های مختلف پرتال‌های داده باز انجام می‌شود. این پرتال‌ها پلتفرم‌های اصلی انتشار و استفاده از مجموعه داده‌ها هستند [۴۸].

۳-۵. محک‌زن‌ها و مدل‌ها برای ارزیابی پیشرفت پرتال‌های داده باز

از زمان راه‌اندازی اولین پرتال داده باز توسط دولت ایالات متحده در سال ۲۰۰۹ و بریتانیا در سال ۲۰۱۰، شمار بیشتری از کشورها چنین طرح‌ها و پرتال‌های داده‌ای ایجاد کرده‌اند. هدف آن‌ها تسهیل استفاده از این داده‌ها توسط عموم مردم است. داده‌ها در فرمت‌ها و حوزه‌های متنوعی ارائه می‌شوند [۵۴]. این مسئله همسو با یافته‌های پژوهش آمبریچ و همکاران [۴۷] است که مشاهده کردند شمار داده‌ها و منابع، پیوسته افزایش می‌یابند. نمونه‌های افزایش محبوبیت پرتال‌های داده عبارت‌اند از: پرتال‌های داده باز دولتی (OGD) [۲۰]، پرتال‌های داده سازمان‌های بین‌المللی و سازمان‌های مردم‌نهاد، پرتال‌های داده‌های علمی و نیز فهرست داده‌های کلان در کسب‌وکارهای بزرگ [۱۴]، [۱۹]. بسیاری از کشورها، از جمله بسیاری از دولت‌های عضو اتحادیه اروپا همراه با برخی دولت‌های محلی (مانند شهرداری‌ها) از این روند تبعیت کرده‌اند [۵۰]. بسیاری از این پرتال‌ها از شبکه جامع آرشيو اطلاعات (CKAN) استفاده می‌کنند که یک پلتفرم داده رایگان و متن‌باز است که توسط Open Knowledge توسعه یافته و حفظ و نگهداری می‌شود. در نتیجه، آن‌ها یک رابط برنامه‌نویسی کاربردی (API) قدرتمند و استاندارد دارند که امکان ترکیب فهرست داده‌های‌شان را فراهم می‌کند. بدین ترتیب می‌توان یک نقطه ورود واحد برای یافتن و استفاده از داده‌های دولتی در سطح جهان ایجاد کرد. دیگران از شبکه آرشيو اطلاعات دروپال (DKAN)، پلتفرم دولت باز، Socrata، Prognoz یا Junar استفاده می‌کنند. کتابخانه‌های دیجیتال و شبکه‌های فهرست داده‌ای مشابه می‌توانند از شرح، بایگانی و یافتن داده در وب پشتیبانی کنند [۱۴]، [۵۰]. علاوه بر پرتال‌هایی که به‌طور رسمی از حمایت بخش دولتی یا خصوصی برخوردار هستند، تعداد زیادی منبع داده باز غیررسمی وجود دارد که توسط شهروندان، اجتماعات یا جمیع‌کنندگان داده پشتیبانی می‌شوند.

کالامپکیس و همکاران [۲۴] مدل‌های مرحله‌ای دولت الکترونیک موجود را بازبینی کردند و مدل مرحله داده باز دولتی را پیشنهاد دادند که راهنمایی برای استفاده مجدد از داده باز دولتی فراهم می‌کند و امکان ارزیابی پیشرفت برنامه‌های مربوطه را فراهم می‌کند. سولار و همکاران [۴۵] یک مدل بلوغ داده باز برای ارزیابی تعهد و قابلیت‌های نهادهای دولتی در پیگیری اصول و شیوه‌های داده باز پیشنهاد کردند. این مدل ساختاری سلسله مراتبی دارد و متشکل از دامنه‌های اینترنتی، دامنه‌های خرد و متغیرهای اساسی است. الکسپلوس و همکاران [۱] مدلی جدید برای پرتال داده دولتی طراحی کردند. این مدل کاربری پرتال‌ها را افزایش می‌دهد. قابلیت‌های آن عبارت‌اند از: پردازش داده، بهبود مدل‌سازی داده (فراداده ساده، زمینه‌مند و دقیق)، نظر گذاشتن بر روی مجموعه داده‌های موجود و اعلام نیاز به مجموعه داده‌های تازه، نمره‌دهی به کیفیت مجموعه داده‌ها، تشکیل گروه توسط کاربران و امکان همکاری و برقراری ارتباط میان آن‌ها، پیونددهی داده، بارگذاری نسخه‌های جدید از مجموعه داده‌های موجود و بهبود بصری‌سازی داده. ون در وال و همکاران [۵۰] کاربردهای اصلی پرتال‌های داده باز را شرح می‌دهند و مدلی مفهومی برای بدل‌ساختن این پرتال‌ها به شالوده داده‌های جهانی برای جامعه اطلاعاتی در وب ارائه می‌کنند. چارالابیدیس و همکاران [۸] روشی برای ارزیابی زیرساخت‌های پیشرفته و نسل جدید داده‌های باز دولتی و پرتال‌های داده باز ارائه کرده و اعتبار آن را می‌سنجند. این روش مبتنی بر تخمین ارزش مدل‌ها براساس نمره‌دهی کاربران است. آن‌ها نتیجه گرفتند که بایستی به ارتقای بارگذاری داده و قابلیت‌های جست‌وجو و بارگیری داده بیشترین اهمیت داده شود، چراکه کاربران پایین‌ترین نمره را به این موارد داده بودند و این موارد تأثیر زیادی بر ایجاد لایه‌های بالاتر دارند.

یکی از اولین بررسی‌های تطبیقی بر روی پرتال داده‌های باز را مالی و همکاران در سال ۲۰۱۰ انجام دادند [۳۳]. آن‌ها قصد داشتند که وجوه اشتراک و هم‌پوشانی‌های ساختار و مستندسازی داده‌ها و نیز چالش‌ها و الگوها را شناسایی کنند. هرچند که تنها هفت پرتال داده از پنج کشور با یکدیگر مقایسه شدند. سایوگو و همکاران [۴۴] از تحلیل محتوای وب برای شناسایی کاربرد دستکاری داده و قابلیت استفاده از پرتال‌های داده باز ۳۵ کشور استفاده کردند. ورما و گوپتا [۵۱] سی پرتال داده باز در سطح کشورها را برای یافتن تنوع فرمت‌های مختلف انتشار داده بررسی کردند. یافته‌های آن‌ها نشان داد که به‌طور کلی توسعه پرتال‌های داده باز از رویکرد افزایشی مشابه مراحل توسعه دولت الکترونیک پیروی می‌کند. علاوه بر آن، آمبریچ و همکاران [۴۸] پیوسته کیفیت ۸۲ پرتال داده باز فعال را که قدرت گرفته از شبکه جامع آرشيو اطلاعات (CKAN) و متعلق به ۳۵ کشور مختلف بودند، مورد نظارت و ارزیابی قرار دادند. برای این منظور آن‌ها شش بُعد کیفیتی و سنج‌های مرتبط را معرفی کردند: قابلیت بازیابی، کاربردی، کامل بودن، صحت، باز بودن و قابلیت برقراری ارتباط. نتایج آن‌ها نشان می‌دهد که اطلاعات رشدی منظم داشته است و جنبه‌های مختلف این پرتال‌ها مانند باز بودن، قابلیت برقراری ارتباط و دسترس‌پذیری فراداده بسیار تنوع دارد. باین حال بررسی آن‌ها مربوط به پرتال‌هایی بود که به سطوح مختلف اداری تعلق داشتند.

زویدروایک و جانسن [۵۷] قابلیت استفاده از سازوکارهای مشارکت و شاخص‌های کیفیت داده برای پرتال‌های داده باز را با استفاده از شش ضابطه ارزیابی کردند، در حالی که لورنچیو [۳۲] ارزیابی کرد که آیا ساختار و سامان کنونی هفت پرتال دولت باز برای شفافیت مورد نیاز برای پاسخ‌گویی کافی هستند یا خیر. این نویسنده مجموعه‌ای از شرایط

را مطرح می‌کند که مبتنی بر مشخصه‌های مطلوب افشای داده‌ای هستند و از ادبیات مربوط به دولت باز و ارزیابی شفافیت استنتاج شده بودند. این شرایط به‌عنوان چارچوبی برای تحلیل ساختار و سامان داده در این پرتال‌ها مورد استفاده قرار گرفت. یانگ و همکاران [۵۳] ساختارهای طبقه‌بندی پرتال‌های داده باز را با بررسی انسجام آن‌ها مقایسه کردند. منظور از انسجام، شباهت مجموعه داده‌های موجود در یک دسته است. براون‌اشوایگ و همکاران [۴] پیمایشی در مورد پلتفرم‌های داده باز دولتی موجود با تمرکز بر جنبه‌های فنی ارائه کردند. آن‌ها بیش از ۵۰ پرتال را مورد مطالعه قرار دادند که توسط دولت‌های ملی، منطقه‌ای یا محلی و نیز سازمان‌های بین‌المللی مدیریت می‌شدند. تمرکز آن‌ها بر مشخصه‌هایی چون استانداردسازی، قابلیت یافتن اطلاعات و خوانایی داده برای رایانه بود.

لنیکا و ماچووا [۳۱] پرتال‌های داده باز ملی منتخب‌شان را ارزیابی کردند. با این حال آن‌ها تنها دولت‌های عضو اتحادیه اروپا را با پنج ضابطه مورد بررسی قرار دادند. لنیکا [۳۰] بعدتر لیست پرتال‌ها را به ۶۷ کشور افزایش داد اما مقایسه‌ای انجام نداد. پتیچاکیس و همکاران [۴۲] منابع داده باز دولتی (OGD) توسعه‌یافته در ۲۷ کشور اتحادیه اروپا را از منظر کارایی، معناساختی و فنی تحلیل کردند محتوای موضوعی، صدور پروانه، چندزبانه‌بودن، کسب داده، یافتن داده، ارائه و فرمت داده مورد توجه آن‌ها بود. آنان نتیجه گرفتند که غالب مجموعه داده‌های منابع داده‌های باز دولت (OGD) اروپایی بدون پروانه‌ای تعریف‌شده یا بدون پروانه باز منتشر می‌شوند و حدود نیمی از این پرتال‌ها تنها زبان کشور خودشان را پشتیبانی می‌کند، در حالی که نیم دیگر چندزبانه هستند (از یک یا تعداد بیشتری زبان خارجی نیز پشتیبانی می‌کنند).

با این حال، هیچ‌یک از پژوهش‌های ذکرشده پیشین کیفیت پرتال‌های داده باز و مجموعه داده‌های مربوطه در سطح ملی را بررسی نکرده‌اند. هم‌چنین این مقالات حداکثر از ۶۰ پرتال داده در سطحی مشابه استفاده کرده‌اند. بر همین اساس و با استفاده از مرور انجام شده یک چارچوب محک‌زنی جدید برای رفع این محدودیت‌ها ارائه می‌شود. تنها از پرتال‌های داده باز سطح کشوری برای بررسی کاربردپذیری این چارچوب پیشنهادی استفاده می‌شود. علاوه‌براین، پرتال‌هایی که کمتر از پنجاه مجموعه داده منتشر کرده بودند حذف شدند، چراکه فاقد محتوای کافی بودند. ۶۷ کشور مورد ارزیابی قرار گرفته‌اند.

