

Open data dissemination at Eurostat: State of the art

Christine Laaboudi*, Martin Karlberg and Maja Islam
Eurostat, European Commission, Luxembourg

Abstract. Eurostat is charged with providing high-quality statistics for Europe, and the current information landscape is making it increasingly challenging to do so. This paper presents the Eurostat dissemination approach (including the traditional dissemination vectors), and thereafter proceeds to present the recent initiatives to make European statistics data and metadata available in the form of Linked Open Data (LOD). After presenting some of the main challenges for open data dissemination (complete reproducibility, availability of high-quality LOD, capacity to consume LOD and achieving meaningful mashups between official statistics LOD and other data sources), it concludes by noting the potential of LOD to foster transparency, reproducibility, collaboration, interdisciplinary research driving scientific advancements, and contributing to a broader understanding of complex scientific challenges.

Keywords: Official statistics, dissemination, linked data, open data, classification, mashup

1. Introduction

1.1. Guiding principles for Eurostat dissemination

The European Statistical System (ESS) is the partnership between the statistical authority of the European Union, i.e. Eurostat (a Directorate-General of the European Commission), and the national statistical institutes (NSIs) of the European Union Member States [1]. Eurostat has various tasks [2], not the least in the field of dissemination, where the tasks include:

- the dissemination of European statistics in accordance with the statistical principles of professional independence, impartiality, objectivity, reliability, statistical confidentiality and cost-effectiveness as defined in the Regulation on European Statistics [1] and as further elaborated in the European Statistics Code of Practice (ESCoP) [3].
- ensuring that European statistics are made accessible to all users in accordance with statistical principles – and in this respect, providing the technical

explanations and the support necessary for the use of European statistics.

Consequently, Eurostat’s overall mission is to provide high-quality statistics for Europe, and to this end, Eurostat develops and promotes standards, methods and procedures that allow the cost-effective production and dissemination of European statistics. It cooperates with international organisations in order to facilitate the global comparability of European statistics [4].

1.2. Past achievements and recent challenges for European statistics dissemination

1.2.1. Trust in an era of disinformation

The spread of disinformation and the increasing sophistication of the phenomenon, the technological transformation, a growing demand for new statistics to measure societal phenomena, users’ changing habits and expectations have all led to a systematic reflection and changes in the way Eurostat and its partners in the ESS collect, produce, disseminate and communicate statistics to their users. Arguably, a major surge in public attention to the ‘fake news’ phenomenon took place in 2016, in connection with the Brexit referendum in the United Kingdom and the presidential election in

*Corresponding author: Christine Laaboudi, Eurostat, European Commission, L-2920, Luxembourg. Tel.: +352 4301 33925; E-mail: christine.laaboudi@ec.europa.eu.

the United States. Coincidentally (or presciently), the Digital communication, User analytics and Innovative products (DIGICOM) project was conceived already in 2015 [5].

This project, which ran from 2016–2019 and brought together participants from nearly all countries of the ESS, included the dissemination and promotion of, and the communication of the value of, European statistics as a reliable basis for evidence-based decision-making and an unbiased picture of society [6]. After the end of DIGICOM, these activities have been mainstreamed, with Eurostat's main strategic goal [4] being to, in the context of growing disinformation, remain an independent and trusted point of reference for statistics and data on Europe, necessary for better decisions, policies and public debate in the European Union.

1.2.2. Open data – a necessary but not a sufficient condition

Eurostat has a long-standing tradition of open data dissemination, with all Eurostat data having been freely available online since 1 October 2004 [7]. However, the mere act of putting data in the public domain does not achieve Eurostat's task of disseminating official statistics in line with the ESCoP [3]. Most notably, ESCoP principle 15 on accessibility and clarity sets out a number of additional requirements in terms of e.g. the presentation of statistics and the corresponding metadata (including the methodology of statistical processes), the use of modern information and communication technology, methods, platforms and open data standards. Principle 15 also covers access to microdata for researchers (for which the DIGICOM project included a specific strand).

In terms of modern technology and open data standards, the DIGICOM project was a frontrunner also in this regard, with a sizeable 'Open Data Dissemination' work package, which aimed to explore and identify solutions in relation to a number of questions that arose with the prospect of disseminating official statistics in a way that would give active users as much freedom as possible to create their own products. The benefit for official statistics producers was clear: active users of official statistics can also be seen as 'redissemulators' who would create tailor-made products and services for their users/clients using official statistics, thereby enhancing data quality by adding value to the statistical information supplied, and help official statistics producers reach new audiences. Arguably, this could also (already through the increased use and perceived utility) be conducive to an increase in trust in official statistics.

The DIGICOM project did also lead to the identification of a number of concrete challenges [8] concerning Linked Open Data (LOD):

- an increased need for documentation and/or standardisation to enhance sharing,
- a number of gaps to fill for ensuring conceptual and syntactical interoperability in the ESS for LOD. Many of the existing ESS assets (data, meta-data and documentation) needed further adaptation before they could be integrated and provide a seamless experience to users.

