

# Data harmonisation as a bottleneck to better integration in the Balkans

Mátyás Jaschitz<sup>a</sup>, Viktória Jánosi<sup>b</sup>, Olivér Kriska<sup>c</sup>

---

## Summary

There is no denying the importance of consistent, dependable data and geographical evidence for any type of monitoring, policy planning, or situation analysis. In transnational scenarios, this is even more valid. Even so, there are still serious gaps in the EU's framework when it comes to the harmonisation of cross-border data, so it is not shocking that issues are also common among the Balkan nations. In this paper, the authors aim to provide a short overview of the data harmonisation situation throughout the Balkans. Following that, the study will focus more on some representative indicators from the perspective of green and digital transitions and cohesion. The Central European Service for Cross-Border Initiatives (CESCI) planning and policy-making inventory is the primary source of some of the potential solutions for a better data harmonisation process that are provided in the final paragraph. These solutions have already been tested and proven effective. The paper ends with a brief related policy recommendation that includes suggestions for territorial governance.

**Keywords:** Western Balkans; integration in the Balkans; data harmonisation; territorial monitoring; territorial planning

## Contact

<sup>a</sup> [matyas.jaschitz@cesci-net.eu](mailto:matyas.jaschitz@cesci-net.eu)  
Central European Service for Cross-Border Initiatives (CESCI), Budapest, Hungary

<sup>b</sup> [viktoria.janos@cesci-net.eu](mailto:viktoria.janos@cesci-net.eu)  
Central European Service for Cross-Border Initiatives (CESCI), Budapest, Hungary

<sup>c</sup> [kriska.oliver@szte.hu](mailto:kriska.oliver@szte.hu)  
Central European Service for Cross-Border Initiatives (CESCI), Budapest, Hungary

## Mapping the issue

There may be several issues when comparing some data sets between different nations and regions. These challenges are not unique to the researchers who plan to analyse such data; they are also faced by national and international organisations and institutions that work to harmonise data and establish mutual links. The issue of data harmonisation may also be described as multidimensional, as it has multiple approaches and facets (Talhofer et al., 2013).

One of the most important facets, which is also the root of the issue, is the disparate territorial level of administration. The administrative division of countries of the West Balkans (on which the statistical data publication is often based) bears similarities, as it was already partially mentioned above. Meanwhile, it is possible to find slightly different structures in the territorial divisions that line up with the NUTS system. In the case of Montenegro, the NUTS 3 statistical units represent the entire nation; in Serbia, however, the regions that meet the NUTS 3 level lack legitimate local governments (Marjanović et al., 2021). In this respect, Bosnia and Herzegovina is the most problematic. It has not yet established the territorial division that corresponds with the NUTS system, despite being a potential candidate country. Further issues are created by the fact that the country is divided into 3 entities because, while the Republic of Serbia is divided into two

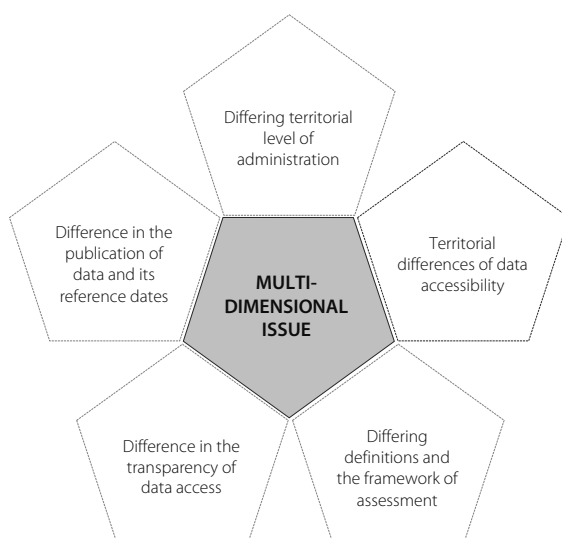
territorial levels (municipalities and the Republic), the Federation of Bosnia includes an in-between administrative level, the cantons. Political reasons also further the issue of NUTS harmonisation, although some possible solutions have already been proposed (Mutabdžija, 2018).