۴. نتایج ارزیابی کیفیت پرتال‌های داده باز

تنها پرتال‌های داده باز در سطح ملی ارزیابی شده‌اند. پرتال‌های بین‌المللی، منطقه‌ای یا محلی داده باز و نیز مؤسسات آماری ملی یا پرتال ادارات که داده باز ارائه می‌کنند در پژوهش ما گنجانده نشده‌اند. مقایسه بر مبنای رده‌بندی شاخص داده‌های باز دولتی (OGD) و شاخص بارومتر داده باز (ODB) در سال ۲۰۱۵ انجام شد. این شاخص‌ها وضعیت داده باز را در کشورهای منتخب ارزیابی می‌کنند. در مجموع ۱۴۰ کشور را پوشش می‌دهند. فرایند روایی و اعتبار پرتال داده باز شامل این مراحل است: یک کلمه کلیدی شامل نام کشورهای ذکرشده در بالا با واژه‌های داده باز و پرتال داده باز در موتور جست‌وجوی گوگل وارد می‌شود؛ کشور منتخب با لیست موجود در سایر منابع مانند وب‌سایت ۱، ۲ و ۳ مقایسه می‌شود و URL شناسایی‌شده پرتال برای بررسی این که در حال حاضر کار می‌کند یا خیر باز می‌شود. این پرتال‌ها که در ابتدای سال ۲۰۱۵ شناسایی شده بودند در مقاله لنیکا [۳۰] آمده‌اند. بعداً در همان

سال یک فرآیند جست‌وجوی دیگر ایجاد شد و ۲۴ پرتال داده‌باز دیگر در سطح ملی یافت شد. تعداد کل پرتال‌های داده‌باز به ۹۱ رسید.

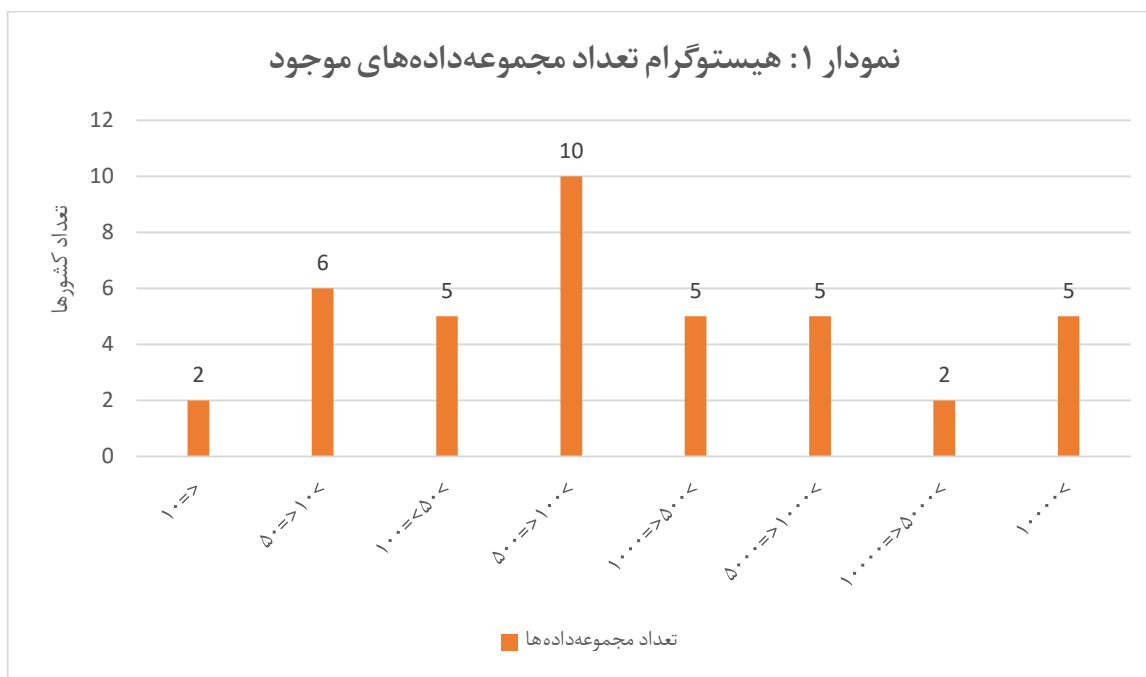
۴ - ۱. ارزیابی کیفیت پرتال‌های داده‌باز از طریق رابط برنامه‌نویسی کاربردی (API)

اولین مرحله ارزیابی کیفیت پرتال‌های داده‌باز، تحلیل محتوای این پرتال‌ها برای مقایسه آن‌هاست. برای این منظور، از رابط برنامه‌نویسی کاربردی (API) برای دسترسی به این پرتال‌ها استفاده می‌شود. از همین رو، تنها پرتال‌هایی که رابط برنامه‌نویسی کاربردی (API) آن‌ها در دسترس است ارزیابی خواهند شد. بیشتر این ۵۲ پرتال داده‌باز قدرت گرفته از شبکه جامع آرشیو اطلاعات (CKAN) یا DKAN^۱ هستند. از انتقال بازنمودی حالت (REST) پست‌مان به‌عنوان ابزار اصلی حصول نتایج با فرمت JSON استفاده شده است.

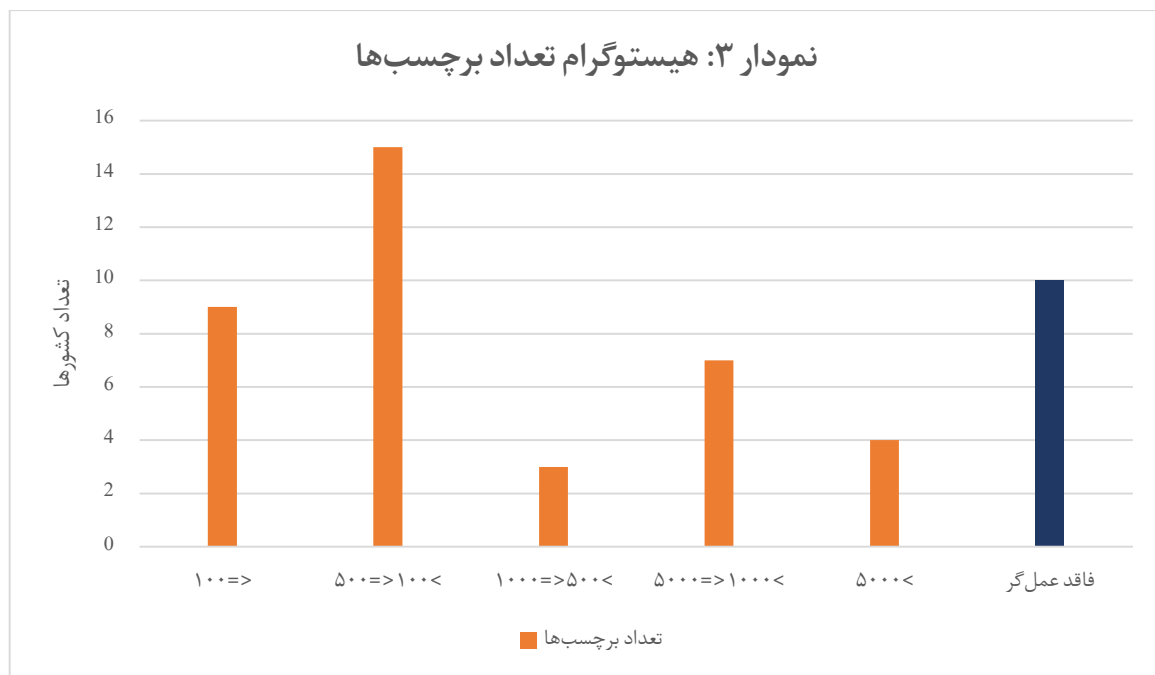
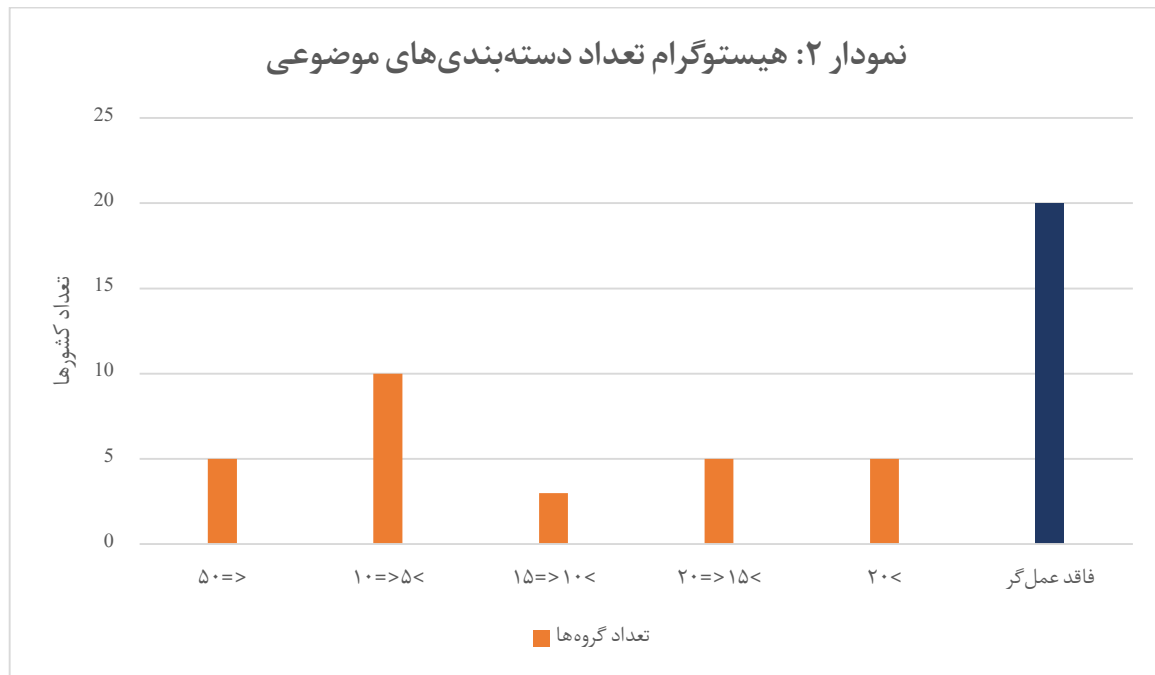
اولین گروه عمل‌گرها متمرکز بر لیست مجموعه داده‌ها، گروه‌ها، سازمان‌ها و سایر موضوعات همچون برچسب‌ها در پرتال بودند. تنها برچسب‌های آزاد (برچسب‌هایی که متعلق به یک مجموعه واژگان نیستند) کنار گذاشته شدند. لیست پروانه نیز پروانه‌های موجود برای مجموعه داده‌های موجود در پرتال‌ها را شامل می‌شد. دومین گروه بر فرآیند جست‌وجو متمرکز بود. جست‌وجو برای بسته‌ها یا منابع، منطبق بر جست‌وجویی بود که ضوابط خاص یک جست‌وجو ایجاد می‌کرد. این کار پارامترهای جست‌وجوی Apache Solr را می‌پذیرد و یک دیکشنری از نتایج به ما ارائه می‌دهد که شامل مجموعه داده‌های مرتبط و مطابق با ضوابط جست‌وجو، دفعات جست‌وجو و نیز اطلاعات چند وجهی می‌شود. تاریخچه فعالیت یعنی تمامی بسته‌های اخیراً در پرتال افزوده شده یا تعویض شده‌اند. دریافت کردن لیستی از حساب کاربران سایت و نقش اعضای گروه‌ها و سازمان‌ها.

