Delivering Official Statistics as *Do-It-Yourself* services to foster produsers' engagement with Eurostat open data

Keywords: open data; data-as-a-service; co-design & co-creation; participation & collaboration; openness and transparency; open source community; citizen statistics; civic epistemology.

1. INTRODUCTION: BEYOND OPENING DATA, ENCOURAGING PARTICIPATION

It is commonly recognised that public organisations, including National Statistical Offices (NSOs), could make better use of data resources [1], in particular in two areas:

- for *data-informed and evidence-based decision-making*: in the era of big data, public organisations should leverage the use of data to inform policy decisions,
- in terms of *stakeholder participation and public transparency*: the growing demands for more openness and transparency in public organisations underpin the movement towards more inclusive and participative decision-making systems.

Because the former area is the primary concern of the NSOs, the European Statistical System (ESS) very early addressed this challenge under different angles while focusing its strategic development towards the improved integration of (existing and new) data sources into statistical processes and products [2]. Instead, the latter area has been mostly addressed under the sole angle of *Open Data*, since it is often believed that by opening up data to the people they aim at serving, NSOs instantaneously become, likewise other public organisations, more transparent and accountable [3].

Opening up data obviously provides the opportunity to involve actors from outside the ESS – say *produsers*, *e.g.*, statisticians, scientists, citizens – and promote innovative, user-centric ways to tackle new and existing policy issues by co-designing statistical products [4]. This also has the potential to increase NSOs efficiency and effectiveness [5]. However, open data alone does not automatically translate to public participation to the decision-making [6]. In most data provision services, there is an overarching top-down ideology since statistical processes are still owned, dictated and designed by NSOs and the final users are only involved as the receivers of the data and/or services[1]. We propose to move away from the current approach by providing tools and software for accessing and using online data, so as to enable sharing best practices, learning from others' experience, adopt common methodologies, enhance cooperation between data producers and data users, and further engage in *Open Data*-driven innovation [3][4].

2. OBJECTIVES: DRIFTING AWAY FROM A TOP-DOWN APPROACH

Often, it is difficult to use data for an innovative purpose that was not anticipated by the NSOs, and the main barriers for it are very much associated with technological factors [1][4]. These barriers include the availability of data in high-quality and easily usable forms, which should be supported by the availability of concrete tools and applications to showcase and facilitate the (re)use of data. Indeed, the value of data comes not only from their availability and accessibility, but from their actual use and "transformation" into insightful or actionable information [3]. To enable the different actors – NSOs and *produsers*, but also private or third sector organisations – to add value to the different phases of the information creation, the proposed approach revolves around the provision of *"Do-It-Yourself"* (DIY) services on top of open data (*e.g.*, disseminated through

[1] When implementing Experimental Statistics, Eurostat invites the user to provide feedback on the final products (as for late 2018, and while the project started in late 2017, less than 20 comments have been posted on the dedicated forum), but does not engage with him/her at any stage of the actual statistical production process.

online database, web-services and/or REST Application Programming Interfaces). Ultimately, the traditional top-down data and service delivery model shall be revised – with all actors viewed as partners in data service creation, undertaking any of the data and/or services designer/producer/user roles – and there should be a focus on collaboration between the different parties.

In this regard, we introduce herein various tools and software[2] that aim at facilitating the use of Eurostat data by better supporting their target users, at least for the most common and generic use cases. They contribute to the development of an open, modular, scalable and interoperable software ecosystem around open data and services whose adoption will result from the (bottom-up) engagement with external actors rather than any (top-down) standard obligation. Based on best practices from the *Open Source Software* community [7], data services can be co-implemented together with *produsers* who can also be involved in extending and monitoring the services through providing feedback. Feedback is a core aspect of *Open Data*-driven public services [6][8], it comes in various forms – *e.g.* as useful contributions and significative improvements like increased robustness, additional features, *etc…* – and it can be received in regards to the data being offered as well as the services themselves. Overall, the tools and software presented here offer the prospect of engaging *produsers* in the creation, implementation, testing and validation of statistical products [8]. Such collaborative approach to statistical production emerges as an important way to innovate services traditionally provided in a top-down manner.

3. DIY DATA SERVICES AND SOFTWARE TO ACCESS AND PROCESS EUROSTAT DATA

Supporting data analytics and visual analytics, which aim eventually at extracting valuable information from data to use in intelligent ways by means of advanced statistical, computational and visualisation techniques, offers big opportunities for *Open Data*-driven innovation for decision-making in public services [3]. While hackathons are nowadays identified as successful initiatives to do so, the line of reasoning: *#opendata + #hackathon = #opengov* we are witnessing may send out the wrong message to the *produsers*' community[3]. Instead, we believe that the sharing of open and reusable (and reproducible) tools and software (see **Figure A**) enable to actively and durably engage with this community [7]. Using DIY tools, *produsers* with advanced skills can mine and analyse open data to explore patterns or discover problems and link them to policy issues [6]. Not only they can fully reproduce experiments, by rerunning or tweaking previous data analyses, they can also judge for themselves if they agree with the analytical choices, possibly identify innocent mistakes and try other routes [5].