The objectives and challenges described in the European Data Strategy [9] – which emphasises the importance of open data in driving innovation, improving public services, and promoting transparency – echo those identified as part of DIGICOM. The strategy sets out several goals related to open data, including:

- Ensuring that data are available for reuse: The strategy calls for making more data available for reuse, particularly in the public sector. This includes making data available under open formats and licenses that allow for reuse and redistribution.
- Promoting data interoperability: The strategy emphasises the importance of data interoperability, which refers to the ability of different systems and applications to link and exchange data with each other. Interoperability can in turn help to promote the reuse and sharing of data.

1.3. Contribution of this paper

As could be seen from Section 1.2.2, there are various unmet challenges which need to be tackled by Eurostat and its ESS partners in order to improve the adherence to ESCoP principle 15 and to contribute to the implementation of the European Data Strategy when it comes to tapping the full potential of open data. In this paper we will focus on recent and ongoing achievements in the field of Eurostat's dissemination to meet those challenges, with an emphasis on LOD.

We start by presenting (in Section 2), the current Eurostat dissemination approach, including the Eurostat website is a major component therein (not the least through the Eurostat data browser) and the key role it plays in the dissemination of European statistics. Thereafter, we widen the scope by presenting data.europa.eu (DEU) [10] – the central point of access to European open data from international, European Union, national, regional, local and geodata portals. This is done in Section 3, which focuses on the collaboration between

Eurostat and the Publications Office of the European Union (OP) when it comes to populating DEU with quality European statistics metadata. Considering the high impact of European statistical classifications, and the various ongoing initiatives in improving their dissemination, we treat them separately in Section 4. We then proceed to present a number of challenges – many of them already being tackled by the official statistics community – in Section 5, and end with some concluding remarks in Section 6.

2. The Eurostat dissemination approach

Guided by the principles set out in 1.1, the Eurostat communication and dissemination strategy [11] defines the operational framework for ensuring that trustworthy European statistics are widely accessible to users and also well understood by anyone looking for reliable data on Europe. It describes the wide range of existing statistical products on offer, highlights the areas that will require further attention in the future, and lists the actions to be taken at different stages of the communication cycle.

2.1. Users in focus

When developing its products and services, Eurostat takes a number of steps to ensure its communication and dissemination actions are user-centered, starting from knowing the Eurostat users and their needs. To this end, the DIGICOM project featured a user analytics work package, which for instance rendered results [8] in the form of a typology of Eurostat users (user personas), guidelines on user analytics and usability guidelines.

Many of the good practices established in the project have been mainstreamed and integrated into the Eurostat approach, which is based on user behaviour analysis as well as user feedback obtained in several ways, from user interactions via the support system, social listening on Eurostat social media channels, to usability testing and user research.

2.2. Metadata as a prerequisite for useful official statistics

Official statistics data are of little use without the accompanying metadata. Metadata provide essential information about the data, enabling users to identify them, to understand the content of the data, to get information on how to access or download them and to assess

the data quality. For maximum utility, they should be offered in both human-readable and machine-readable format.

Accordingly, principle 15 of the ESCoP states that data should always be made available with supporting metadata. Eurostat fully subscribes to this principle, and consistently integrates the appropriate metadata in all its vectors of dissemination [12]. This includes [13]:

- structural metadata which are used to represent the structure of the dataset (dimensions, attributes, variables), as well as
- reference metadata that describe statistical concepts, methodologies used for the generation of data or evaluation of the quality.

2.3. Dissemination vectors

2.3.1. Eurostat website

The full range of Eurostat's products and services is provided on the Eurostat public website (ec.europa.eu/eurostat) free of charge, as has been the case since 2004 [7]. The Eurostat website was revamped in 2022, with the aim of facilitating data access with new and improved features and making it clearer and more user friendly as well as fully in line with the European Commission Web Accessibility Directive [14].

2.3.2. Eurostat data browser

The main tool for data dissemination is the data browser [15]. The data browser was launched in November 2022, and is the result of a multiannual project to overhaul the Eurostat dissemination chain. The data browser allows users to customise, visualise and extract statistical data in an easy and interactive way. The data browser gives users easy access to data and metadata and gives them control of their user experience by allowing them to easily customise data visualisations and save their favourite views for later use, download data in a wide range of formats, including Excel, SDMX 2.0 and 2.1, TSV and JSON-stat, easily share their datasets through bookmarks and social media, and more. New users benefit from a comprehensive online support system, which includes a 'first visit' guide [15].

2.3.3. Machine-to-machine access to Eurostat data

While the data browser offers a traditional ('point and click') user interface, more advanced users (including for instance research institutions and private enterprises active in the data economy) need to access data without unnecessary manual operations. Therefore, Eurostat

offers an application programming interface (API) for data access [16]. Given the increasing use of the R programming language in official statistics, Eurostat has developed the **restatapi** package [17] to allow R users an easy way to retrieve data via the Eurostat API.