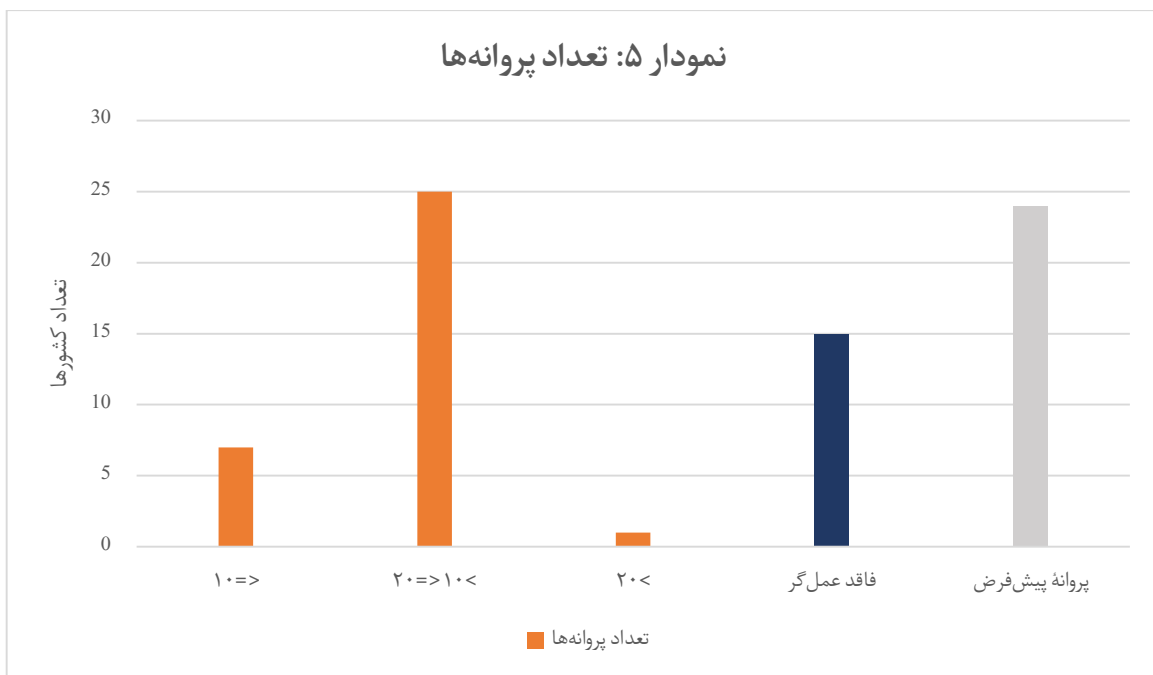
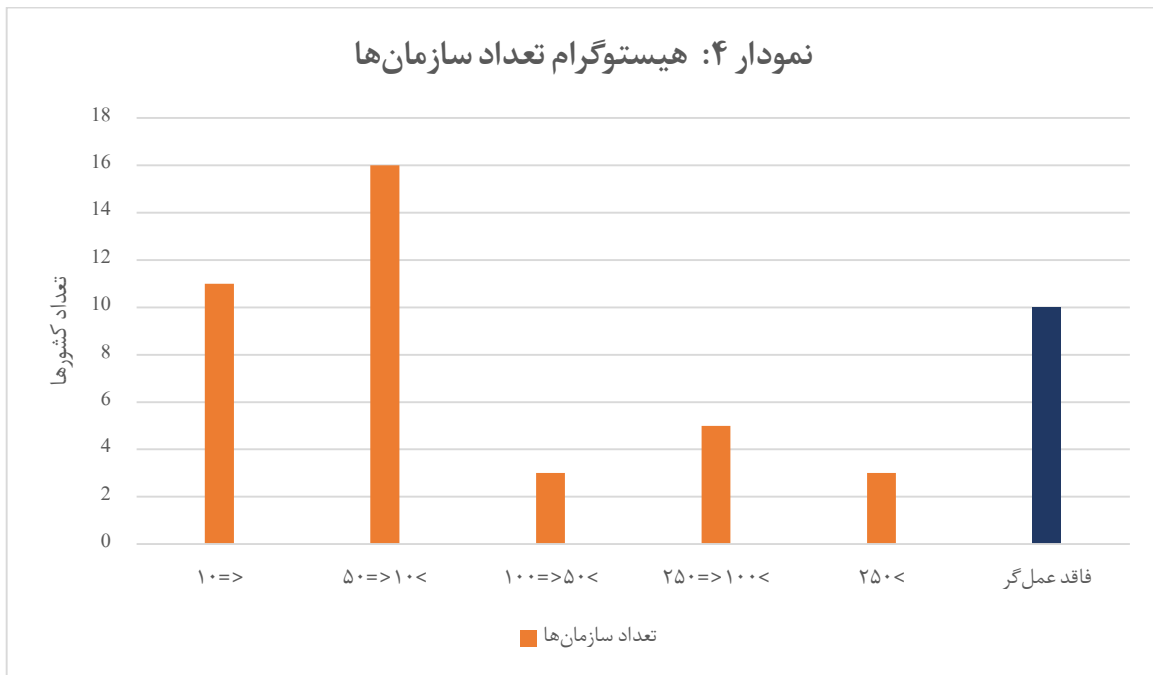
نمودار ۱ هر گروه کشورها را بر مبنای اندازه پرتال داده‌باز نشان می‌دهد. اغلب کشورها بین ۱۰۰ تا ۵۰۰ مجموعه داده دارند. کانادا، ایالات‌متحده و بریتانیا در مقایسه با سایر کشورها داده‌باز بیشتری در اختیار عموم گذاشته‌اند. این نتایج همسو با رتبه‌بندی شاخص جهانی داده‌های باز (GODI) و شاخص بارومتر داده‌باز (ODB) هستند. در رتبه‌بندی شاخص جهانی داده‌های باز (GODI)، بریتانیا، دانمارک و فرانسه بیشترین گشودگی اطلاعات را دارند. در رتبه‌بندی سال ۲۰۱۵ شاخص بارومتر داده‌باز (ODB)، بالاترین میزان گشودگی متعلق به بریتانیا، ایالات‌متحده و سوئد است.

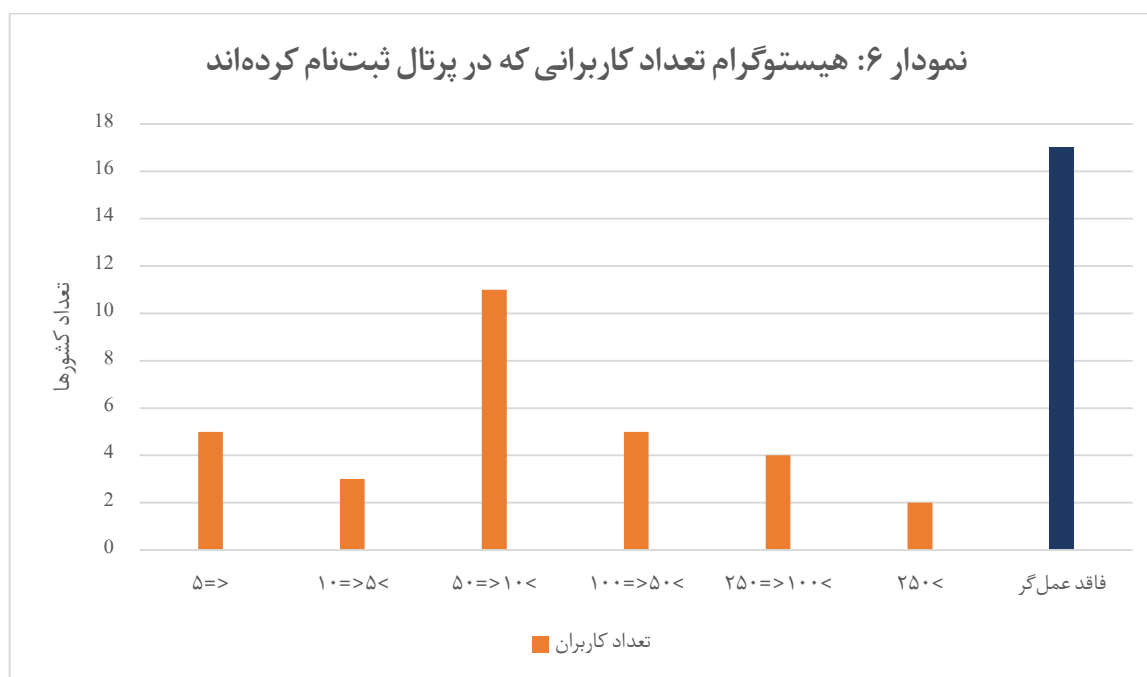
۱ DKAN: یک پورتال داده‌باز مبتنی بر دروپال است و آن هم بر CKAN بنا شده است. م.



در نمودار ۲ تعداد گروه‌های موضوعی پرتال‌ها قابل مشاهده است. این گروه‌های موضوعی می‌تواند برای تقسیم مجموعه داده‌ها به دسته‌های مختلف مورد استفاده قرار گیرد. با این حال، این عملگر الزامی نیست و به همین خاطر ۲۰ پرتال فاقد آن هستند. استرالیا، برزیل، پاراگوئه، رومانی و سوئد بیش از ۲۰ گروه عرضه کرده‌اند. نمودار ۳ تعداد برچسب‌های مرتبط با مجموعه داده‌ها را نشان می‌دهد. نمودار ۴ گروه کشورها را براساس تعداد سازمان‌های آن‌ها که در پرتال مشارکت دارند نشان می‌دهد. معمولاً تا ۵۰ سازمان داده‌های خود را در این پرتال‌ها ارائه کرده‌اند. مجموعه داده‌های بیش از ۲۵۰ سازمان در پرتال داده باز فنلاند، سوئد و بریتانیا در دسترس است. در هر دو نمودار ۱۰ پرتال داده باز این عملگر را نداشته‌اند. تعداد پروانه‌های نشر و استفاده مجدد از داده باز در نمودار ۵ آمده است. مجموعاً ۱۵ پرتال فاقد این عملگر از طریق رابط برنامه‌ویسی کاربردی (API) هستند و ۲۴ پرتال تنها از ۱۵ پروانه پیش فرض استفاده می‌کنند، که توسط شبکه جامع آرشیو اطلاعات (CKAN) عرضه شده است.







پراستفاده ترین پروانه‌ها عبارت‌اند از: پروانه و تخصیص مشترکات داده باز در حوزه عمومی (Open Data Commons)، پروانه پایگاه داده باز مشترکات داده باز (Open Data Commons)، پروانه Public Domain Dedication and License، پروانه اشتراک خلاقانه (Creative Commons CCZero)، و پروانه دولت باز بریتانیا (UK Open Government License)، پروانه مشترکات خلاقانه (Creative Commons Attribution License) و پروانه مستندسازی رایگان GNU، پروانه اسناد همگانی خلاقانه یادکرد (Creative Commons Attribution License) و پروانه مشترکات خلاقانه غیرتجاری (Creative Commons Non-Commercial)، روی هم رفته ۲۶ پروانه در ۲۵ پرتال قابل دسترس است. تنها هشت کشور براساس فضای حقوقی کشور خود اقدام به طراحی پروانه کرده‌اند: استرالیا (سه نوع پروانه)، اتریش، مشترکات خلاقانه (سه نوع پروانه)، کانادا (چهار نوع پروانه)، استونی (دو نوع پروانه)، فنلاند (نُه نوع پروانه)، ایسلند (دو نوع پروانه)، هلند (هشت نوع پروانه)، لهستان (شش نوع پروانه)، رومانی و پاراگوئه هر دو یک نوع پروانه. با این حال، تنها برخی مجموعه داده‌ها با این پروانه‌ها منتشر می‌شوند. این مسئله در مقاله پتی چاکی و همکاران [۴۲] مطرح شده است.

