Economic Commission for Europe
Conference of European Statisticians
67th plenary session
Paris, 26-28 June 2019
Item 2 (a) of the provisional agenda
New data sources – accessibility and use
Session 1: Accessing new data sources

A strategy to improve administrative data quality in Albania

Note by the National Institute of Statistics of Albania

Summary

Administrative data are becoming an important source of information for official statistics to lower costs and response burden. Coordinated efforts from administrative data owners and national statistical offices are a must to assess the quality of administrative data and improve it based on standard frameworks, definitions and methodologies.

This paper focuses on the quality dimensions for administrative data. A list of hyper-dimensions will be suggested to be used by the public institutions to measure the quality of administrative data sources. The problems, quality issues and challenges will be explained as well as a possible contribution of the National Institute of Statistics of Albania in improving quality of these data. A roadmap for quality assurance for administrative data will be proposed.

To improve the quality of administrative data the public institutions shall use the framework and roadmap proposed. To have an effective and coordinated approach the proposed roadmap shall be integrated in the Statistical Literacy Strategy as well as in the third National Strategy for Development and Integration 2020-2025.

This document is presented to the 2019 Conference of European Statisticians seminar on “New data sources – accessibility and use”, session 1 “Accessing new data sources” for discussion.
I. Introduction

1. In the times where technology is having a major impact in everyone’s everyday life, the role and volume of data produced is huge. It is up to public and private organizations to leverage the use of the data produced for data-based decision making. To keep pace with time public institutions and governments need to have a data centric approach to adapt to the rapidly changing policies and data landscape. New types of skills are needed and new data management strategies need to be implemented.

2. The Albanian Government has taken a strong commitment in promoting digital initiatives as a powerful tool that leads to modernized governance, increased know-how for a more open society, and a sustainable economic development for its citizens.

3. Albania’s cross-cutting strategy “Digital Agenda of Albania 2015-2020” aims at a society based on knowledge and information, through the consolidation of digital infrastructure in the whole territory of the Republic of Albania; improvement of the quality of online public services and increase of transparency of governance1. In this framework three main directions have been established: the creation and promotion of electronic services for citizens, businesses and administration; the use of ICT for the purposes of education, and the consolidation of the digital infrastructure in the whole territory of the Republic of Albania.

4. On the other hand, in the Development Strategy of INSTAT 2017-20302 the second objective aims to use appropriate data sources and instruments to improve the production and quality of indicators and statistical services. More specifically sub-objective 2.1 is focused on increasing the use of administrative data and administrative records for production of official statistics. Since various administrative systems generate a vast amount of data, INSTAT will be focused on increasing the use of administrative data for statistical production by lowering costs and burden. To achieve this objective, a continuous coordination is needed with administrative data providers, taking into consideration the differences in concepts, definitions and methodologies, timeliness, accuracy of specific variables, formats and identifiers.

5. Information derived from data is crucial for the decision making process. To produce it in a more efficient way, existing data sources must be used as much as possible and the new systems built must be designed taking into consideration also the need for statistics production and indicators. A major coordination is needed as well among different donors and international organizations assisting in the designing and creation of new systems and electronic public services. The public institutions have the responsibility to guarantee that the data collected or generated through different systems are used as much as possible to generate information to support decision making and the public good, taking into consideration the protection of personal data and confidentiality. There are different fields where administrative data is lacking or the data suffers from low quality. These data are often collected or stored not taking into consideration best practices or methodologies making almost impossible their use for producing statistical information.

6. Having a data centric approach and empowering the use of administrative data will lead towards a transformational process of helping the government to make decisions and institutions achieve their objectives in a more effective and efficient way.

---

II. Quality dimension for administrative sources

7. The terminology for quality assessment of administrative data sources has evolved over time. In 2006\(^3\), the authors Batini and Scannapieco, referred to the criteria of the quality of these sources with the term category. While authors Daas et al., 2008b\(^4\), link quality criteria with three hyper-dimensions: source, data and metadata, used by statistical offices or institutions to measure the quality of administrative data sources. Each of these three hyper-dimensions has several dimensions, and the dimensions are measured by some quality indicators. Measurement of indicators is achieved by combining one or more qualitative or quantitative methods.

8. The list of dimensions of each hyper-dimension which will be suggested by INSTAT to be used by the institutions to measure the quality of administrative data sources is given in Figure 1.

Figure 1
Dimensions and hyper-dimensions to measure quality

9. The Source hyper-dimension should provide information about the following dimensions: data provider, contacts with the administrative source, purpose of data collection, possibility of their use to produce official statistics, measurement of the burden of units that provide this information, legal basis, protection of confidentiality, data collection procedures and their transmission according to the criteria set out in the legal basis.