3.1. Offering processing services to data analysts

The *pyrostat* module and the *java4eurostat* library aim at providing a data access layer for developers – in *Python* and *Java* respectively, similarly to the *R* package *eurostat* – to query *Eurostat* online database programmatically so as to process *Eurostat* data (filter/manipulate/export the data, explore/merge with other data sources, *etc…*). For that purpose both tools provide with a number of functions/methods to load statistical data – *e.g.*, into a hypercube structure or a dataframe – and index it for efficient in-memory computations. They not only aim at being (re)used by the data analysts, but also at being

[2] All tools and software, namely *pyrostat*, *java4eurostat*, *happyGISCO* and *eurostat-map.js*, are available on Eurostat domain in *github*: https://github.com/eurostat.

[3] Excerpt from The true concept of Open Government: *"We gave you some of our data (awesome), we want you to do stuff with it (nice, thank you), and hence we now have Open Government (not quite)"*.

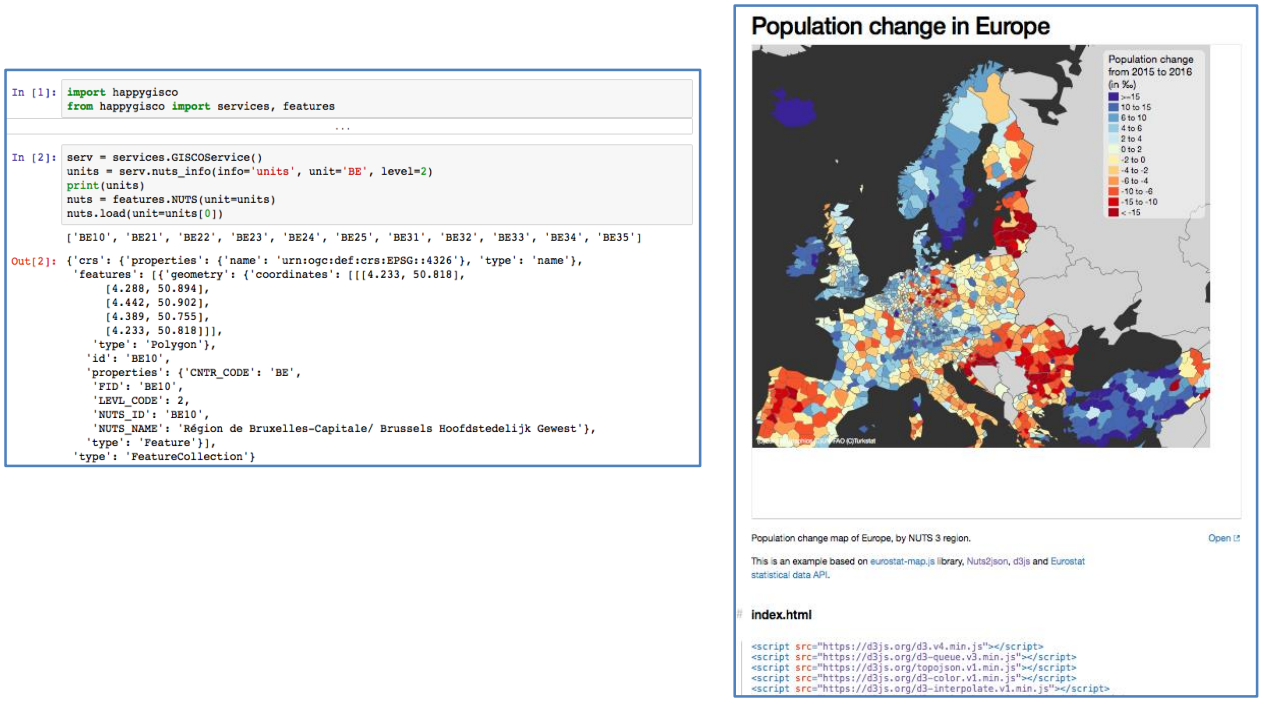

Figure A: Examples of DIY services for Eurostat open data. Left: *happyGISCO* allows the user to run GISCO web-services and query its API directly from a ready-to-use interactive computing platform (*e.g.*, a *Jupyter* notebook distributed through a *docker* virtualised container); in that case, NUTS datasets are fetched and used as-is into the full-stack operation. **Right:** *eurostat-map.js* provides with a set of reusable cartographic styles for showing genuine Eurostat figures. These styles concern choropleth maps, proportional symbol maps, categorical maps, *etc…*; in this example, a map population change over Europe is generated in a web page and the script to reproduce the map is also provided; note that this library uses the *NUTS2json* tools which provide reusable web-formatted versions of NUTS datasets.