این‌ها همگی در ۲۷ پرتال در دسترس هستند. پروانه اسناد همگانی داده باز (Open Data Commons Attribution License) و پروانه مستندسازی رایگان GNU، پروانه اسناد همگانی خلاقانه یادکرد (Creative Commons Attribution License) و پروانه مشترکات خلاقانه غیرتجاری (Creative Commons Non-Commercial)، روی هم رفته ۲۶ پروانه در ۲۵ پرتال قابل دسترس است. تنها هشت کشور براساس فضای حقوقی کشور خود اقدام به طراحی پروانه کرده‌اند: استرالیا (سه نوع پروانه)، اتریش، مشترکات خلاقانه (سه نوع پروانه)، کانادا (چهار نوع پروانه)، استونی (دو نوع پروانه)، فنلاند (نُه نوع پروانه)، ایسلند (دو نوع پروانه)، هلند (هشت نوع پروانه)، لهستان (شش نوع پروانه)، رومانی و پاراگوئه هر دو یک نوع پروانه. با این حال، تنها برخی مجموعه داده‌ها با این پروانه‌ها منتشر می‌شوند. این مسئله در مقاله پتی چاکی و همکاران [۴۲] مطرح شده است.

همچنین ۳۹ کشور عملگر بسته جست‌وجو از طریق API شان عرضه کرده‌اند. ۳۸ کشور عملگر جست‌وجوی منبع و ۳۸ کشور عملگر بسته‌های اخیراً افزوده یا تعویض شده در پرتال را ارائه می‌دهند. علاوه بر این، عملگر لیست کاربران تنها می‌تواند در اختیار کاربران احراز هویت شده قرار گیرد. با این حال چنان‌که در نمودار ۶ مشاهده می‌شود تنها ۱۷

کشور کاربران را احراز هویت می‌کنند. سایر کشورها اطلاعاتی همچون نام، تاریخ ایجاد حساب، هاش ایمیل^۱، تاریخچه فعالیت اعلان‌های ایمیلی، وضعیت، دفعات ویرایش، شمار بسته‌های مورد استفاده کاربران را در اختیار قرار می‌دهند. دست‌آخر، سه نقش اصلی برای اعضای هر پرتال تعریف شده است که قدرت گرفته از شبکه جامع آرشیو اطلاعات (CKAN) است: مدیر، ویرایشگر و عضو.

۴-۲. ارزیابی کیفیت پرتال‌های داده باز با استفاده از چارچوب محک‌زنی

براساس مرور ادبیاتی که در بخش‌های پیشین انجام شد، سه تحول تکنولوژیکی تأثیری مستقیم بر چارچوب محک‌زنی پیش‌نهادی داشتند. اولین تحول تکنولوژیکی رشد فناوری‌های پهن‌دامنه و سرعت ابزارهای رایانه‌ای است. این‌ها به نهادهای بخش دولتی اجازه می‌دهند که با مجموعه داده‌های بزرگ‌تری کار کنند و خدماتی چون جست‌وجوی شخصی و استفاده کارآمدتر از داده ارائه کنند. دومین تحول تکنولوژیکی استفاده گسترده از رسانه‌های اجتماعی و پلتفرم‌های ویکی برای تولید محتوا، تبادل ایده و بهترین سرمشق‌ها یا به‌اشتراک‌گذاری، بحث و همکاری در مورد اطلاعات است. استفاده از این فناوری‌ها توسط نهادهای بخش دولتی، شهروندان و کسب‌وکارها استفاده از داده دولتی و رابطه میان ذی‌نفعان را متحول ساخته است. آخرین تحول تکنولوژیکی ظهور فناوری‌های داده‌های کلان برای تحلیل، گردآوری و پردازش حجم وسیعی از داده‌های دولتی است که منجر به افزایش استفاده‌های احتمالی از داده‌های دولتی و نشر و اشتراک آن‌ها شده است. این تغییر تأثیر مثبتی بر تلفیق این داده‌ها و افزایش همکاری با استفاده از داده‌های بیشتر و معتبرتر داشته است.

چارچوب محک‌زنی‌ای که در جدول ۱ آمده است مبتنی بر مرور نظام‌مند ادبیات و تجربه نویسندگان و دانش حاصل در بخش نخست ارزیابی پرتال‌های داده است. این چارچوب رویکرد باتیتی و همکاران [۲] در مورد سنجه‌ها و ابعاد کیفیت است. چارچوب پیشنهادی دو بخش دارد. بخش اول بر ویژگی‌های کلی تمرکز دارد که شامل بعد فنی، بعد دسترس‌پذیری و دسترسی و بعد ارتباطات و مشارکت است. بخش دوم ویژگی‌های کلی مجموعه داده‌ها و کیفیت فراداده‌ها را ارزیابی می‌کند. در مجموع، ۲۸ ضابطه (سنجه) کامل تعریف شده است. این چارچوب اولین بار در مقاله لنیکا [۳۰] معرفی شد اما براساس بحث‌ها و بازخوردهای دریافتی اصلاح شد. برخی سنجه‌ها تعریف شدند و مفصل‌تر شرح داده شدند و یک ضابطه جدی به ویژگی‌های کلی مجموعه داده‌ها افزوده شد.

تمامی ضوابط به پرسش تبدیل شدند تا در یک پرسشنامه آمده و به کاربران ارائه شود، هم‌چون: *این پرتال داده باز اطلاعاتی درباره فرد صاحب اختیاری که دارنده این پرتال است و مدل حکمرانی یا چارچوب نهادی پشتیبان مدل‌های ارائه داده در اختیار می‌گذارد.* این پرسش با یک طیف پنج امتیازی لیکرت برای سنجش موافقت یا مخالفت با گزاره پرسیده شدند (۱= کاملاً مخالفم، ۲= مخالفم، ۳= نظری ندارم، ۴= موافقم، ۵= کاملاً موافقم). هر پرسش می‌تواند امتیازی بین یک و پنج بگیرد و امتیاز کل می‌تواند بین ۲۸ و ۱۴۰ باشد.

۱ Email hash: یک نوع عملیات رمزنگاری است. م.

پرسشنامه ارزیابی ابتدا توسط دو استاد دانشگاه که تجربه زیادی در پژوهش کمی راجع به حوزه سامانه اطلاعاتی داشتند آزمون شد. آن‌ها این پرسش‌نامه را قابل فهم تشخیص دادند و مشکلی گزارش نکردند. سپس ۱۰ دانشجوی تحصیلات تکمیلی در مورد قابلیت‌های پرتال‌های داده‌باز و ابعاد کیفی‌شان آموزش دیدند. جلسه آموزشی که در حدود یک ساعت طول کشید متشکل از دو بخش نظری و عملی بود. بخش نظری شامل یک ارائه پاورپوینتی درباره پرتال‌های داده‌باز، الزامات ارزیابی و ضوابط محک‌زنی مورد استفاده برای ارزیابی بود. در پی این ارائه یک جلسه پرسش و پاسخ انجام شد. بخش عملی شامل فرآیندی نمونه برای انجام ارزیابی بود. بعد از پایان این دوره آموزشی، آن‌ها پرسش‌نامه‌ها را به صورت کتبی پر کردند. سپس داده‌های ارزیابی کاربر گردآوری شد و مقدار میانگین برای هر پرتال را به دست آوردیم. نتایج در جدول ۲ ارائه شده‌اند. در سه ستون اول، تمامی ابعاد ویژگی‌های کلی پرتال همراه با مجموع این امتیازات آمده است. امتیاز میانگین برای ویژگی‌های کلی مجموعه داده‌ها را می‌توان در ستون پنجم مشاهده کرد. پس از آن امتیاز کل آمده است.

جدول ۱: چارچوب محک‌زنی برای ارزیابی کیفیت پرتال‌های داده باز

I. ویژگی‌های کلی پرتال داده باز	
لیست معیارها	شرح الزامات ارزیابی کیفیت
۱. بُعد فنی	
۱-۱. اختیار و مسئولیت	پرتال‌ها بایستی اطلاعاتی را درباره فرد صاحب اختیاری که مجری پرتال است و نیز درباره مدل حکمرانی یا چهارچوب نهادی که از مدل تدارک داده‌ها پشتیبانی می‌کند، ارائه دهند. [۳۲]، [۴۴]، [۴۷].
۱-۲. سیستم مدیریت داده	پرتال‌ها بایستی اطلاعاتی را درباره سیستم مدیریت داده‌ها که برای قدرت بخشیدن به پرتال به کار می‌رود، ارائه دهند. [۲۵]، [۳۳]، [۴۴].
۱-۳. زبان	پرتال‌ها بایستی زبان‌های بیشتری را پوشش دهند تا بتوانند کاربران (توجه بیشتری جلب کنند و کیفیت کلی پرتال را تقویت کنند. [۸]، [۴۲].
۱-۴. رایگان بودن	مطابق با پروانه باز، پرتال‌ها بایستی تمامی مجموعه داده‌ها و خدمات را به‌طور رایگان و بدون هیچ محدودیتی در دسترس قرار دهند. [۴۰]، [۵۲].
۲. بُعد دسترس پذیری	
۲-۱. تعداد مجموعه داده‌ها	پرتال‌ها بایستی تعداد مجموعه داده‌هایی را که در اختیار دارند، اعلام کنند. [۲۵]، [۳۳].
۲-۲. تعداد اپلیکیشن‌ها (استفاده مجدد)	پرتال‌ها بایستی تعداد اپلیکیشن‌هایی را که با استفاده مجدد از داده‌های باز ایجاد شده‌اند، اعلام کنند. [۴۷].
۲-۳. موتور جست‌وجو (فیلتر)	پرتال‌ها باید ساختار شفاف و روشنی داشته باشند. همچنین بایستی قابلیت‌های جست‌وجو و ابزارهای انتخاب قوی تعریف کنند. برای این منظور آن‌ها بایستی ضوابط مختلفی برای جست‌وجو براساس دسته‌بندی و با فیلترهای مختلف تعریف کنند [۸]، [۲۷]، [۳۲]، [۳۳]، [۴۲]، [۴۷].
۲-۴. رابط برنامه‌نویسی کاربردی (API)	پرتال‌ها باید برای ذی‌نفعان رابط برنامه‌نویسی کاربردی (API) عرضه کنند تا بتوانند با استفاده از داده‌های باز اپلیکیشن ایجاد کنند [۸]، [۳۳]، [۴۲]، [۴۷].
۲-۵. حساب کاربری	پرتال‌ها باید از امکان ایجاد حساب کاربری پشتیبانی کنند تا صفحات و اطلاعات مشاهده شده قابل شخصی‌سازی باشند [۸].
۲-۶. دسته‌بندی‌های موضوعی	پرتال‌ها باید مجموعه داده‌های‌شان را دسته‌بندی موضوعی کنند [۲۷]، [۴۲].
۲-۷. برچسب‌ها (کلیدواژه‌ها)	دسته‌بندی (موضوعی) بایستی متمایز از برچسب‌ها (کلیدواژه‌ها) باشد [۳۳]. باید برای نشان دادن داده‌های هم‌نوع و متعلق به دسته‌های مشابه از برچسب‌های مشابه استفاده کرد [۸]، [۲۷]، [۳۲].
۳. بُعد ارتباطات و مشارکت	
۳-۱. تالار گفت‌وگو (بازخورد)	پرتال‌ها باید به کاربران امکان ارسال بازخورد در مورد داده‌های‌شان را بدهند. همچنین تالار گفت‌وگو ایجاد کنند تا کاربران بتوانند در آن‌ها بحث کرده و ایده‌های‌شان را به اشتراک بگذارند [۸]، [۴۲]، [۴۷]، [۵۷].