The issue of data harmonisation might also be approached with the help of an XY function graph, where Y, the vertical axis, indicates the territorial differences in data accessibility. The fact that different countries have different meanings for the territorial division—for example, in some countries the lowest territorial level (LAU) denotes separate municipalities (therefore municipal level data is available), while in other countries it denotes a community of municipalities—could lead to frictions. Additional disparities may manifest on the vertical axis merely because one nation provides a specific set of data or set of data concerning LAU territorial units, whereas another only provides this at the NUTS 3 level. The horizontal axis, denoted by X, shows the actual data. It is a common problem that the publicly available set of data is different in each country. Additional issues are brought about by the disparate definitions and the assessment framework for the data and indicators (Burkhauser-Lillard, 2005; Talhofer et al., 2013). For instance, the definition of unemployment may vary depending on which nation-state compares the number of job seekers to which demographic group. For instance, the Republic of Serbia's (RBS) Statistical Office publishes data on the unemployment rate related to the 15–74 age group, while the Statistical Office of Montenegro (MONSTAT) calculates it related to the age group above 15.

Another common problem is the disparity in the publication of data and/or its reference date. Regarding this, variations in the population data may result even from the fact that certain countries report their data referring to the last day of the year (e.g., December 31, 2021), while others do so referring to the first day of the year (e.g., January 1, 2022). Naturally, if one is cautious, such data can be compared, but managing, juxtaposing, and displaying such data may be difficult due to the variations over the years.

Another issue, albeit minor one given the previous point, is the ease of access, transparency, and format of the data. All of these factors alleviate and contribute to the ease with which

Figure 1. Multi-dimensional issue of data harmonisation (Our own editing)



Source: Authors' own elaboration.

the data can be processed and harmonised across countries. In this sense, variations in the ways that various nations report their own data can be found. Certain data might be available to download in the form of charts, while other data might be published in PDFs or similar formats that are difficult to adapt. Territorial statistical data may have codes or identification numbers (as in the case of Serbia), which facilitates processing and data visualisation on a GIS-based map. In other cases, however, the downloadable chart simply includes the name of the territorial unit (e.g., Montenegro).

Our own research confirms the above-mentioned issues, but other sources also report similar problems (Cierpiak-Wolan, 2018). The project titled 'Border Region Data Collection', which includes several border regions across Europe, has made an effort to use a variety of data sources for its investigation into cross-border labour migration. These sources include statistical offices, mobile networks, the labour force survey (LFS), and their own assessments. Overall, the typical issues with data sets also surfaced, impacting the data's vertical and horizontal characteristics as previously mentioned. Joint, integrative research is made significantly more difficult in some countries because some data is often only available at the nation-state level (LFS) or is not available or does not exist (Cierpiak-Wolan, 2018). We looked at the types of data that are available in the publications and online databases on the

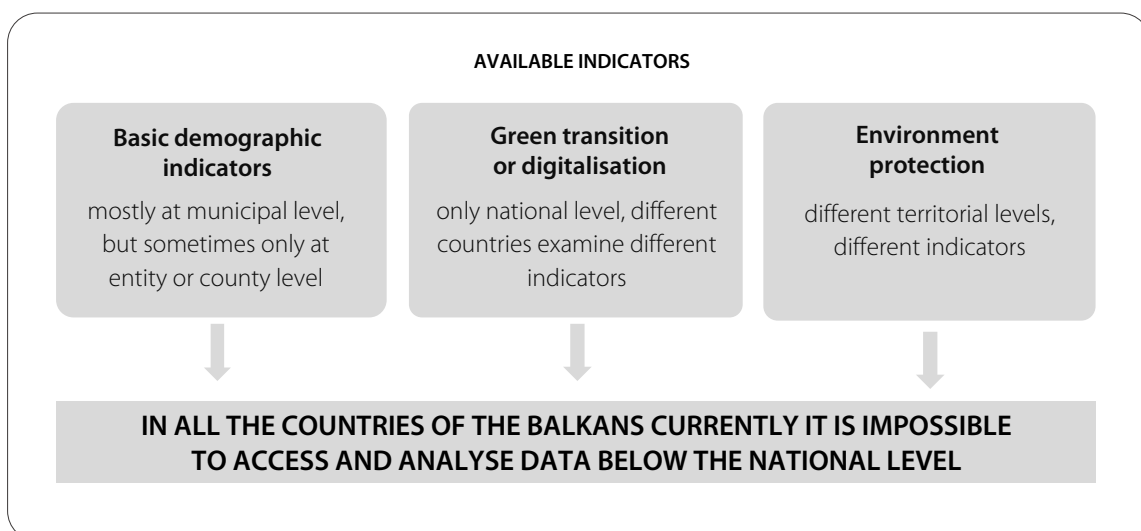
websites of the statistical offices of the Western Balkan countries.