3. Open data dissemination of European statistics via data.europa.eu

3.1. Exposing European statistics datasets on data.europa.eu

In 2017, Eurostat started uploading the catalogue of its data on the EU Open Data Portal, a predecessor to DEU [10]. The catalogue includes a description of the Eurostat datasets in RDF (Resource Description Framework) format including links to the distributions or visualisation of the datasets, and their reference metadata in different formats (SDMX, CSV, TSV).

Today, over 8 000 Eurostat datasets are published on DEU – the main point of access for European open data that aims to improve access to open data, foster high-quality open data publication at all levels, and create impact through open data reuse – with Eurostat feeding the portal twice daily. It should be noted that the datasets themselves do not reside on the portal – DEU provides links to the data and serves as an entry point, allowing Eurostat datasets to be discovered in various ways.

3.2. Common vocabularies for describing European statistics datasets

Each description in DEU follows the Data Catalogue vocabulary Application Profile (DCAT-AP) [18] specifications, which provides a common vocabulary used for describing the resources in data catalogues with the objective to enhance the data findability and to promote the reusability.

While DCAT-AP provides specifications for describing any type of dataset from the public sector, StatDCAT-AP [19], which is an extension of DCAT-AP, enables the description of statistical datasets within the statistical domain. It provides a dissemination vocabulary for statistical open data, defining several additions to the DCAT-AP model that can be used to describe the structure of the statistical datasets such as the dimensions and attributes, units of measurement, quality annotations, the number of data series or the length of time series. To enrich the descriptions of European statistics datasets, Eurostat is currently working on the integration of StatDCAT into DEU.

3.3. Identification of all European statistics datasets through persistent identifiers

3.3.1. Persistent identifiers for dataset descriptions

Each dataset description in the data catalogue is identified by a unique persistent identifier (PID) that is both human- and machine-readable. For instance, the PID <http://data.europa.eu/88u/dataset/zissonexqubduu7v16ta> leads to the description of the Eurostat dataset ‘Production of milk powder’.

While these PIDs are unique, they are specific to DEU. In contrast, Digital Object Identifiers (DOIs) are in common use, and hence more commonly recognised – in particular in the scientific community. A DOI is a specific type of PID and is composed of unique strings of characters used to permanently identify a digital asset – such as a dataset or a scientific article. They are often found on the internet in the form of a link which enables any potential user to reliably locate a digital asset.

3.3.2. Persistent identifiers for datasets in the form of DOIs

In February 2023, Eurostat started assigning DOIs to its datasets to permanently identify them. The datasets published by Eurostat are assigned DOIs in the unique namespace <https://doi.org/10.2908>, with 10.2908 being the prefix applied for all Eurostat data. As an example, the DOI **10.2908/TAG00039** [20] resolves directly to the dataset ‘Production of milk powder’ in the Eurostat data browser.

As the official DOI registration agency for the institutions, bodies, offices and agencies of the European Union, the OP registers the DOIs and their metadata at DataCite [21], a nonprofit organisation that provides persistent identifiers for research data. DataCite has its own metadata schema [22], which offers core metadata properties chosen for an accurate and consistent identification of a resource for citation and discovery purposes. DataCite creates PIDs for its dataset descriptions in a consistent manner. For instance, the aforementioned DOI 10.2908/TAG00039 does, when appended to a common stem, constitute the PID of the DataCite description [23] of the European statistics dataset ‘Production of milk powder’ [20].

Eurostat foresees to assign DOIs to all European statistics datasets, and thereafter add the European statistics DOIs to those dissemination vectors (see 2.3) for which this is practically feasible.

The advantages of using DOIs for European statistics datasets include the following:

- A unique and persistent identifier for datasets ensures that the data can be identified and accessed

by other researchers, even if the data are moved to a new location.

- Improving the discoverability of datasets.
- Helping researchers track citations of a specific dataset and avoid citation errors, such as citing a different dataset with the same title. Researchers can cite any source data that they have reused or integrated (multiple sources) into a new dataset.
- Fostering consistency and interoperability across different data repositories and platforms.
- Facilitating the tracking of usage metrics and analytics for research datasets, which can provide insights into how the data are being used and shared by the research community.

To summarise, the use of DOIs is essential in ensuring that research datasets are easily accessible, discoverable, and citable, which in turn helps to facilitate the advancement of science and innovation.

4. Linked open data dissemination of statistical classifications

4.1. The key role of classifications for unlocking the potential of LOD

When coupled with the appropriate metadata architecture, metadata have the potential to improve findability, accessibility, storage, preservation, analysis, comparison, reproducibility, inconsistency identification, correct interpretation, visualisation, data linkage, assessment and ranking of the quality of data and avoiding unnecessary duplication of data [24].

Statistical classifications (used for standardising concepts in a statistical domain) constitute one key category of structural metadata, as they are necessary for the production of reliable, comparable and methodologically sound official statistics. As described by Hoffmann and Chamie ([25] p. 2), classifications group and organise information meaningfully and systematically into a standard format and involve an exhaustive and structured set of mutually exclusive and well-described categories. Therefore, they play a crucial role in organising, integrating, and leveraging the potential of LOD – enhancing, *inter alia*, data discovery and search as well as data integration, interoperability and comparability.