۲-۳. فرم درخواست	پرتال‌ها باید فرم درخواستی ارائه بدهند که بتوان از طریق آن‌ها نوع یا فرمت دیگری از داده‌باز را درخواست داد [۲۷]، [۳۳]، [۵۷].
۳-۳. راهنما (قابلیت استفاده)	پرتال‌ها باید یک قسمت راهنمای سایت باکیفیت داشته باشند تا بتوان به کمک آن به خوبی از پرتال استفاده کرد و قابلیت استفاده از آن افزایش یابد [۸]، [۴۴]، [۴۷].
۳-۴. سؤالات متداول	پرتال‌ها باید یک قسمت سؤالات متداول داشته باشند تا از طریق آن هر نوع مشکل احتمالی حل شود [۴۷].
۳-۵. رسانه اجتماعی	پرتال‌ها باید به پلتفرم رسانه‌های اجتماعی متصل باشند تا از این طریق یک کانال توزیعی برای داده‌های باز شکل بگیرد. کاربران و تأمین‌کنندگان داده‌های باز می‌توانند یکدیگر را از فعالیت‌های‌شان و از آن‌چه به واسطه یک مجموعه داده آموخته‌اند یا با آن انجام داده‌اند، مطلع سازند [۴۲]، [۴۷]، [۵۷].
II. ویژگی‌های کلی مجموعه داده‌ها	
۱. عنوان و شرح	مجموعه داده‌ها باید همراه با یک شرح در مورد چگونگی گردآوری و هدف گردآوری‌شان، ارائه شوند [۸]، [۲۷]، [۴۰].
۲. نشردهنده	بایستی نشردهنده مجموعه داده مشخص باشد تا بتوان اصلت منبع آن را به تأیید رساند [۳۳].
۳. تاریخ نشر و به‌روزرسانی آن	مجموعه داده‌ها باید یک برچسب تاریخ یا دوره زمانی داشته باشند. تمامی اطلاعات موجود در مجموعه داده‌ها بایستی به‌روزرسانی شده باشد [۲۷]، [۳۲]، [۳۳]، [۴۰]، [۵۲].
۴. پروانه	بایستی اطلاعات پروانه استفاده از داده‌های منتشر شده همراه با یک مجموعه داده بیاید. مجموعه داده‌هایی که پروانه باز نداشته باشند، داده باز محسوب نخواهند شد [۳۳]، [۴۰]، [۴۲]، [۵۲].
۵. پوشش جغرافیایی	بایستی مشخص شود که پوشش جغرافیایی یک مجموعه داده در سطح ملی، محلی یا منطقه‌ای است [۳۳]، [۴۲].
۶. نشانی وب (URL) مجموعه داده‌ها	بایستی URL هر مجموعه داده در دسترس باشد [۳۳]، [۴۲].
۷. اندازه مجموعه داده‌ها	حجم (فایل) هر مجموعه داده باید مشخص شده باشد [۲۵]، [۳۳].
۸. تعداد دفعات مشاهده داده‌ها	تعداد دفعات مشاهده هر مجموعه داده باید در دسترس باشد [۴۷].
۹. تعداد دفعات بارگیری داده‌ها	تعداد دفعات بارگیری هر مجموعه داده باید در دسترس باشد [۴۷].
۱۰. فرمت خوانا برای رایانه	مجموعه داده‌ها بایستی در فرمت‌های ارائه شوند که تحلیل و ویرایش‌شان همچون قالب‌های شناخته شده قابل بارگیری آسان باشد [۸]، [۲۷]، [۳۳]، [۴۰]، [۴۷]، [۵۲].
۱۱. بصری‌سازی	قابلیت بصری‌سازی برای داده‌ها باید تعبیه شده باشد. (برای مثال بصری‌سازی در قالب نمودار یا نقشه) [۴۲].
۱۲. نمره‌بندی از جانب کاربران و پیام‌های گفت‌وگو	بایستی قابلیت گردآوری نمره‌دهی و نظرات کاربران در مورد هر مجموعه داده وجود داشته باشد یا بتوان در مورد نتایج استفاده از داده بحث کرد [۸]، [۵۷].

جدول ۲: نتایج ارزیابی کیفیت پرتال‌های داده باز

کشور	پورتال داده باز	I.				نمره کلی
		۱	۲	۳	جمع	
بریتانیا	سایت ۴	۱۹	۳۴	۲۳	۷۶	۱۲۹,۶
هند	سایت ۵	۱۹,۵	۲۹,۸	۲۱,۴	۷۰,۷	۱۲۶
ایالات متحده	سایت ۶	۱۸,۵	۳۴	۲۰,۱	۷۲,۶	۱۲۴,۶
استرالیا	سایت ۷	۱۸,۲	۳۳,۱	۱۹,۸	۷۱,۱	۱۲۰,۴
اتریش	سایت ۸	۱۶,۷	۳۱,۵	۲۰,۶	۶۸,۸	۱۱۴,۲
کانادا	سایت ۹	۱۷,۹	۲۹,۴	۲۱	۶۸,۳	۱۱۱,۳
فرانسه	سایت ۱۰	۱۶,۵	۲۹,۹	۱۵,۳	۶۱,۷	۱۱۰,۷
روسیه	سایت ۱۱	۱۲,۴	۲۸,۲	۲۱,۳	۶۱,۹	۱۱۰,۶
مولداوی	سایت ۱۲	۱۷,۶	۲۹,۴	۱۵,۵	۶۲,۵	۱۰۹,۴
کرواسی	سایت ۱۳	۱۵,۵	۳۰,۸	۲۰,۶	۶۶,۹	۱۰۷,۲
آلمان	سایت ۱۴	۱۵,۳	۳۳	۱۷,۲	۶۵,۵	۱۰۶,۶
پاراگوئه	سایت ۱۵	۱۷,۵	۳۲,۲	۱۹,۱	۶۸,۸	۱۰۵,۸
یونان	سایت ۱۶	۱۵	۲۹,۳	۱۶,۲	۶۰,۵	۱۰۴,۲
فنلاند	سایت ۱۷	۱۵,۷	۲۶	۱۱,۱	۵۲,۸	۱۰۴,۲
شیلی	سایت ۱۸	۱۶,۱	۲۷,۵	۱۱,۵	۵۵,۱	۱۰۳,۹
برزیل	سایت ۱۹	۱۵,۳	۲۸,۴	۱۷,۲	۶۰,۹	۱۰۲,۵
سویس	سایت ۲۰	۱۸,۵	۲۹,۲	۱۵,۳	۶۳	۱۰۲,۴
نیوزیلند	سایت ۲۱	۱۵,۱	۲۸,۹	۱۸,۶	۶۲,۶	۱۰۲,۳
ایرلند	سایت ۲۲	۱۸,۱	۲۸,۲	۱۵,۹	۶۲,۲	۱۰۲,۱
نپال	سایت ۲۳	۱۵,۳	۲۴,۴	۱۷	۵۶,۷	۱۰۰,۲
تایوان	سایت ۲۴	۱۲	۲۵,۵	۱۹,۴	۵۶,۹	۹۹,۹
پرتغال	سایت ۲۵	۱۴,۶	۲۶,۴	۱۷,۴	۵۸,۴	۹۹,۶
تانزانیا	سایت ۲۶	۱۵,۴	۲۷,۴	۱۷,۸	۶۰,۶	۹۸,۸
ژاپن	سایت ۲۷	۱۶,۳	۲۶,۶	۱۰,۴	۵۳,۳	۹۸,۷
بلاروس	سایت ۲۸	۱۱,۶	۳۱,۲	۱۵,۴	۵۸,۲	۹۸,۴
بلغارستان	سایت ۲۹	۱۸,۷	۲۳,۸	۱۶,۸	۵۹,۳	۹۷,۹
کره جنوبی	سایت ۳۰	۱۵,۲	۲۶,۹	۱۷,۲	۵۹,۳	۹۷,۸
اروگوئه	سایت ۳۱	۱۵,۸	۳۲,۸	۱۰,۴	۵۹	۹۷,۷
سوئد	سایت ۳۲	۱۸,۳	۲۶,۵	۱۲,۹	۵۷,۷	۹۷,۶
اسپانیا	سایت ۳۳	۱۲,۸	۲۸,۲	۱۷,۲	۵۸,۲	۹۷,۳
ایتالیا	سایت ۳۴	۱۵,۱	۲۲,۷	۱۹,۱	۵۶,۹	۹۶,۵
کلمبیا	سایت ۳۵	۱۱,۴	۲۳,۷	۱۸,۹	۵۴	۹۶,۴
اندونزی	سایت ۳۶	۱۴,۳	۲۷,۶	۱۵,۴	۵۷,۳	۹۵,۷
سنگاپور	سایت ۳۷	۱۳,۱	۲۹,۲	۱۱,۷	۵۴	۹۴,۶