First, we looked at how easily accessible the population, the number of births and deaths, and migration are as basic demographic indicators. There is a great deal of variation in accessibility, even for such fundamental data. Such data is available at the municipal level in North Macedonia, Serbia, Montenegro, Kosovo, and Croatia, but only at the county level in Albania and entity level in Bosnia-Herzegovina. Much of the data for Kosovo and Bosnia-Herzegovina is only available in the native tongue, which makes it challenging to comprehend and compare with data from other countries. Data collection in Bosnia-Herzegovina is challenging due to the unreliable and frequently unavailable statistical office website.

We frequently come to a standstill when attempting to analyse data on a more focused topic, like digitalization or the green transition. European or global databases contain some national-level data on this subject, but not enough to perform macro-regional analyses.

We can also find data on the websites of national statistical offices, but the overwhelming majority of the data found here is at the national level. Some countries have access to more data than European databases do but comparing them can be demanding because different countries look at different metrics related to the same topic.

Figure 2. Available indicators in the Balkan



Source: Authors' own elaboration.

We scrutinized the accessibility of data on environment and environmental protection at the county level in the WB countries. Data on basic environmental indicators such as temperature, precipitation, and air pollution are also available at a lower level than the national level for North Macedonia and Montenegro. For Albania, waste management data is available, and for Croatia, county-level data on environmental protection expenditures and revenues is available. In Kosovo, waste and water management and water contamination data are available, but only at the national level. Serbia and Bosnia-Herzegovina also have comprehensive data on this topic, but only at the national level.

It is clear that the data that is currently available is inconsistent. While we can locate comparable data for two to three, or even four countries, it is currently not feasible to do so at a level smaller than the national level throughout the entire region.

### Attempts at solutions

Several examples of successful and functional data harmonisation are known across Europe. One, or rather two, good examples are harmonised data submissions of statistics and mapping. The Nordic Statistics mostly report multi-faceted data at a nation-state level, from basic demographic data to health services and education to the changing of price levels and biodiversity. The Nordic Council of Ministers provides funding for the Statisticon, a separate organisation that is in charge of maintaining and coordinating the statistical database with assistance from the national statistical offices. The national statistical offices, Eurostat, the OECD, the UN, and other organisations that deal with statistics are the database's sources of statistical data<sup>1</sup>.

The Nordic states also produced Nordmap, an interactive map service. In this instance, Nordregio, which directly reports to the Nordic Council of Ministers, is in charge of data harmonisation. The service is primarily available at the lowest administrative and statistical levels (LAU levels) and provides economic, labour market, and demographic data<sup>2</sup>.

An additional illustration would be the GIS map service stretching across the Grande Region/Großregion and attempts to provide the general public with access to the region's interactive, thematic maps, data, and geoportal. The

website's coordination and content are under the jurisdiction of Luxembourg's Ministry of Energy and Spatial Planning. There are different ways to access data: either at the territorial level of the LAU or at some aggregated levels (like the level of cantons or other municipal communities)<sup>3</sup>.

Both EU members and prospective members must adjust to the new "nomenclature of territorial units for statistics (NUTS)." The NUTS system allows member states to harmonise their own administrative territorial division by providing the minimum and maximum population numbers of a territorial unit. Since the goal of the system is to provide data in a harmonised manner, which is best accomplished by adhering to already existing territorial divisions, member states and candidate states typically compare an administrative territorial division that already exists to the NUTS.

At the same time, variations between nations can be found. In the EU, territorial units at the local level that are smaller than the NUTS 3—the smallest NUTS system—are referred to as "local administrative units" (LAUs). These are the smallest and most inferior territorial units under a municipal system found in every nation.

Because it is the unit where the relationship between data and population is "tightest," this territorial level is crucial. The more comprehensive the collection of local data, the more effective the strategic intervention and supporting evidence can be. Local data is particularly significant when it comes to evaluations of impact and efficiency. However, depending on how each nation defines this territorial unit (LAU) and what degree of spatiality it has, there may be significant differences between them. This could refer to a single municipality in some nations (like Hungary), but it could also refer to a collection of municipalities with distinct names in other nations (like Serbia's "opština" or Romania's "comună"). When considering the matter from the standpoint of the West Balkans, it can be stated that the nations have similar approaches to the LAU territorial units, i.e., a community with municipal rights is made up of multiple municipalities.