Eurostat has a high level of knowledge and experience in the development of classifications and is the custodian of several sectoral and transversal European statistical classifications used to produce European statistics [26]. Eurostat is also responsible for covering the

European dimension of the international statistical classifications (ISIC, CPC) that are reference classifications for European statistical classifications (NACE, CPA) under its responsibility. As illustrated in Part I of the NACE Rev. 2 introductory guidelines [27], each statistical classification typically exists in a statistical ecosystem, where it is normally interlinked with other classifications – either structurally, or by means of correspondence tables.

4.2. Streamlining the dissemination via the Euro SDMX registry

Since the early 2000s and until 2023, Eurostat disseminated the statistical classifications used for the production of European statistics (as well as relevant correspondence tables involving those classifications) via the ‘Eurostat Reference and Management of Nomenclatures’ platform (RAMON). Once the decision was taken to phase out RAMON, Eurostat seized the opportunity to streamline and modernise the way in which statistical classifications are disseminated.

One of the ways in which the dissemination of classifications was upgraded was via the Euro SDMX Registry [28]. This registry is the Eurostat implementation of the SDMX Registry specifications as published by the Statistical Data and Metadata Exchange (SDMX) initiative. To streamline its dissemination of statistical classifications, Eurostat has converted all the classifications previously available in RAMON into SDMX/XML format and is now disseminating them via the Euro SDMX Registry.

4.3. Dissemination of statistical classifications as LOD

Eurostat also pursued a second approach by converting the main classifications used for the production of European statistics into RDF format and exposing them as LOD in EU Vocabularies [29] and in Cellar (the semantic repository of the OP) [30]. This was done with the aim of increasing data FAIRness (Findability, Accessibility, Interoperability and Reusability) in the ESS and beyond.

4.3.1. Formatting statistical classifications for LOD dissemination – from SDMX to RDF

Eurostat bases its second approach on the SDMX terminology, reinterpreted in the context of LOD. While there is no single formal RDF ontology that provides a full one-to-one equivalent for the SDMX Information Model, the most relevant ontology that can cover the modelling of statistical classifications is the Extended

Knowledge Organization System (XKOS) [31] which is an extension (for representing statistical classifications) of the Simple Knowledge Organization System (SKOS) [32] that meets domain-relevant community standards and best practices. XKOS is derived from the generic statistical information model (GSIM) [33], a terminology and a conceptual model that defines the concepts relevant to structuring statistical classification metadata. In relation to the SDMX artefacts, XKOS has the added advantage of being compliant with the semantic web technologies and allowing a richer description of the resources, rendering them interoperable and machine-readable [34].

4.3.2. Persistent identifiers for European statistical classifications

Eurostat LOD classifications are defined in the domain ‘data.europa.eu’, with one namespace assigned per classification (for example, <http://data.europa.eu/ux2>, where ‘ux2’ identifies the NACE classification), one identifier per version of the classification (for example, <http://data.europa.eu/ux2/nace2/>, where ‘nace2’ identifies NACE Rev. 2), and with a persistent URI (Uniform Resource Identifier) being assigned to each type of resource: Classification, Classification item forming part of a Classification, Classification Level, Correspondence Table and its concept associations.

4.3.3. Technology stack for LOD dissemination of European statistical classifications

For the storage and dissemination of Eurostat classifications in RDF, a suite of four semantic platforms offered by the OP is used, building on three operational pillars:

- (1) reference data maintenance
 - **VocBench** [35], a web-based collaborative semantic application enabling the creation and maintenance of generic RDF datasets and, in particular, controlled vocabularies (thesauri, classifications/nomenclatures or code lists). In addition to SKOS, the default data model in VocBench, it supports the integration of different ontologies such as XKOS applied for modelling statistical classifications, and the Core Ontology for Official Statistics (COOS) [36] describing the production of statistics, applied for modelling metadata catalogues (e.g. statistical methodologies, standards or organisations). Furthermore, it eases the creation and validation of correspondence tables between two classifications or concordance between two versions of the same classification.

- (2) visualisation
 - **ShowVoc** [37], an additional web-based collaborative semantic application, allowing an easier display and browsing of datasets maintained with VocBench,
- (3) storing for sharing and re-use,
 - **Cellar** [30], a large semantic dissemination knowledge graph (Triple Store), exposes all OP documents (EU publications, EU legislation) and their metadata as LOD, and is the main point of interaction for systems and applications consuming the data through a SPARQL Endpoint [38] and an API.
 - **DEU** [10], the central point of access of European open data, gives access to the description of the classification datasets with a link to their distributions for download.

To summarise, these tools jointly provide a solid back-end for data owners (including Eurostat) to maintain and expose their data assets, a user-friendly front-end for users to discover and view the data assets, and the necessary infrastructure for programmatic access (to allow automatic re-use of these assets).