کشور	پورتال داده باز	I.				II.	نمره کلی
		۱	۲	۳	جمع		
اسلواکی	سایت ۳۸	۱۸,۵	۲۶	۱۲	۵۶,۵	۳۷,۹	۹۴,۴
السالوادور	سایت ۳۹	۱۳,۴	۲۳,۱	۲۰,۵	۵۷	۳۷,۱	۹۴,۱
رومانی	سایت ۴۰	۱۸,۲	۲۶,۱	۱۰,۲	۵۴,۵	۳۹,۲	۹۳,۷
مراکش	سایت ۴۱	۱۷,۹	۳۰	۱۰,۹	۵۸,۸	۳۳,۹	۹۲,۷
هلند	سایت ۴۲	۱۳,۳	۲۷,۵	۱۰,۵	۵۱,۳	۴۱,۴	۹۲,۷
فیلیپین	سایت ۴۳	۱۵,۴	۲۸,۹	۱۳	۵۷,۳	۳۵,۲	۹۲,۵
غنا	سایت ۴۴	۱۵,۳	۲۲,۵	۱۸,۴	۵۶,۲	۳۶,۲	۹۲,۴
دانمارک	سایت ۴۵	۱۴,۳	۲۴,۹	۱۸	۵۷,۲	۳۵	۹۲,۲
پاکستان	سایت ۴۶	۱۷,۱	۲۸,۶	۱۰,۶	۵۶,۳	۳۴,۹	۹۱,۲
نروژ	سایت ۴۷	۱۳,۲	۳۲	۱۲,۴	۵۷,۶	۳۳,۱	۹۰,۷
لهستان	سایت ۴۸	۱۶,۱	۲۶,۳	۱۰,۵	۵۲,۹	۳۷,۷	۹۰,۶
کنیا	سایت ۴۹	۱۶,۸	۲۳,۹	۲۰,۷	۶۱,۴	۲۸,۸	۹۰,۲
سريلانكا	سایت ۵۰	۱۵,۲	۲۲,۶	۱۹,۳	۵۷,۱	۳۲,۹	۹۰
مقدونیه	سایت ۵۱	۱۳,۲	۲۳,۱	۱۳,۵	۴۹,۸	۳۸,۵	۸۸,۳
بلژیک	سایت ۵۲	۱۴,۵	۱۹,۹	۱۵,۷	۵۰,۱	۳۸,۱	۸۸,۲
تایلند	سایت ۵۳	۱۰,۷	۲۱,۵	۱۹,۵	۵۱,۷	۳۶,۵	۸۸,۲
جمهوری چک	سایت ۵۴	۱۶,۴	۲۶,۵	۶,۴	۴۹,۳	۳۸,۱	۸۷,۴
ازبکستان	سایت ۵۵	۱۵,۷	۲۵,۴	۱۲,۸	۵۳,۹	۳۳,۲	۸۷,۱
قزاقستان	سایت ۵۶	۱۷,۶	۲۵,۱	۱۰,۵	۵۳,۲	۳۳,۵	۸۶,۷
قبرس	سایت ۵۷	۱۴,۵	۱۸,۷	۱۷,۴	۵۰,۶	۳۵,۴	۸۶
صربستان	سایت ۵۸	۱۶,۶	۲۵	۶,۴	۴۸	۳۷,۵	۸۵,۵
اوکراین	سایت ۵۹	۱۳,۳	۲۵,۷	۱۴,۹	۵۳,۹	۳۱,۲	۸۵,۱
کاستاریکا	سایت ۶۰	۱۳,۷	۲۳,۱	۱۳,۴	۵۰,۲	۳۴,۷	۸۴,۹
اوگاندا	سایت ۶۱	۱۴,۵	۲۲,۴	۱۲,۶	۴۹,۵	۳۴,۷	۸۴,۲
عربستان سعودی	سایت ۶۲	۱۵,۹	۱۹,۲	۱۵,۱	۵۰,۲	۳۳,۸	۸۴
هنگ کنگ	سایت ۶۳	۱۶,۸	۱۸,۲	۱۵,۱	۵۰,۱	۲۸,۹	۷۹
مکزیک	سایت ۶۴	۱۱,۷	۲۲,۴	۱۳,۳	۴۷,۴	۳۰,۲	۷۷,۶
برونئی	سایت ۶۵	۱۲,۶	۲۱,۸	۱۵,۴	۴۹,۸	۲۶,۶	۷۶,۴
مالزی	سایت ۶۶	۱۳	۱۹,۷	۱۰,۶	۴۳,۳	۳۰,۵	۷۳,۸
لیتوانی	سایت ۶۷	۸,۱	۲۶,۲	۱۰,۳	۴۴,۶	۲۷,۵	۷۲,۱
بورکینافاسو	سایت ۶۸	۱۳,۵	۲۱,۵	۱۰,۴	۴۵,۵	۲۶,۵	۷۱,۹
تونس	سایت ۶۹	۱۲,۴	۱۵	۱۹	۴۶,۴	۲۳,۳	۶۹,۷
فلسطین اشغالی	سایت ۷۰	۰,۱	۱۵,۱	۱۳,۴	۳۸,۶	۳۰,۱	۶۸,۷

۵. یافته‌ها و بحث

بریتانیا در میان کشورهای ارزیابی شده بالاترین امتیاز را به دست آورد که همسو با شاخص جهانی داده‌های باز (GODI) و شاخص بارومتر داده‌های باز (ODB) است. این شاخص‌ها و گزارش‌های مربوط به آن‌ها بریتانیا را دارای پیشرفته‌ترین اکوسیستم داده باز می‌دانند. این نکته با یافته‌های هم‌اشتات و همکاران که در [۱۸] آمده است نیز همسو است. بعد از بریتانیا، هند قرار دارد که با کیفیت‌ترین مجموعه داده‌ها را ارائه می‌کند. پس از آن ایالات متحده و استرالیا هستند. غالب پرتال‌های ارزیابی شده قدرت گرفته از شبکه جامع آرشیو اطلاعات (CKAN) هستند، اگرچه که همگی از آخرین نسخه آن استفاده نمی‌کنند. از این‌رو، امتیاز پایین‌تری گرفتند. نمونه این کشورها جمهوری چک است. با این حال، از سوی دیگر، بهترین نتایج نصیب کشورهای شد که پرتال داده باز آن‌ها قدرت گرفته از CKAN و DKAN بود. ب‌عبارت دیگر از توسعه یافته‌ترین اقتصادها مانند فرانسه، آلمان و ایالات متحده.

اساساً پرتال‌های داده بازی که بهترین عملکرد را داشتند مانند پرتال‌های بریتانیا، هند، ایالات متحده، استرالیا و اتریش عملگرها گوناگونی در آن‌ها تعریف شده بود. عملگرهایی مانند جست‌وجو، درخواست، مشاهده، بارگیری، تحلیل، تلفیق، بصری‌سازی و مباحثه درباره انواع مجموعه داده‌ها و نیز روندها و مفاهیم در مورد داده‌های باز. برای مثال، یک کاربر می‌تواند در منوی اصلی زبان را تغییر دهد، داده جست‌وجو کند، سپس نتایج را فیلتر کند و بر روی یکی از مجموعه داده‌ها زوم کرده و آن را به دقت تحلیل کند. اغلب این پرتال‌ها امکان استفاده شخصی‌سازی شده از عملگرها موجود را با استفاده از ایجاد حساب کاربری در اختیار می‌گذارند. ارزیابی هم‌چنین نشان داد که برخی پرتال‌ها امکان بالایی برای مداخله فراهم می‌کنند و امکاناتی دارند از قبیل: وبلاگ برای تسهیل به اشتراک‌گذاری منابع و اطلاعات، تالار گفت‌وگو برای تسهیل بحث درباره اپلیکیشن‌ها و فناوری‌های مرتبط و گروه‌های مجازی. در مجموع، برخی از این پرتال‌های داده باز یک نقطه دسترسی واحد برای سیاست‌ها و برنامه‌های دولت می‌سازند و داده‌های دولتی و فناوری‌های تازه را ارائه می‌دهند، مانند استفاده کنونی دولت از پلتفرم‌های رسانه‌های اجتماعی. گذشته از این‌ها، اگرچه تمامی پرتال‌ها نوعی فراداده برای هر مجموعه داده ارائه می‌کنند، سودمندی این‌ها اغلب محدود است چراکه تمامی بخش‌های ممکن فراداده تکمیل نشده است، برخی بخش‌های فراداده خالی می‌ماند یا یافتن آن‌ها در پرتال دشوار است.

براساس ارزیابی‌ای که در بخش‌های پیش ارائه شد، می‌توان نتیجه گرفت که اغلب پرتال‌های داده باز در جهان گام‌های مهمی به سمت عمومی ساختن تعدادی مجموعه داده جالب و مفید برداشته‌اند. بدین گونه، یافته‌های ارائه شده در این مقاله همسو با یافته‌های مقاله الکسیلوس و همکاران [۱] است که گزارش می‌دهد که اولین نسل پرتال‌های داده باز تنها عملگرهای ساده برای جست‌وجو و بارگیری داده توسط کاربران و نیز بارگذاری داده توسط تأمین‌کنندگان داشتند. غالب این پرتال‌ها امکان جست‌وجوی ساده متنی و نیز عمل‌گرها جست‌وجوی موضوعی برای یافتن مجموعه داده‌ها را می‌دهند. تنها برخی از بهترین پرتال‌ها اخیراً از وب معناسناختی بهره گرفته‌اند. با وب معناسناختی می‌توان از خدمات جست‌وجوی مبتنی بر معنا بهره برد و تنها شمار کمی از پرتال‌ها امکان مشاهده مجموعه داده‌ها بر روی نقشه را می‌دهند. از جمله نمره‌دهی مجموعه داده و ثبت نظر یا انواع نمودارها. با این حال هیچ عملگری برای پردازش

مجموعه داده‌ها برای بهبود آن‌ها، تغییر آن‌ها برای نیازهای خاص یا پیوندزنی آن‌ها با سایر مجموعه داده‌های (عمومی یا خصوصی) و برای بارگذاری-نشر نسخه‌های جدید آن‌ها یا بارگذاری مجموعه داده‌های خود کاربران وجود ندارد. بنابراین، این مقاله نشان می‌دهد که اصلاحات مهمی باید انجام شود تا باز بودن اطلاعات در برخی کشورهای بهبود یابد و ارزش اجتماعی و اقتصادی حاصل از آن افزایش یابد. توصیه‌هایی در این زمینه در پایین می‌آید.

چنان‌که یافته‌های ارزیابی اطلاعات نشان داد، کیفیت پرتال‌ها تحت تأثیر نسخه سامانه مدیریت داده است، به‌ویژه در مورد شبکه جامع آرشیو اطلاعات (CKAN). پرتال‌های قدرت‌گرفته از CKAN و DKAN امتیاز بالاتری گرفتند. CKAN نسخه‌های متفاوتی دارد که از حیث ویژگی‌ها و سطوح خدمات با یکدیگر تفاوت دارند. هر نسخه عملگرهای مختلفی دارد که می‌تواند کیفیت پرتال و نیز کیفیت مجموعه داده و فراداده‌های مربوطه را ارتقا دهد. بدین‌سان، می‌توان توصیه کرد که سایر پرتال‌ها باید خدمات خود را به CKAN یا DKAN منتقل کنند. این مسئله به‌ویژه در مورد کشورهای در حال توسعه مصداق دارد.

همچنین، مجموعه داده‌های بیشتری، با تنوع موضوعی بالاتر از جمله داده‌های اقتصادی / مالی درباره نحوه هزینه‌کرد دولت بایستی ارائه شود که می‌تواند به افزایش شفافیت و پاسخ‌گویی منجر شود. همچنین این موضوع در مورد فعالیت‌های اقتصادی کسب‌وکارها مصداق دارد. در عین حال، باید تأکید بیشتری بر ارائه مجموعه داده‌هایی شود که برای کیفیت زندگی شهروندان اهمیت دارند. مجموعه داده‌هایی درباره اشتغال، کشاورزی و گردشگری، محیط‌زیست و برنامه‌ریزی. می‌توان تأکید بیشتری بر روی استفاده از فرمت‌های ساختارمند و خوانا برای رایانه در انتشار مجموعه داده‌ها و فراداده‌ها کرد (بکارگیری استانداردهای فراداده‌ای موجود) و نیز بر روی پشتیبانی از چارچوب توصیف منبع (RDF) و اسپارکل (SPARQL) که جست‌وجو و یافتن راحت‌تر مجموعه داده‌ها و نیز پیوندزنی و تلفیق داده‌های باز از چندین منبع را میسر می‌سازند. این‌ها به افزایش سودمندی و ارزش پرتال‌ها برای ذی‌نفعان مختلف می‌انجامد. وانگهی، باید تعامل با تأمین‌کنندگان داده ممکن باشد؛ برای نمونه کاربران بتوانند از تأمین‌کنندگان داده درخواست داده کنند و پس از استفاده از این داده به تأمین‌کنندگان بازخورد دهند. اطلاعات درباره پروانه‌های استفاده از مجموعه داده‌ها در اکوسیستم‌های داده باز اهمیت دارد، چراکه کاربران داده باز نیاز دارند بدانند که آیا این پروانه به آن‌ها استفاده اجازه از داده به‌نحو مورد نظرشان را می‌دهد یا نه. همچنین اطلاعات مربوط به کاربران ثبت‌نام کرده بایستی تنها در اختیار کاربر احراز هویت شده قرار بگیرد. این عملگر شاید از طریق رابط برنامه‌نویسی کاربردی (API) در دسترس نباشد. دست‌آخر، تعامل‌پذیری میان پرتال‌های ارزیابی شده را می‌توان با ارائه فراداده‌هایی که شناسه‌ها و واژه‌های مشترک دارند ارتقا داد. این فقدان هم‌آهنگی ضرورت ایجاد استانداردهای کیفیت را شدت می‌بخشد.