integrated into larger – *Python* or *Java* -based – systems to access and process Eurostat data. The possibility to adjust the current products to more specific needs is offered and encouraged. In a constantly evolving data ecosystem, these tools support new modes of production of e-*Official-Statistics* since perfectly configured and ready-to-use computing environments can be distributed with any newly published *Official Statistics* to the public. This could lead to improved decision-making and service provision in the NSOs, in particular through harnessing information, knowledge, skills and perspectives beyond the ESS community, that are traditionally unavailable.

3.2. Supporting multipliers with visualisation tools

Both *happyGISCO* module and *eurostat-map.js* library – developed in Python and Java respectively – make possible multiple cartographic representations based on Eurostat geographical data (*e.g.*, NUTS areas) on various platforms (*e.g.*, web-based interfaces or interactive notebooks) while covering the most basic and generic need of the target users (*e.g.*, different map projections, different tiling basemaps, different map scales, different NUTS levels, different map feature types, *etc*...). For instance, *eurostat-map.js* enables users to quickly create and customise interactive web maps displaying *Eurostat* data directly – retrieved *on-the-fly* from *Eurostat* online database – with various predefined cartographic styles. It does not only support the user in the map design process, it also provides with a more secure design frame, where major pitfalls of cartographic design are avoided. Besides, it offers the web-cartographer the possibility to adopt innovative cartographic representations that go far beyond what has been shown on maps so far. Through supporting the dissemination of Eurostat data and products, these tools further

encourage multipliers (*e.g.*, data journalists, citizen statisticians, *etc*…) to use and share *Official Statistics*.

4. CONCLUSION

The proposed approach – hereby illustrated with software to access and process Eurostat online database – represents a shift in how *Open Data*-driven innovation can be fostered in NSOs by putting a large emphasis on the engagement of external actors, *e.g.*, *produsers*, in the statistical production process[4]. Driven by the opening and sharing of all assets – making not only the data, but also the methods, tools, software and services as well as the decisions open – it enables collaboration and increases bottom-up, participative forms of statistical service design, production and dissemination. At the same time, engaging actors such as scientists, citizens, and businesses in the co-creation of statistical products promote the openness and transparency of the NSOs, and increase the perceived legitimacy of the decision-making.

Because the engagement of *produsers* is also seen as a way of improving, in the future, efficiency and effectiveness in NSOs by aligning services to actual users' needs and interests, one major challenge regards the current answer to immediate needs and legacy issues, as well as long-term problems and potential future requirements for statistical production. In this context, a new assignment for the NSOs may consist in further developing, *i.e.*, populating, and supporting, *e.g.* funding, the – open, modular, scalable and interoperable – software ecosystem around open data and services, so as to accelerate their adoption, and orchestrate the life of the major components of this ecosystem, depending on the level of usage and the evolving requirements of the users.

REFERENCES

[1] European Commission (2013): Powering European public sector innovation: Towards a new architecture, doi: 10.2777/51054.

[2] European Statistical System Committee (2018): Bucharest memorandum on Official Statistics in a datafied society (Trusted Smart Statistics), report of *DGINS Conference*.

[3] OECD (2018): Open Government Data report – Enhancing policy maturity for sustainable impact, doi: 10.1787/9789264305847-en.

[4] Kalampokis E. *et al.* (2016): Open Statistics: The rise of a new era for Open Data?, doi: 10.1007/978-3-319-44421-5_3.

[5] Pollitt C., Bouckaert G. & Loeffler E.(2007): Making quality sustainable: Co-design, co-decide, coproduce, co-evaluate, report of the *Quality conference*.

[6] Nambisan S. & Nambisan P. (2013): Engaging citizens in co-creation in public services: Lessons learned and best practices, IBM Center for *The Business of Government*.

[7] Grazzini J. *et al.* (2018): *"Show me your code, and then I will trust your figures"*: towards software-agnostic open algorithms in statistical production, in Proc. *Quality* conference.

[8] Wijnhoven A.B. *et al.* (2015): Open government objectives and participation motivations, doi:10.1016/j.giq.2014.10.002.

[4] More generally, it represents a much-needed upgrade to the current understanding of public service creation. It can result in more user-friendly and effective public services, improve the quality of decision-making, promote greater trust in public institutions and thus enhance public value [1].