A general but excellent example of statistical data harmonisation between countries is Eurostat, which is the statistical office of the EU. Its principal duty is to generate and disseminate uniform and equivalent statistical data on a European scale;

in order to do so, it is also required to harmonise terminology, procedures, and frameworks.

The national statistical offices themselves are in charge of gathering data, and the data that the countries send to Eurostat must go through a rigorous verification process that determines whether the data was actually collected using the standardised, specified methods, whether it satisfies the requirements, and ultimately whether it can be considered reliable and comparable. From the collected, validated, and arranged data on the EU and the euro zone, Eurostat generates aggregated data. Along with the raw data, this is also available on their website<sup>4</sup>.

Eurostat is the office of Europe that has the largest harmonised database, but since it operates within the EU, it cannot offer a comprehensive solution to the data harmonisation issue in the West Balkans. Albania, North Macedonia, Montenegro, and Serbia are candidate members of the region, while Bosnia-Herzegovina and Kosovo are only potential candidates.

A data harmonisation process has been initiated in these countries due to their candidate membership and potential candidate status; however, it is still in its early stages with respect to both thematic and territorial aspects. Only a few indicators, and only at the national level, are accessible to non-members of Eurostat's official database.

However, because this approach completely misrepresents intraregional or urban-rural differences, the Eurostat data structure (mainly because of the typical national and NUTS 2 databases broken down by region) can only analyse macroregional processes restrictedly.

The INSPIRE directive is a legislative document that contains the basic approaches and rules for the establishment of the GIS infrastructure of the EU. Its objectives are to guarantee data accessibility, exchange member state harmonised data, and supply the required technological foundation (Hintz, 2012; Villa et al., 2012). It arose from the need to address cross-border environmental issues. In addition, the INSPIRE directive is essential to the digital and green transitions.

For the past 20 years, the ESPON has been working on building a knowledge base about European territorial dynamics. The ESPON 2020 cooperation programme aims to maintain the availability of European-level, comparable, systematic, and

reliable territorial evidence, to ensure that the requisite policy knowledge is applied, and, by doing so, to enhance the effectiveness of the EU's cohesion politics. There are numerous approaches to this, but we will only cover a few here, particularly with regard to statistical data and the instruments used to harmonise it. In addition to the online resources, the ESPON website offers a number of publications on this subject<sup>5</sup>.

The ESPON 2020 Database Portal provides harmonised and accurate data in the form of an online database, which allows decision-makers, experts, researchers, and other parties concerned to study several territorial issues. This portal gathers, manages, and publishes international statistical and GIS data from ESPON projects and from other regional databases like Eurostat, in order to provide users with easy access to global statistical and geographic information in one location<sup>6</sup>.

Another excellent example is the European and Macro-regional Territorial Monitoring Tool, or MRS. ESPON, for short, is a tool that aims to provide a stable online platform that is easy to use to continuously observe the territorial development trends and models in Europe and its macro-regions. The data integrated into the tool can be combined into five geographic modules: Europe, the Baltic region, the Danube region, the Adriatic and Ionian region, and the Alpine region<sup>7</sup>.

In February 2019, CESC started the elaboration of the territorial analysis of the transnational programme encompassing 14 countries and the Danube Region. This task proved to be greatly instructive, which might be fundamental in an initiative that aims to harmonise the data between the countries of the Balkans.

During the creation of the database that would substantiate the elaboration process, the main challenge was to harmonise the data from the 14 countries. At a national level, highly standardized and harmonised data is available for download. Our primary sources of information were the databases of Eurostat and the World Bank. The national-level data, however, were far too general and inaccurately depicted the variations in an indicator across a nation or macro-region. Since it was crucial to have the most detailed data possible for the elaboration—and this is also true in the context of the Balkans—we also analysed data at the regional and NUTS 3 levels. It would have been ideal in some cases to analyse