افراد صاحب اختیاری که مسئول این پرتال‌های داده باز هستند باید به نیازهای خاص کسانی که در پی داده‌های مربوط به پاسخ‌گویی هستند، توجه کنند و ساختارها و سازوکارهای مناسب برای این منظور را تعبیه کنند [۳۲]. در عین حال، عملکرد منابع موجود داده باز دولتی (OGD) باید بهبود یابد: ابزارهای جست‌وجوی داده ارتقا یابند (به‌نحوی که کاربران احتمالی سریع‌تر و راحت‌تر مجموعه داده مورد نیاز خود را بیابند)، بصری‌سازی داده (مانند نقشه و نمودار؛ تا کاربر احتمالی بتواند فهم اولیه سریع و راحتی از مجموعه داده کسب کند و تصمیم بگیرد که تحلیل

دقیق‌تر به کارش می‌آید یا نه)، و بازخوردهای کاربران (به‌نحوی که کاربران داده باز دولتی بتوانند به تأمین‌کنندگان این داده‌ها بازخورد دهند) - در مورد کیفیت مجموعه داده‌ای که استفاده کرده‌اند، نقاط ضعف و اصلاحات لازم، نیاز به مجموعه داده‌های دیگر. چراکه همکاری بین کاربران و تأمین‌کنندگان داده برای ایجاد ارزش از آن‌ها ضروری تلقی شده است [۲۱]، [۲۲]، [۳۲]، یا [۵۶]، [۵۸].

برخلاف پژوهش‌های [۳۰]، [۳۱]، [۳۳]، [۴۲]، [۴۴]، [۴۸]، [۵۱] این مقاله بین کیفیت یک پرتال و کیفیت مجموعه داده‌ها و فراداده‌های این پرتال تمایز قائل می‌شود. وانگهی، از نمونه پژوهشی بزرگتری بهره می‌برد. نهایتاً یافته‌های مهمی درباره بهبود کیفیت این پرتال‌ها، به‌ویژه در عرصه امنیت و حفاظت از اطلاعات دارد.

محدودیت‌ها و اعتبار نتایج این تحقیق شاید تحت‌تأثیر برنامه‌ها، قوانین و مقررات مرتبطی باشد که اجازه استفاده از داده جز در موارد تنظیم شده یا برای منظوری که داده برای آن تهیه شده بود را ندهند. مشخصاً، تنها داده‌هایی که مطابق با این شرایط باشد در پرتال‌های داده باز مورد بررسی یافت شود. باین‌حال، در چارچوب پیشنهادی این مسئله لحاظ شده و غالب ارزیابی انجام شده در مورد مجموعه داده‌ها براساس شرح فراداده بوده است. همچنین این چارچوب محک‌زنی عملگری ارزیابی نکرده است که نیازمند کاربر احراز هویت شده باشد. گذشته از این‌ها، پرتال‌های ارزیابی شده ساختارهای سامان‌دهی و دسته‌بندی پیچیده (و بعضاً مبهمی) که مرتب تغییر می‌کنند و اجازه تحلیل از منظر یک کاربر عادی را نمی‌دهند (که دانش داخلی از هر پرتال ندارد). تمامی ضوابط در چارچوب محک‌زنی دارای روایی مشابه هستند. باین‌حال، احتمالاً بعضی از آن‌ها اهمیتی بیشتر در مقایسه با باقی داشته باشند. اگرچه این شرایط در ادبیات اولویت‌بندی نشده‌اند، می‌توانند بر رده‌بندی نهایی اثرگذار باشند. برای مثال، اگر مشخصه‌های کلی پرتال داده باز وزن ۲/۳ داشته باشد و مشخصه‌های کلی مجموعه داده وزن ۱/۳، رده‌بندی نهایی بدین‌صورت است: بریتانیا، ایالات متحده، هند، استرالیا، اتریش، کانادا، پاراگوئه، کرواسی، روسیه و فرانسه.

نکته مهم دیگر کاربرانی است که از آن‌ها خواسته شود پرتال‌های مورد نظر را ارزیابی کنند. در این مورد، تنها دانشجویان تحصیلات تکمیلی مشارکت کردند. اگرچه ذی‌نفعان دیگری مانند دولت‌های محلی و ایالتی، کسب‌وکارها و شهروندان نیز وجود دارند. بدین‌سان، پژوهش‌های آتی باید تمرکز بیشتری بر مطالعه آن‌ها، شرایطشان و دلیلی که برای نظرشان در مورد ارزش این پرتال‌ها دارند باید بررسی شود. همچنین، پرتال‌های انتخاب‌شده در سطح منطقه‌ای یا محلی بخش دولتی باید ارزیابی شوند. دست‌آخر، ضرورت وجود استانداردهایی در مورد کیفیت داده دوباره مطرح می‌شود.

۶. نتیجه‌گیری

این مقاله چارچوب محک‌زنی برای ارزیابی کیفیت پرتال‌های داده باز در سطح ملی ارائه کرد. همچنین مرور مفصلی بر روی مسائل، چالش‌ها و فرصت‌های این پرتال‌ها صورت گرفت. برای این منظور مرور نظام‌مند ادبیات و روش‌های تحلیل محتوا و چندبعدی با تکنیک‌های کمی بکار گرفته شد.



پرتال‌های داده‌باز واسطه بین داده دولتی از یک‌سو و کاربران از سوی دیگر است. اساساً هر پرتال باید ظاهری مرتب با یک نوار جست‌وجو در صفحه اصلی و اطلاعاتی درباره افراد صاحب اختیار پرتال داشته باشد. همچنین محتوای آن باید در قالب دسته‌ها و برجسته‌های ساده و شفاف ارائه شود. به غیر از قراردادن داده‌ها در اختیار ذی‌نفعان، این پرتال‌ها باید ایده‌ها و بازخوردهای شهروندان را دخیل کنند. توسعه کارآمد این پرتال‌ها ارزیابی نظام‌مند کیفیت را برای شناخت بهتر آن‌ها و ارزیابی انواع ارزشی که خلق می‌کنند، ضروری می‌سازد. بنابراین، چارچوب پیشنهادی عمده‌ترین ایفای سهم این مقاله است. باین‌حال، ارزیابی کیفیت پرتال‌های داده‌باز نه تنها برای درک ارزش پرتال‌ها مهم است، بلکه برای پیشرفت‌های بیشتر نیز مهم می‌باشد. یافته‌های سایر ارزیابی‌ها را می‌توان برای ارتقای مجموعه داده‌های باز و پرتال‌های داده‌های باز مورد استفاده قرار داد تا آن‌ها کاربردی‌تر شوند و استفاده از آن‌ها افزایش یابد. دست‌آخر، برخی مشاهدات و توصیه‌ها از حیث عملی ارتقای پرتال‌های داده‌باز از این ارزیابی استنتاج شده است.

باین‌حال، شماره مجموعه داده‌های آنلاین و پیچیدگی پرتال‌های داده‌های باز و عملگرهای آن‌ها تفاوت دارد و نشان از فقدان یک‌دستی دارد. ایجاد استانداردهای کیفی ضروری است. به‌طور خاص بریتانیا، هند و ایالات‌متحده مجموعه داده‌های زیادی منتشر کرده و وب‌سایت‌های پیشرفته‌ای راه‌اندازی کرده‌اند. مشخص است که کیفیت داده و تنظیمات زبانی بسیار مهم هستند اما مهم‌ترین مسئله ارائه داده به‌شکلی قانونی با استانداردهای کیفی و بروز برای کسانی است که حق استفاده از آن در قالب ابزار، اطلاعات یا فراداده را دارند. بنابراین، نتایج این ارزیابی کیفی با استفاده از چارچوب جدید محک‌زنی می‌تواند برای بهبود کیفیت پرتال‌های داده‌های باز و نیز زیرساخت، استراتژی‌ها و ابتکار عمل‌ها در زمینه داده مورد استفاده قرار گیرد.

منابع

- [1] C. Alexopoulos, et al., Designing a second generation of open data platforms: Integrating open data and social media, in Proceedings of the 13th IFIP WG 8.5 International Conference (EGOV 2014), 2014, pp. 230-241. [[Links](#)]
- [2] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, Methodologies for data quality assessment and improvement, ACM Computing Surveys, vol. 41, no. 3, pp. 1-52, 2009. [[Links](#)]
- [3] R. Bílková, R. Máchová and M. Lnénicka, M., Evaluating the impact of open data using partial least squares structural equation modeling, Scientific Papers of the University of Pardubice - Series D, Faculty of Economics and Administration, vol. 22, no. 34, pp. 29-41, 2015. [[Links](#)]
- [4] K. Braunschweig, J. Eberius, M. Thiele, and W. Lehner, The state of open data - limits of current open data platforms, in Proceedings of the 21st World Wide Web Conference 2012, Web Science Track at WWW'12, ACM, New York, 2012, pp. 1-6. [[Links](#)]
- [5] S. Buchholtz, M. Bukowski and A. Sniegocki, Big and Open Data in Europe: A Growth Engine or a Missed Opportunity. Varsava: demosEUROPA, 2014. [[Links](#)]
- [6] I. Caballero, M. Serrano and M. Piattini, A data quality in use model for big data, in Advances in Conceptual Modeling (M. Indulska and S. Purao, Eds.). Atlanta, GA: Springer, pp. 65-74, 2014. [[Links](#)]
- [7] C. Calero, A. Caro and M. Piattini, An applicable data quality model for web portal data consumers, world wide web, vol. 11, no. 4, pp. 465-484, 2008. [[Links](#)]
- [8] Y. Charalabidis, E. Loukis and C. Alexopoulos, Evaluating second generation open government data infrastructures using value models, in Proceedings of the 47th Hawaii International Conference on System Sciences, IEEE, Hilton Waikoloa, 2014, pp. 2114-2126. [[Links](#)]
- [9] M. Chen, S. Mao and Y. Liu, Big data: A survey, Mobile Networks and Applications, vol. 19, no. 2, pp. 171-209, 2014. [[Links](#)]
- [10] P. Colpaert, S. Joye, P. Mechant, E. Mannens, and R. van de Walle, The 5 stars of open data portals, in Proceedings of the 7th International Conference on Methodologies, Technologies and Tools Enabling E-Government (MeTTeG13), University of Vigo, Spain, 2013, pp. 61-67. [[Links](#)]
- [11] D. Cowan, P. Alencar and F. McGarry, Perspectives on open data: Issues and opportunities, in Proceedings of the 2014 IEEE International Conference on Software Science, Technology and Engineering, IEEE, Ramat Gal, Israel, 2014, pp. 24-33. [[Links](#)]
- [12] R. Cyganiak and F. Maali. (2015, May) Use cases and requirements for the data catalog vocabulary. W3C. [Online]. Available: <https://dvcs.w3.org/hg/gld/raw-file/default/dcat-ucr/index.html> [[Links](#)]
- [13] S. S. Dawes and N. Helbig, Information strategies for open government: Challenges and prospects for deriving public value from government transparency, in Proceedings of the 9th IFIP WG 8.5 International Conference (EGOV 2010), Springer, Heidelberg, 2010, pp. 50-60. [[Links](#)]
- [14] I. Ermilov, et al., Linked open data statistics: Collection and exploitation, in Knowledge Engineering and the Semantic Web (P. Klinov and D. Mouromtsev, Eds.). Berlin Heidelberg: Springer, 2013, pp. 242-249. [[Links](#)]
- [15] C. P. Geiger and J. von Lucke, Open government and (linked) (open) (government) (data), eJournal of eDemocracy & Open Government, vol. 4, no. 2, pp. 265-278, 2012. [[Links](#)]