5. Challenges to be overcome for open data dissemination

5.1. Reproducibility through blockchain technology

It should be noted that even with DOIs fully implemented for all European statistics datasets, there may be issues with reproducibility, since European statistics data currently are not versioned. For instance, whenever data are revised (for instance to replace preliminary data with final data or to correct for errors), previously disseminated data are ‘overwritten’. In such a case, a researcher or policy analyst trying to reproduce the results of a previous analysis will – even when using the exact same DOI and the exact same selection criteria – arrive at different results.

While this reproducibility issue could technically be resolved by taking a ‘snapshot’ of each disseminated Eurostat dataset whenever it is being updated, this would generate large volumes of near-identical data of little public utility. Eurostat is therefore currently considering using a more ‘lightweight’ approach based on blockchain technology [39]. The approach would essentially entail the following:

- Eurostat injects ‘hashes’ (digital fingerprints) of

each disseminated version of a dataset into a blockchain.

- Any researcher or analyst A interested in ensuring the reproducibility of their results would have to download the Eurostat data that they use and (using their own infrastructure) save those data.
- To credibly demonstrate the reproducibility of their findings the researcher/analyst A would then (on top of sharing their code and the DOI of the data that they have used) also have to share the thus saved Eurostat data in unaltered form.
- Although the data to which the DOI resolves might have been updated, any other researcher/analyst B wishing to reproduce the findings could then verify that the data used are indeed authentic by checking that the ‘hash’ of the data shared by A does appear in the ‘Eurostat blockchain’.

Thereby, all researchers and analysts wishing to achieve demonstrable reproducibility of their results would have the means to do so. Moreover, by researchers and analysts only saving the dataset versions underlying their analyses, there is no wasteful use of storage space for the various intermediate versions of data that nobody ever uses.

5.2. High-value datasets

An important step towards new ESS capacity development and an increased quality of the open data dissemination by its NSIs was taken through the adoption of a list of high-value datasets (HVDs) for statistics [40]. Member States must disseminate these HVDs (i) for free, (ii) in machine-readable format – and made available through (iii) APIs and (iv) for bulk download. To support national authorities in their dissemination of HVDs, guidelines [41] have been issued on how to use DCAT-AP [18] for a dataset that is subject to the requirements set out in the regulation [40] on HVDs.

While the ESS NSIs already do disseminate their statistics in machine-readable format for free, a number of NSIs do not yet expose their data via APIs or bulk download facilities. As dissemination of HVDs by Member States is mandatory, this will serve as an incentive for those NSIs of the ESS that do not yet have facilities for making their data available through APIs (and via bulk download) to develop their infrastructure. This could possibly be done through collaboration among NSIs to achieve a standardised approach and economies of scale. Once an infrastructure for disseminating statistics HVDs is in place, it will benefit the full range of products of the NSI, since it could be used for the dissemination of all their datasets – not just the HVDs.

5.3. Linked open data challenges for official statistics

5.3.1. Making statistical data available in RDF format

The transformation of existing data into LOD increases the opportunities for further collaboration between Eurostat and the NSIs of the ESS for developing, reusing, and linking reference and derived classifications. The main challenge for enabling this interoperability remains the availability of these statistical classifications in RDF. A successful interoperability use case is the availability of correspondence tables established between international and EU statistical classifications (NACE – ISIC, CPA – CPC), accessible remotely on EU Vocabularies [29] or on the Caliper platform [42], a project run by the Food and Agriculture Organization of the United Nations (FAO).

5.3.2. An international Community of Practice

Under the auspices of the ESS Standards Working Group, the LOD Community of Practice (LOD CoP) was launched in April 2023. This initiative, coordinated by Eurostat, includes nine ESS NSIs (the Netherlands, Latvia, Croatia, Hungary, Spain, France, Italy, Finland, Denmark, and Norway), Statistics Canada and the FAO. The LOD CoP is developing use cases and recommendations for:

- linking structural metadata to statistical datasets,
- linking statistical classifications,
- defining specifications for a common API for retrieving classifications and correspondence tables,
- linking statistical datasets across data catalogues.

5.3.3. Consuming LOD

It is not worth investing resources in LOD assets if they are ultimately not used by at least some categories of official statistics users. A part of the reason potentially hampering the use of LOD may be that there are not enough relevant datasets around – so the various initiatives already achieved and underway (as described in 4 and 5.3.1) are a crucial first step.

However, another hurdle to overcome before there could be greater uptake of LOD in the official statistics user community concerns the difficulty to consume them. Zeginis et al. [43] suggest that ‘*to unleash the full potential of [LOD] we need to facilitate the interaction with [LOD] and hide most of the complexity*’. As an example, even official statistics users with considerable IT skills might struggle with non-traditional query languages such as SPARQL – if they do have experience in query languages, it would typically be those of the ‘SQL’ variety. Some initiatives are already underway to remedy this.