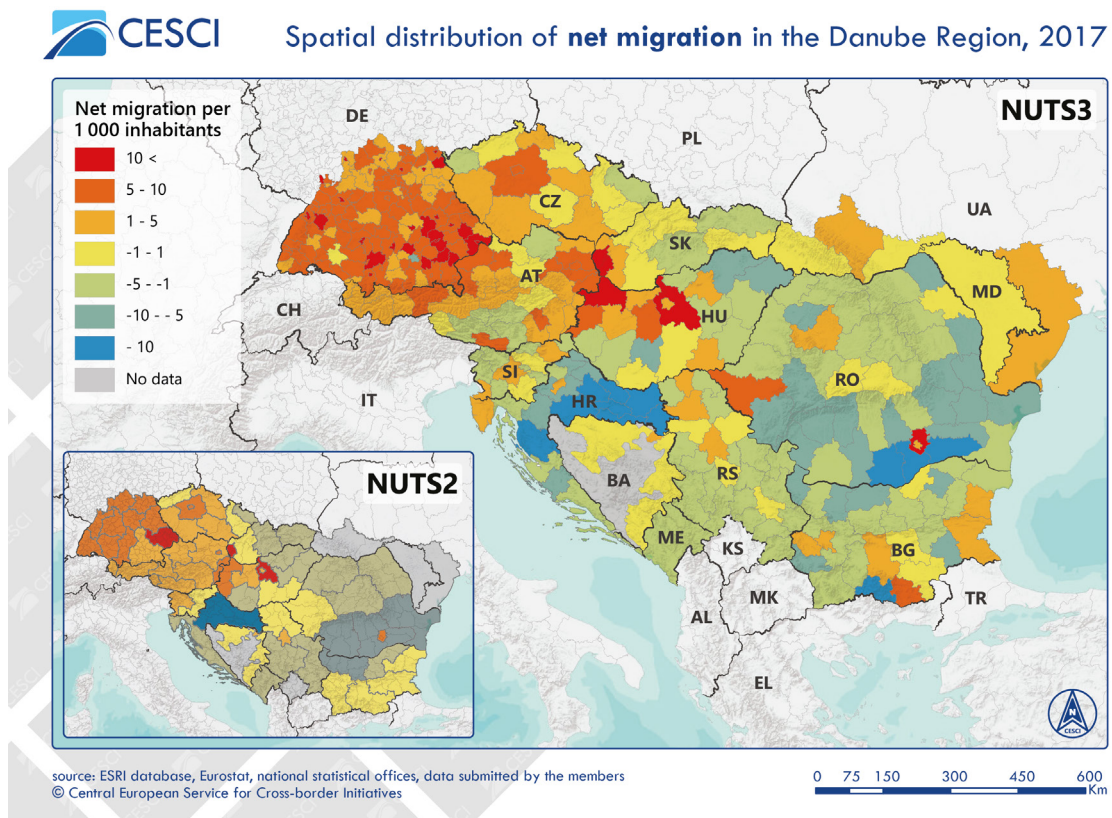
the indicators at a lower territorial level, like the district or municipal levels, but there are no comparable databases that were made with the same methodology.

Nonetheless, the fact that nine of the analysed nations are EU members meant that harmonised data at the county and regional levels could also be found on the Eurostat website, which was a huge assistance. It should be highlighted, though, that even the regional-level data was far too general, and as the figure below illustrates, this can lead to the significant county-level data being misrepresented (Figure 1).

The actual difficulty lay in adding the data from the five non-EU members—Ukraine, Moldova, Serbia, Bosnia-Herzegovina, and Montenegro—to the data that had been downloaded from Eurostat. The databases available on the websites of each nation's statistical office and statistical publications were the main sources from which this type of data was taken. Another problem was that in many cases, even though a particular

indicator could have been found, the approach or time frame used to collect the data differed. Mostly because of our methodology, we collected all the comparable and easily accessible data sources and found every systemic gap (indicator, nation, year). Upon request, the missing data was sent to us by each country's competent bodies, through the programme's Task Force and with the help of the MA/JS, who were able to catalyse the competent bodies of each country on the official channels with regard to the specific indicators, which helped us establish a more complete database. In total, 631 indicators were requested in this manner; offices from 7 countries (BA, CZ, HU, SK, SI, RO, HR) provided us with 100% of the indicators, but the other countries were also able to provide us with nearly all of the data. We created the greatest Danube regional database ever during this process, and the maps we created had very little inconsistency and no blank spaces. We consulted with experts and stakeholders as the territorial strategy was being implemented. We now had appropriate supplements for the primarily quantitative analysis thanks to the

Figure 2. Available indicators in the Balkan



Source: Authors' own elaboration.

qualitative data we acquired. In the end, the language of the programme priorities and particular objectives was created by synthesising the qualitative and quantitative evidence<sup>8</sup>.

