



Economic Commission for Europe**Conference of European Statisticians****Sixty-ninth plenary session**

Geneva, 23-25 June 2021

Item 2 (b) of the provisional agenda

Post-Covid: sustaining organisational and product innovation in national statistical offices.**Innovation in national statistical offices organization and working arrangements****Covid-19 Operational Data Management – Production model
for new generation statistics*****Prepared by Lithuania***Summary*

The document reviews the experience of Statistics Lithuania (SL) in managing the operational data of the Covid pandemic, which is vital for the pandemic control, monitoring and decision-making with intention to preserve human health and lives. The emergency situation highlighted the importance of having a space in the state where the entire data ecosystem can be governed to meet the information needs, from data collection to outsourcing, and flexible user rights management in order to make decisions “here and now”.

This important task strengthened SL as an institution and increased the efficiency of its activities, which the organization lacked before this challenge. However, although it has had a huge impact on the growth of the reputation and brand of SL, this step represents a long-term strategic change in the main activity of the organization – the production of official statistics. An efficient data management model developed in a short time is being introduced to produce official and experimental statistics. This allows to expect significant changes in the level of detail and relevance of statistics, as well as new quality services for the state, science, business and society. The timely decisions of SL and quick actions to take the operational management of Covid statistics created the preconditions for the implementation of a much higher ambition – to fundamentally change the role of SL and become a state data steward.

This document is presented to the Conference of European Statisticians’ session on “Post-Covid: sustaining organisational and product innovation in national statistical offices - Innovation in national statistical offices organization and working arrangements” for discussion.

* The present document was submitted late due to resources constraints.



I. Introduction: overview of the pre-Covid situation in public data management

1. Public sector data in Lithuania is fully decentralized. This results in a great diversity of administrative sources in technological and functional terms. The high fragmentation of data sources and its negative impact on being able to use the data became apparent with the onset of the Covid pandemic. In the case of Lithuania, collecting and processing operational data was a key problem in pandemic management. And it was the Covid pandemic that inspired decisive and rapid decisions in the field of state data management. SL became a central actor, demonstrating its ambition to take on a fundamentally new role as a data steward.

2. What is the current state data management situation in Lithuania? Officially there are 96 registers and 285 state information systems. These registers and information systems are managed by 139 different managers and operated by 168 main managers. Most systems and registers are based primarily on the internal operational objectives, owners' or managers' financial capacity, competence, etc., and not necessarily taking into account the need of systematic approach to state data, the wider context, users' needs and so on.