- [16] H. S. Hansen, L. Hvingel and L. Schroder, Open government data - A key element in the digital society, in *Technology-Enabled Innovation for Democracy, Government and Governance* (A. Ko, C. Leitner, H. Leitold, and A. Prosser, Eds.). Berlin: Springer, 2013, pp. 167-180. [[Links](#)]
- [17] T. Heath and C. Bizer, Linked data: Evolving the web into a global data space, *Synthesis Lectures on the Semantic Web*, vol. 1, no. 1, pp. 1-136, 2011. [[Links](#)]
- [18] M. Heimstadt, F. Saunderson and T. Heath, Conceptualizing open data ecosystems: A timeline analysis of open data development in the UK, in *CeDEM14: Proceedings of the International Conference for E-Democracy and Open Government 2014*, Edition Donau-Universitat Krems, Krems, 2014, pp. 245-255. [[Links](#)]
- [19] B. Hyland and D. Wood, The joy of data-a cookbook for publishing linked government data on the web, in *Linking Government Data* (D. Wood, Ed.). New York: Springer, 2011, pp. 3-26. [[Links](#)]
- [20] M. Janssen, Y. Charalabidis and A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, *Information Systems Management*, vol. 29, no. 4, pp. 258-268, 2012. [[Links](#)]
- [21] T. Jetzek, M. Avital and N. Bjorn-Andersen, Generating value from open government data, in *Proceedings of 34th International Conference on Information Systems: ICIS 2013*, Bepress, Berkeley, 2013, pp. 1-20. [[Links](#)]
- [22] T. Jetzek, M. Avital and N. Bjorn-Andersen, Generating sustainable value from open data in a sharing society, in *Proceedings of IFIP WG 8.6 International Conference on Transfer and Diffusion of IT (TDIT 2014)*, Berlin, Springer, 2014, pp. 62-82. [[Links](#)]
- [23] E. Kalampokis, E. Tambouris and K. Tarabanis, Linked open government data analytics, in *Proceedings of the 12th IFIP WG 8.5 International Conference (EGOV 2013)*, Springer, Heidelberg, 2013, pp. 99-110. [[Links](#)]
- [24] E. Kalampokis, E. Tambouris and K. Tarabanis, Open government data: A stage model, in *Proceedings of the 10th IFIP WG 8.5 International Conference (EGOV 2011)*, Springer, Heidelberg, 2011, pp. 235-246. [[Links](#)]
- [25] M. Kostovski, M. Jovanovik and D. Trajanov, Open data portal based on semantic web technologies, in *Proceedings of the 7th South East European Doctoral Student Conference*, University of Sheffield, Greece, 2012, pp. 1-13. [[Links](#)]
- [26] J. Kucera and D. Chlapek, Benefits and risks of open government data, *Journal of Systems Integration*, vol. 5, no. 1, pp. 30-41, 2014. [[Links](#)]
- [27] J. Kucera, D. Chlapek and M. Necasky, Open government data catalogs: Current approaches and quality perspective, in *Technology-Enabled Innovation for Democracy, Government and Governance* (A. Ko, C. Leitner, H. Leitold, and A. Prosser, Eds.). Berlin: Springer, 2013, pp. 152-166. [[Links](#)]
- [28] T. Lebo, et al., Producing and using linked open government data in the TWC LOGD portal, in *Linking Government Data* (D. Wood, Ed.). New York: Springer, 2011, pp. 51-72. [[Links](#)]
- [29] S. M. Lee, T. Hwang and D. Choi, Open innovation in the public sector of leading countries, *Management Decision*, vol. 50, no. 1, pp. 147-162, 2012. [[Links](#)]
- [30] M. Lnénicka, An in-depth analysis of open data portals as an emerging public e-service, *International Journal of Social, Behavioral, Educational, Economic and Management Engineering*, vol. 9, no. 2, pp. 589-599, 2015. [[Links](#)]

- [31] M. Lnénicka and R. Máchová, Open (big) data and the importance of data catalogs and portals for the public sector, in Proceedings in Global Virtual Conference: The 3rd International Global Virtual Conference (GV-CONF 2015), EDIS - Publishing Institution of the University of Zilina, Zilina, 2015, pp. 143-148. [[Links](#)]
- [32] R. P. Lourenço, An analysis of open government portals: A perspective of transparency for accountability, Government Information Quarterly, vol. 32, no. 3, pp. 323-332, 2015. [[Links](#)]
- [33] F. Maali, R. Cyganiak and V. Peristeras, Enabling interoperability of government data catalogues, in Proceedings of the 9th IFIP WG 8.5 International Conference (EGOV 2010), Springer, Heidelberg, 2010, pp. 339-350. [[Links](#)]
- [34] R. Máchová and M. Lnénicka, Exploring the emerging impacts of open data in the public sector, in Proceedings of the 20th International Conference Current Trends in Public Sector Research, Masaryk University, Brno, 2016, pp. 36-44. [[Links](#)]
- [35] R. Máchová and M. Lnénicka, Reframing e-government development indices with respect to new trends in ICT, Review of Economic Perspectives, vol. 15, no. 4, pp. 383-411, 2015. [[Links](#)]
- [36] U. Maier-Rabler and S. Huber, Open: the changing relation between citizens, public administration, and political authority, eJournal of eDemocracy & Open Government, vol. 3, no. 2, pp. 182-191, 2011. [[Links](#)]
- [37] S. Martin, M. Foulonneau, S. Turki, and M. Ihadjadene, Open data: Barriers, risks and opportunities, in Proceedings of the 13th European Conference on eGovernment (ECEG 2013), Academic Conferences and Publishing International Limited, Reading, 2013, pp. 301-309. [[Links](#)]
- [38] A. Marton, M. Avital and T. Blegind Jensen, Reframing open big data, in Proceedings of the 21st European Conference on Information Systems, Utrecht University, Netherlands, 2013, pp. 1-12. [[Links](#)]
- [39] H. Maylor and K. Blackmon, Researching Business and Management. New York: Palgrave-Macmillan, 2005. [[Links](#)]
- [40] Methodology. (2013, October) Global open data index by open knowledge. Global Open Data Index. [Online]. Available: <http://index.okfn.org/methodology/> [[Links](#)]
- [41] M. Petticrew and H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide, Malden: Blackwell Publishing, 2006. [[Links](#)]
- [42] M. Petychakis, O. Vasileiou, C. Georgis, S. Mouzakitis, and J. Psarras, A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective, Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no. 2, pp. 34-47, 2014. [[Links](#)]
- [43] L. Reggi and C. A. Ricci, Information strategies for open government in Europe: EU regions opening up the data on structural funds, in Proceedings of the 10th IFIP WG 8.5 International Conference (EGOV 2011), Springer, Heidelberg, 2011, pp. 173-184. [[Links](#)]
- [44] D. S. Sayogo, T. A. Pardo and M. Cook, A framework for benchmarking open government data efforts, in Proceedings of the 47th Hawaii International Conference on System Sciences, IEEE, 2014, pp. 1896-1905. [[Links](#)]
- [45] M. Solar, G. Concha and L. Meijueiro, A model to assess open government data in public agencies, in Proceedings of the 11th IFIP WG 8.5 International Conference (EGOV 2012), Springer, Heidelberg, 2012, pp. 210-221. [[Links](#)]

- [46] J. M. Tien, Big data: Unleashing information, Journal of Systems Science and Systems Engineering, vol. 22, no. 2, pp. 127-151, 2013. [[Links](#)]
- [47] B. Ubaldi, Open government data: Towards empirical analysis of open government data initiatives, OECD, France, Working Papers on Public Governance, no. 22, OECD Publishing, 2013. [[Links](#)]
- [48] J. Umbrich, S. Neumaier and A. Polleres, Quality assessment & evolution of open data portals, in Proceedings IEEE International Conference on Open and Big Data, IEEE, Rome, 2015, pp. 1-8. [[Links](#)]
- [49] United Nations, United Nations E-government Survey 2014: E-Government for the Future We Want. New York: UN Publishing Section, 2014. [[Links](#)]
- [50] S. Van der Waal, et al., Lifting open data portals to the data web, in Linked Open Data - Creating Knowledge Out of Interlinked Data (S. Auer, V. Bryl and S. Tramp, Eds.). Cham: Springer International Publishing, 2014, pp. 175-195. [[Links](#)]
- [51] N. Verma and M. P. Gupta, Open government data: More than eighty formats, in Proceedings of the 9th International Conference on E-Governance (ICEG 2012), CSI, Cochin, India, 2012, pp. 207-216. [[Links](#)]
- [52] World Wide Web Foundation, Open Data Barometer Global Report: Second Edition. Washington DC: World Wide Web Foundation, 2015. [[Links](#)]
- [53] H.-C. Yang, C. S. Lin and P.-H. Yu, Toward automatic assessment of the categorization structure of open data portals, in Multidisciplinary Social Networks Research (L. Wang, S. Uesugi, I.-H. Ting, K. Okuhara, and K. Wang, Eds.). Berlin: Springer Heidelberg, 2015, pp. 372-380. [[Links](#)]
- [54] Z. Yang and A. Kankanhalli, Innovation in government services: The case of open data, in Grand Successes and Failures in IT: Public and Private Sectors (Y. K. Dwivedi, H. Z. Henriksen, D. Wastell, and R. De', Eds.). Berlin: Springer, 2013, pp. 644-651. [[Links](#)]
- [55] S. Zillner, et al. (2014) Public deliverable of the EU-project BIG (318062; ICT-2011.4.4). BIG. [Online]. Available: http://big.atosresearch.eu/sites/default/files/content-files/deliverables/BIG_D2_3_2.pdf [[Links](#)]
- [56] A. Zuiderwijk and M. Janssen, A coordination theory perspective to improve the use of open data in policymaking, in Proceedings of the 12th IFIP WG 8.5 International Conference (EGOV 2013), Springer, Heidelberg, 2013, pp. 38-49. [[Links](#)]
- [57] A. Zuiderwijk and M. Janssen, Participation and data quality in open data use: Open data infrastructures evaluated, in Proceedings of the 15th European Conference on E-Government 2015 (ECEG 2015), Academic Conferences and Publishing International Limited, Reading, UK, 2015, pp. 351-359. [[Links](#)]
- [58] A. Zuiderwijk, M. Janssen and C. Davis, Innovation with open data: Essential elements of open data ecosystems, Information Polity, vol. 19, no. 1, 2, pp. 17-33, 2014. [[Links](#)]



پژوهشکده مطالعات توسعه

ارزیابی کیفیت پرتال‌های داده باز در سطح ملی

مجموعه داده‌باز و شفافیت - ۱

کد گزارش: ۱۳۷ - شماره مسلسل: ۹۹۳۱

نویسنده: رناتا ماچووا و مارتین لنیکا

مترجم: سیاوش امجدی

مرکز توانمندسازی حاکمیت و جامعه جهاد دانشگاهی