The analysed region possesses considerable experience and has initiated efforts to synchronise and align the disparate territorial information (Trkulja-Dabović, 2021). In the Western Balkans, efforts to harmonise each nation's spatial data infrastructure began in 2008, and since then, a number of projects that were organised from outside the region have been carried out. The aforementioned projects, namely INSPIRATION, IMPULS, and MATRA, are designed to accelerate the implementation of the INSPIRE directive (Trkulja-Dabović, 2021).

The INSPIRATION project, which took place in the Western Balkans during 2012 and 2013, aimed to revise spatial data infrastructure legislation, aid in its implementation, enhance education and training, and raise awareness about NSDI/INSPIRE. It is unclear, though, whether the primary goal of coordinating SDI implementation with the INSPIRE Directive was accomplished because the project website does not provide updates on results or impacts. The benefits mentioned are rather general and don't specify any tangible changes in either the legal framework or technical infrastructure for SDI in the region. This information gap may have resulted from the project's two-year duration, a lack of funding, or even from institutional, political, or technological obstacles that prevented the project's successful execution in the beneficiary countries.

The IMPULS project, running from 2014 to 2018, aimed to advance national spatial data infrastructure (NSDI) by developing legislation and creating national catalogues and services. Coordinated by Sweden's Lantmäteriet and the Croatian Geodetic Institute and funded by the Swedish International Development Cooperation Agency, the project was timely. By the end of the 2010s, participating countries had updated their NSDI laws to align with the EU's INSPIRE Directive. The project achieved several milestones: it improved legal and institutional frameworks in line with INSPIRE, boosted technical capacities for NSDI data management, and increased public and sectoral demand for NSDI data. Regional cooperation was also strengthened. However, the evaluation report highlighted the need for more political support, sustainability mechanisms, user-oriented approaches, and data harmonisation at regional and European levels (Naik, 2019).

In most cases, the coordination and harmonisation of the territorial data of the Western Balkans were initiated and overseen by outside parties; there are only a few internal, own-initiation cooperations (Trkulja – Dabović, 2021), which, as a matter of fact, would be essential for a successful cooperation and for an adaptation to each other. The Territorial Governance in the Western Balkans (TG-WeB) platform may also help with this (Trkulja-Dabović, 2021).

### Contours of a policy recommendation

As the evaluation above illustrates, despite many significant but incomplete accomplishments, the implemented territorial harmonisation among the Balkan countries is still in its infancy with respect to the levels that permit macro-regional or cross-border regional analyses.

This type of multilateral data harmonisation basically stems from two problems in the Balkans:

1. The fact that the countries are at varying stages of European integration and the harmonisation of laws and procedures is at varying stages regarding depth and progress, which foreshadows further differences, further challenges the prevalent or inherited differences in data gathering.
2. Based on the results obtained thus far, it appears that the harmonisation efforts are not typically local, "initialised internally," or based on shared interests. Rather, they were initiated in response to external incentives, primarily ad hoc project grants from the EU, which left their accomplishments non-systematic and non-durable and had little to no effect on national institutions, authorities, and legislation.

Naturally, the Balkan data harmonisation efforts that have been completed thus far are helpful, crucial in light of the PILOT nature, and necessary. However, given the foregoing, we also think that it would be unrealistic to expect these ad hoc projects to resolve the systemic problems; regrettably, we also think that it would be unrealistic to anticipate that the unified EU accession of the Balkan states would resolve this issue.

In place of and in addition to these, the Balkans require a "internal," "own," and "individual" data harmonisation platform in order to maximise the above-mentioned benefits. This platform

would only be up to par if it could permanently integrate all competent (in terms of governance) national institutional levels, from the competent ministries to the national statistical offices and the authorities in charge of managing each type of data, by acknowledging the shared interests of the region, organising themselves in a multilateral manner, and operating without outside incentives.

In the absence of consistently valid community regulation, the primary goal of a platform based on multi-level governance may be to bridge the anomalies resulting from EU accession processes at different speeds.