10. Metadata hyper-dimension should provide information about the following dimensions: definitions and classifications used, population and focus units of the collected data, changes in time of the main concepts used. One of the dimensions of this hyper-dimension will be also data check, which will greatly target the quality of the data when they are combined with other administrative data or modifications over the years.

11. To link administrative data each of them shall use same identifiers for the focus units. NIPT is used for businesses in Albania, which is a unique code that identifies the enterprise in several administrative sources such as: National Business Center, Tax Authority, etc. With regard to individuals, they have an identification code, which is also used in some administrative data sources such as: Tax Authority, Social and Health Insurance.

---


12. Data hyper-dimension should provide information regarding the coverage of the administrative data, whether this source has under or over coverage, possibility of linkage to other administrative data, measuring the response at the record level or specific questions, identifying outlier or the imputation of missing values, measuring relative standard error for the main variables, and checks with other administrative sources. This hyper-dimension determines the quality of the administrative source and the changes it has had over the years.

13. This model should be used by statistical offices and institutions by following the logical flow of Source, Metadata and Data without changing their order. If the quality of the administration source is not evaluated the assessment of the other two hyper-dimensions cannot happen.

III. Challenges related to administrative data

14. The work with different administrative data has shown that many problems should be solved to improve the quality of administrative data and use these data for statistical production and decision making:
   • The administrative data providers need to have a better understanding of the importance of data they collect/generate
   • The need for common identifiers among different data sources
   • Standard roadmap need to be developed to guide administrative data providers to ensure quality
   • Quality assurance framework especially for administrative sources is a new area and staff need to improve their knowledge in this area
   • Fear of being assessed in regard to quality
   • Quality assessment by internal or external body.

15. INSTAT can contribute to improve the quality of administrative data through:
   • Organization of meetings, workshops, conferences to raise awareness on the importance of the quality assurance of administrative data
   • Preparation of training curricula to be used for training on quality assurance and quality assessment
   • Trainings in structural and referential metadata
   • Quality assurance framework developed for the entire National Statistical System
   • Development of a roadmap with milestones to achieve quality need to be developed to guide administrative data providers to assure quality
   • Continuous support in implementing quality assurance
   • Definition of common vocabulary for most important variables.

IV. The roadmap for quality assurance for administrative data

16. To assure quality of administrative data the roadmap in Figure 2 is being proposed. This roadmap is in line with INSTAT objectives towards increasing the use of administrative data for statistical production and will be a separate component of INSTAT Statistical Literacy Strategy. It will be proposed to be integrated in the third National Strategy for Development and Integration 2020-2025, to be coordinated with government plans in this area.
A. Quality Framework and standards

17. A list of indicators and dimensions to be used by institutions to assess the quality of administrative sources they collect/generate. For each of the dimensions and hyper-dimensions, the state in which they are and the target to be reached, will be evaluated. INSTAT should constantly suggest institutions the standards to be followed in various aspects, to ensure the reliability and the desired level of quality.

B. Set objectives and prepare a work plan

18. Analyzing problems with each of the administrative data sources and current situation with the intended one, is the next step to improve the quality of data sources by preparing a work plan. This work plan should contain improvement actions that must be derived by the proposed standards.

C. Capacity building of the staff

19. To ensure that all measures specified in the action plan are successfully implemented, institutions should give proper training to staff involved in processes that do not function properly. The best way to improve staff skills is to first assess gaps or needs and then build training modules where needed.

D. Evaluation

20. The assessment of progress phase will analyze the state after the improvement actions have been implemented for each of the problems raised in the previous stages. This process will provide more information on the progress made towards achieving the desired target in relation to all the hyper-dimensions defined in the framework. One of the main aspects of this phase will be the improvement plans for the next round of data collection/generation the same administrative source.

E. Adjustments and lessons learned

21. The improvement actions undertaken by the institutions to improve the quality of administrative sources will be included as new practices in their work. This will ensure that all the actors involved in this process will be following the decided steps continuously and not just temporary.

22. As for the lessons learned, these will serve as tools for training other organizations that want to achieve the same level of quality.
V. Conclusions

23. Administrative data is one of the main resources for increasing efficiency and effectiveness on the production of official statistics. It is important to increase awareness of administrative data owners on the importance of the information they have. This shall not be limited to the primary scope for which the administrative data is collected/generated. Concrete actions need to be taken to monitor and improve the quality of administrative data. In this regard a roadmap for quality assurance is being proposed by INSTAT as part of the Statistical Literacy Strategy to be integrated in the third National Strategy for Development and Integration 2020-2025. Coordination and planning are two keywords where NSI and Government need to be focused to reach concrete objectives to improve quality of administrative data, which will have a direct impact in the quality of production of statistical indicators. Capacity building is a very important aspect, new types of skills need to be introduced as well as new concepts and standards.