First, Eurostat has developed an R package for automatically generating or updating candidate correspondence tables between two classifications. As described by Karlberg et al. [44], this package is currently being extended to facilitate data ingestion through a function directly accessing classifications and correspondence tables data via SPARQL Endpoint APIs (Cellar of the OP [30] or Caliper of the FAO [42]). Apart from meeting the most pressing needs of official statistics users ('getting the data'), it also serves a didactic purpose: the SPARQL code used to retrieve the classifications (and correspondence tables) is also returned by the function – thus allowing users to see what SPARQL code of relevance to them looks like. Ideally, this will allow official statistics users with good general coding skills to figure out how to tweak the SPARQL code themselves so that they can apply it for other purposes.

Moreover, as part of the ongoing collaboration between Eurostat and the OP [45], an initiative has been launched to better tailor ShowVoc [37] to statistical classifications through improved formatting, adapted terminology (replacing 'LOD jargon' with use terminology commonly used in the classification community) and relevant ergonomics.

5.3.4. Going beyond official statistics

While the initiatives described in this paper focus on official statistics, it has to be borne in mind that to truly unleash the full potential of LOD, a wider group of use cases should also be brought into the picture. In a crisis situation, a policymaker might wish to rapidly get dashboard-like information on a phenomenon from whatever source is available. In principle, LOD was conceived for situations like this, and is designed to allow 'mashups' of different sources, such as those envisaged by DiFranzo et al. [46].

Exposing official statistics data as LOD opens up for this opportunity – but the organisations disseminating official statistics might need to reflect on the scope of their role therein. For instance: is it the role of the official statistics community to guide key users (policymakers and policy analysts) in their use (so that the mashups use quality data whenever available), or does our responsibility stop with putting the LOD 'out there'?

6. Conclusions

Eurostat does, like the official statistics community in general, have a longstanding tradition of open data

dissemination. However, just disseminating data is not enough. Section 1 discusses the communication and dissemination activities, challenges and actions required to ensure trust in official statistics. As outlined in Sections 3 and 4, Eurostat has recently taken considerable steps to add to its regular dissemination vectors (described in Section 2) by making key data assets available in LOD format. Linked Open Data can provide substantial support to scientific research by offering a wealth of interconnected and openly accessible data resources and formats.

Linked Open Data empower scientific researchers by providing a framework for accessing, integrating, and analysing interconnected data. It fosters transparency, reproducibility, collaboration, and interdisciplinary research, driving scientific advancements and contributing to a broader understanding of complex scientific challenges. However, while LOD offer many benefits, several challenges still need to be addressed in order to create an enabling environment for exploiting their full potential. This includes, but is not limited to, enhancing data quality, privacy, curation, full accessibility, interoperability, and sustainability.

Addressing these challenges requires a concerted effort from a range of stakeholders, including policymakers, data providers, and users of open data. As illustrated in Section 5, Eurostat is committed to continue working in this area in collaboration with the key actors at EU level, such as the OP, as well as partners in the ESS and worldwide. Thereby, European statistics will become even more widely accessible to anyone looking for reliable data on Europe – and (as discussed in Section 5.3.4) their interoperability and integration with other sources will become possible.

References

- [1] European Union. Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics. Official Journal of the European Union. 31.3.2009 (L 087): 164. Available from: <https://eur-lex.europa.eu/eli/reg/2009/223/2015-06-08/> [Accessed 19 October 2023].
- [2] European Commission. Commission Decision of 17 September 2012 on Eurostat (2012/504/EU). Official Journal of the European Union. 18.9.2012 (L 251): 49-52. <https://eur-lex.europa.eu/eli/dec/2012/504/oj/> [Accessed 19 October 2023].
- [3] Eurostat. European Statistics Code of Practice – revised edition 2017. Luxembourg: Publications Office of the European Union. doi: 10.2785/798269 [Accessed 19 October 2023].
- [4] Eurostat. Strategic Plan 2020–2024, published 26 October 2020. Available from: https://commission.europa.eu/publications/strategic-plan-2020-2024-eurostat_en [Accessed 19 October 2023].