Based on shared interests and equipped with adequate expertise, this platform could allow Balkan countries (regardless of stage in the accession process) to logically and gradually incorporate EU (Eurostat) data collection methods, protocols, and rules into their own systems.

As soon as the platform's policy-level results are developed into data series that can be used in practise, the development of an online "Balkan Eurostat" interface, along with data series, textual and/or diagram or map-type analyses, and reports, can be considered.

The establishment of such a platform implies a high-level and unified political will from each country. However, the Balkans are among the world's most densely bordered regions, and long-term success is impossible without the cooperation and shared intelligence of the smaller states. It is impossible to carry out a more significant national territorial development intervention in this region without causing cross-border external effects, even unintentionally. It makes sense that these forces would be advantageous if they reinforced one another rather than cancelled each other out. But without implementing fundamental data harmonisation, this is not feasible.

## References

- Burkhauser, R. – Lillard, D., 2005. The contribution and potential of data harmonization for cross-national comparative research. *Journal of Comparative Policy Analysis: Research and Practice*, 7:4, pp. 313-330
- Cierpiak-Wolan, M., 2018. Towards coherent statistical systems of transborder areas – Case study on internal and external borders of European Union. *Studia i Materiały. Miscellanea Oeconomicae*, 19:4, pp. 25-35.
- Hintz, D. 2012. Data Harmonization Principles and Development Approaches as Applied to INSPIRE SDIs. INSPIRE-GMES Information Brochure. Technische Universität München, 8 p.
- Marjanović, M. – Ristić, D. – Miličević, M., 2021. New European Regionalism: Regional Planning in the Ex-Yugoslavian Countries. – In Berisha, E., (ed) et al., 2021. *Governing Territorial Development in the Western Balkans*, Springer, pp 43-66.
- Mutabdžija, G., 2018. CONCEPT AND IMPORTANCE OF STATISTICAL REGIONS ON THE EXAMPLE OF BOSNIA AND HERZEGOVINA. *Journal of the Geographical Institute Jovan Cvijic SASA* 68:2, pp. 265-280.
- Naik, A., (2019): End Evaluation of Regional Project Impuls (Geo-Spatial Data Infrastructure). SIDA, [https://cdn.sida.se/publications/files/sida62270en-end-evaluation-of-regional-project-impuls-geo-spatial-data-infrastructure.pdf], pp. 116.
- Talhofer, V. – Hoskova-Mayerova, S. – Rybansky, M. – Hofgmann, A., 2013. Harmonization of Geospatial Data. *AWERProcedia Information Technology & Computer Science*, [http://archives.un-pub.eu/index.php/P-ITCS/article/view/1807/1597], pp 541-546.



Trkulja, S. – Dabović, T., 2021. Supranational Frameworks for Territorial Governance and Spatial Planning in the Western Balkans. In Berisha, E., (ed) et al., 2021. *Governing Territorial Development in the Western Balkans*, Springer, pp 265.287.

Villa, P. – Molina, R. – Gomasasca, M. A., 2012. Data Harmonisation in the Context of the European Spatial Data Infrastructure: The HUMBOLDT Project Framework and Scenarios. *Earth Observation of Global Changes (EOGC)*, pp. 179–191.

## Notes

- <sup>1</sup> Nordic Statistics. Available at: <https://www.nordicstatistics.org/>
- <sup>2</sup> Nordmap. Available at: <http://nordmap.se/>
- <sup>3</sup> SIG-GIS. Available at: <https://www.sig-gr.eu/de.html>
- <sup>4</sup> Eurostat: Data validation. Available at: <https://ec.europa.eu/eurostat/web/main/data/data-validation> (Accessed: 12 July 2022)
- <sup>5</sup> ESPON 2020 Cooperation Programme. Available at: <https://www.espon.eu/programme/espon/espon-2020-cooperation-programme> (Accessed: 12 July 2022)
- <sup>6</sup> ESPON DATABASE. Available at : <https://www.espon.eu/tools-maps/espon-database> (Accessed: 12 July 2022)
- <sup>7</sup> MRS. ESPON TOOL. Available at: <https://www.espon.eu/tools-maps/mrs-espon> (Accessed: 12 July 2022)
- <sup>8</sup> DTP: Programming process. Available at: <https://www.interreg-danube.eu/about-dtp/new-funding-2021-2027/programming-process> (Accessed: 18 July 2022)

---

## Disclosure statement

No potential conflict of interest was reported by the author.