- [5] Eurostat. Digital communication, User analytics and Innovative products (ESS.VIP DIGICOM) project Business case; 2015. Available from: <https://cross-legacy.ec.europa.eu/content/essvip-digicom-business-case-version-100> [Accessed 19 October 2023].
- [6] Karlberg M, Czumaj E, de Jong-de Heer H, Gálvez Moraleda A, Hagenkört-Rieger S, Pinto Martins J, McCuirc E, Mordant G, Mottura G, Orjala H, Schulz T, Tomaschek B, Vichi M. DIGICOM – an unprecedented collaboration on the dissemination and communication of European statistics. Online proceedings of the 2021 Conference on New Techniques and Technologies for Official Statistics (NTTS 2021). Available from: https://coms.events/NTTS2021/data/abstracts/en/abstract_0092.html [Accessed 19 October 2023].
- [7] Eurostat. Free access to all Eurostat data and publications. Eurostat news release 148/2004 of 13 December 2004. Available from: <https://ec.europa.eu/eurostat/web/products-euro-indicators/-/1-13122004-ap> [Accessed 19 October 2023].
- [8] Eurostat. Digital communication, User analytics and Innovative products (DIGICOM) – Project End Report; 2020. Available from: <https://europa.eu/!bQ76kU> [Accessed 19 October 2023].
- [9] European Commission. A European strategy for data – Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions of 19 February 2020. Brussels: European Commission. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066> [Accessed 19 October 2023].
- [10] European Commission. Documentation of data.europa.eu (DEU) – Version 1.0; 2022. Available from: https://dataeuropa.gitlab.io/data-provider-manual/pdf/documentation_data-europa-eu_V1.0.pdf [Accessed 19 October 2023].
- [11] Eurostat. Eurostat Communication and Dissemination Strategy 2021–2024. Available from: <https://europa.eu/!mqC8D3> [Accessed 26 October 2023].
- [12] Eurostat. Metadata overview. [homepage on the internet]. Luxembourg: Eurostat. Available from: <https://ec.europa.eu/eurostat/web/metadata/> [Accessed 26 October 2023].
- [13] Díaz Muñoz P. The role of Statistical Data and Metadata Exchange in global statistical infrastructure. *Statistical Journal of the IAOS*. 2008; 25: 47-54.
- [14] European Union. Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies. *Official Journal of the European Union*. 2.12.2016 (L 327): 1-15. Available from: <https://eur-lex.europa.eu/eli/dir/2016/2102/oj> [Accessed 26 October 2023].
- [15] Eurostat. Data Browser First Visit Guide. [homepage on the internet]. Luxembourg: Eurostat. Available from: <https://wikis.ec.europa.eu/display/EUROSTATHELP/Data+browser+first+visit> [Accessed 26 October 2023].
- [16] Eurostat. Eurostat SDMX RESTful API. [homepage on the internet]. Luxembourg: Eurostat. Available from: <https://ec.europa.eu/eurostat/api/dissemination/swagger-ui> [Accessed 26 October 2023].
- [17] Mészáros M, Weinand S. restatapi: Search and Retrieve Data from Eurostat Database; 2023. R package version 0.22.1. Available from: <https://github.com/eurostat/restatapi> [Accessed 04 January 2024].
- [18] DCAT Application Profile for data portals in Europe [homepage on the internet]. Brussels: European Commission. Available from: <https://europa.eu/!q63dF3> [Accessed 7 November 2023].
- [19] Pellegrino M. StatDCAT-AP: Representing statistical metadata by using the ‘DCAT application profile for data portals in Europe’; 2017. Paper presented at the Joint UNECE/UN-GGIM Workshop on Integrating Geospatial and Statistical Standards. Available from: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2017/mtg3/2017-UNECE-topic-i-EC-StatDCAT-ap-pape__1_.pdf [Accessed 7 November 2023].
- [20] Eurostat 2023. Production of milk powder. [European statistics dataset]. doi: 10.2908/TAG00039.
- [21] Datacite [homepage on the internet]. <https://datacite.org/> [Accessed 7 November 2023].
- [22] DataCite Metadata Schema [homepage on the internet]. The hyperlink is broken, it must be [https:// schema.datacite.org/](https://schema.datacite.org/) [Accessed 7 November 2023].
- [23] DataCite 2023. Production of milk powder [homepage on the internet]. <https://commons.datacite.org/doi.org/10.2908/tag00039> [Accessed 7 November 2023].
- [24] Zuiderwijk A, Jeffery K, Janssen M. The Potential of Metadata for Linked Open Data and its Value for Users and Publishers. *EJournal of EDemocracy and Open Government (JeDEM)*. 2012; 4(2): 222-244. doi: 10.29379/jedem.v4i2.138.
- [25] Hoffmann E, Chamie M. *Standard Statistical Classifications: Basic Principles*. New York: United Nations Statistics Directorate. Available from: https://unstats.un.org/unsd/classifications/bestpractices/basicprinciples_1999.pdf [Accessed 8 August 2023].
- [26] Classifications [homepage on the internet]. Luxembourg: Eurostat. Available from: <https://ec.europa.eu/eurostat/web/metadata/classifications> [Accessed 7 November 2023].
- [27] Eurostat. NACE Rev. 2 – Statistical classification of economic activities in the European Community; 2008. Luxembourg: Publications Office of the European Union. Available from: <https://europa.eu/!QtpWF4> [Accessed 7 November 2023].
- [28] Euro SDMX Registry [homepage on the internet]. Luxembourg: Eurostat. Available from: <https://webgate.ec.europa.eu/sdmxregistry/> [Accessed 7 November 2023].
- [29] EU Vocabularies – Eurostat statistical classifications [homepage on the internet]. Luxembourg: Publications Office of the European Union. <https://op.europa.eu/en/web/eu-vocabularies/eurostat> [Accessed 7 November 2023].
- [30] Publications Office of the European Union. Cellar – The semantic repository of the Publications Office. doi: 10.2830/654692. [Accessed 4 January 2024].
- [31] Cotton F, Gillman DW, Joque Y. XKOS – An RDF Vocabulary for Describing Statistical Classifications. *IASSIST Quarterly*. 2015; 38(4): 47-57. doi: 10.29173/iq900.
- [32] Miles A, Pérez-Agüera JR. SKOS: Simple knowledge organisation for the web. *Cataloging & Classification Quarterly*. 2007; 43(3-4): 69-83. doi: 10.1300/J104v43n03_04.
- [33] United Nations Economic Commission for Europe. Generic Statistical Information Model (GSIM): Statistical Classifications Model. Paper presented at the 2015 Meeting of the Expert Group on International Statistical Classifications. Available from: <https://unstats.un.org/unsd/classifications/expertgroup/egm2015/ac289-22.PDF> [Accessed 7 November 2023].
- [34] Dzikowski C, Cotton F. Best Practices for describing statistical classifications with XKOS; 2023. Online paper of the DDI alliance. Available from: <http://linked-statistics.github.io/xkos/xkos-best-practices.html> [Accessed 7 November 2023].
- [35] Stellato A, Fiorelli M, Turbati A, Lorenzetti T, van Gemert W, Dechandon D, Laaboudi-Spoiden C, Gerencsér A, Waniart A, Costetchi E, Keizer J. VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons. *Semantic Web*. 2020; 11(5): 855-881. doi: 10.3233/SW-200370.

- [36] Rizzolo F, Gillman D, Vucko F. COOS – A Core Ontology for Official Statistics; 2023. Online UNECE paper. Available at <https://linked-statistics.github.io/COOS/coos.html> [Accessed 7 November 2023].
- [37] Stellato A, Fiorelli M, Turbati A, Lorenzetti T. A EU workflow for seamless maintenance and publication of data, metadata and legal acts. Paper presented at the 3rd National Conference on Artificial Intelligence (Ital-IA 2023). Available from: <https://ceur-ws.org/Vol-3486/173.pdf> [Accessed 7 November 2023].
- [38] Eurostat. SPARQL Queries – Short User Guide [homepage on the internet]. Available from: <https://europa.eu/!NcJfQY> [Accessed 10 November 2023].
- [39] Tessitore C. The usage of blockchain technology for official statistics; 2023. Book of abstracts of the 2023 conference on New Techniques and Technologies for Official Statistics (NTTS 2023); 449-451. Available from: https://cros-legacy.ec.europa.eu/content/NTTS2023_en [Accessed 7 November 2023].
- [40] European Commission. Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use. Official Journal of the European Union 20.1.2023 (L 19): 43-75. Available from: http://data.europa.eu/eli/reg_impl/2023/138/oj [Accessed 10 November 2023].
- [41] Fragkou P. DCAT-AP High Value Datasets. 19 June 2023. SEMIC community online document. Available from: <https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd/> [Accessed 14 November 2023].
- [42] Caracciolo C, Aubin S, Jonquet C, Amdouni E, David R, Garcia L, Whitehead B, Roussey C, Stellato A, Ferdinando V. 39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition. CODATA Data Science Journal, Committee on Data for Science and Technology (CODATA). 2020; 19(1): 47. doi: 10.5334/dsj-2020-047.
- [43] Zeginis D, Kalampokis E, Roberts B, Moynihan R, Tambouris E, Tarabanis K. Facilitating the Exploitation of Linked Open Statistical Data: JSON-QB API Requirements and Design Criteria. In: Capadisli S, Cotton F, Dong XL, Guha RV, Haller A, Hitzler P, Kalampokis E, Kejriwal M, Lecue F, Sivakumar D, Szekely P, Troncy P, Witbrock M, editors. Joint Proceedings of the 2017 International Workshops on Hybrid Statistical Semantic Understanding and Emerging Semantics, and Semantic Statistics (HybridSemStats 2017). Aachen: CEUR Workshop Proceedings; 2017. Available at: <https://ceur-ws.org/Vol-1923/article-11.pdf> [Accessed 10 August 2023].
- [44] Karlberg M, Chasiotis V, Stavropoulos P, Laaboudi C, Mészáros M, Nasiopoulou D-A. An R package for automatically generating candidate correspondence tables between classifications. Statistical Journal of the IAOS. 2023; 39(4): 995-1009. doi: 10.3233/SJI-230039.
- [45] Dechandon D, Laaboudi C. A successful collaboration on classifications and what is in it for the statistical community; 2023. Paper presented at the 2023 conference on New Techniques and Technologies for official statistics (NTTS 2023). Available from: <https://europa.eu/!6yjHrq> [Accessed 2 June 2023].
- [46] DiFranzo D, Graves A, Erickson JS, Ding L, Michaelis J, Lebo T, Patton E, Williams GT, Li X, Zheng JG, Flores J, McGuinness DL, Hendler J. The Web is My Back-end: Creating Mashups with Linked Open Government Data; 2011. In: Wood, D. (eds) Linking Government Data. New York: Springer. doi: 10.1007/978-1-4614-1767-5_10.