3. Thus, the problems from which we started to implement our role as a state data steward are clear:

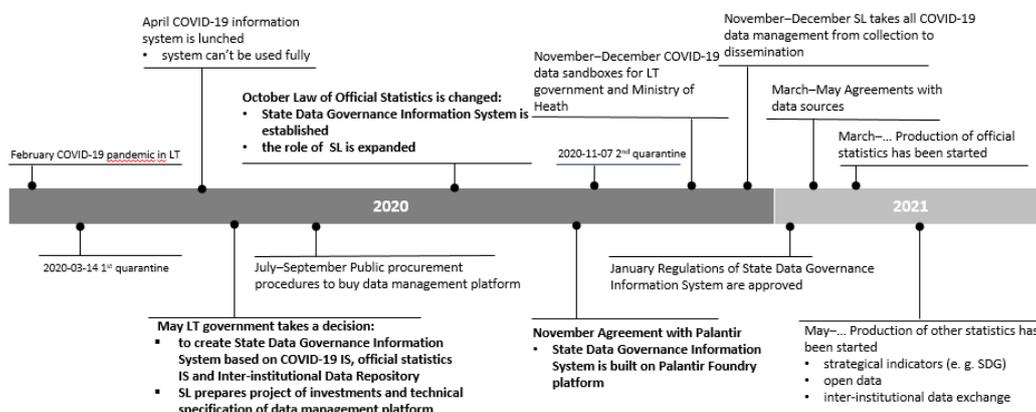
- Isolated, fragmented data do not create much value. Valuable information is obtained by combining data from different areas, analysing, and so on. However, the already mentioned great number of data sources based on a wide range of technological solutions means that public sector data are disabled. This means that they are intended and can be used in practice only for the closed internal purposes of their owners. Most registers and information systems are closed in nature and focused just on their owners' needs. Data opening or re-using processes are not centralized.
- Inflexible and slow official statistics. Official statistics are produced on the basis of the EU and national legislation, in strict accordance with established methodologies and deadlines. This ensures the quality of statistical information. However, nowadays users need timely and granular data more than they need highly accurate data. On average, we prepare annual statistical information in 150 days, quarterly – 58 days, monthly – 24 days, which is not acceptable in the context of pandemic or other possible disasters. Traditional statistical surveys do not allow high efficiency to be achieved and also increase the statistical reporting burden. For these reasons, the use of administrative sources is one of the key solutions. This allows producing "faster" statistics that reflect socio-economic changes in near real time. Therefore, public data becomes more and more important in our production. And here we face with another huge problem: having the right to collect any administrative data for our official statistics purposes, we are unable to unlock their full potential as we do not have the right to use them for other purposes, unlike the owners of these data (with respect to GDPR, of course).
- Ineffective data exchange, re-use and opening initiatives. Both the public sector and business own huge data resources. However, the sharing of these data is complicated due to the already mentioned fragmentation of data sources. Users are obliged to contact individual data owners or managers to obtain data from the public sector, in accordance with the different procedures. Businesses as well are not motivated to share their high-value big data with the state. Existing data opening initiatives do not solve the fragmentation of public sector data, do not increase accessibility. In addition, the opening of data increases the costs for data owners in preparing and providing open data or making it possible to re-use. In addition, the opening of data further increases the costs for data owners in preparing and providing open data, while the value of these open data remains questionable due to quality deficiencies. Fragmented data collection and opening up hinders science and business from using these public information resources to develop products of high economic and social value.

4. The problems mentioned above result in SL initiative to solve them by creating the Data Governance System (or state data lake, better to say), which was launched in November

2020. SL acquired a ready-made data management platform, which allows creating a sustainable and flexible data ecosystem with cloud infrastructure services. As already mentioned, the main impetus for the idea of the Data Governance System was (and still is) the Covid pandemic.

Figure 1

Timeline: from idea to implementation.



5. Milestones of timeline: the Government decision on the establishment of the State Data Governance System, amendment of the Law on Official Statistics and purchase of a data management platform enabling the information system. We are currently in the process of primary data assimilation from data sources and production. That means, we sign agreements with data owners, organize the collection of data from their systems and start the transfer of statistical surveys and the production of statistical information in the State Data Governance System. In parallel, we establish models for ensuring data confidentiality, quality assessment, and service provision. All this is done in the context of an existing Covid data management process. This means that what we aim to achieve on a large scale has already been achieved with Covid data.

II. Legal basis: a timid step forward

6. Under the new provisions of the [Law on Official Statistics](#), SL becomes the owner and manager of the State Data Governance System, the scope of which covers a wider data space and purposes of use than official statistics. Thus, SL becomes like an actor of two roles: SL uses the new system to produce official statistics, and at the same time SL gets the right to collect data and to produce information other than official statistics.

7. It should be noted that we see that the amendment to the law has slightly increased the ability of SL to meet the information needs of the state and society, but not enough, as there are significant contradictions and obstacles in terms of data opening and re-use. The provisions of the Law on Official Statistics allow the use of State Data Governance System resources only for statistical purposes. It means that:

- Data cannot be provided for administrative purposes or it requires long bureaucratic procedures (e.g. signing tripartite agreements between the data owner, SL and the data user), which makes impossible to implement the main idea of the State Data Governance System – fast exchange and provision of data. Real examples: an effective analytical space for research on the black economy conducted by the State Tax Inspectorate could be realized in the State Data Governance System, but without additional legal regulation, the State Data Governance System cannot be used for such a purpose. An analogous situation is with the provision of detailed data to municipalities with intention to help them in controlling and local decision-making on Covid pandemic: technologically, the State Data Governance System can provide operational and detailed (depersonalised or encrypted) data for decisions related with outbreaks' management, preventive measures, logistics of vaccination, etc., but the current legal framework does not allow it.

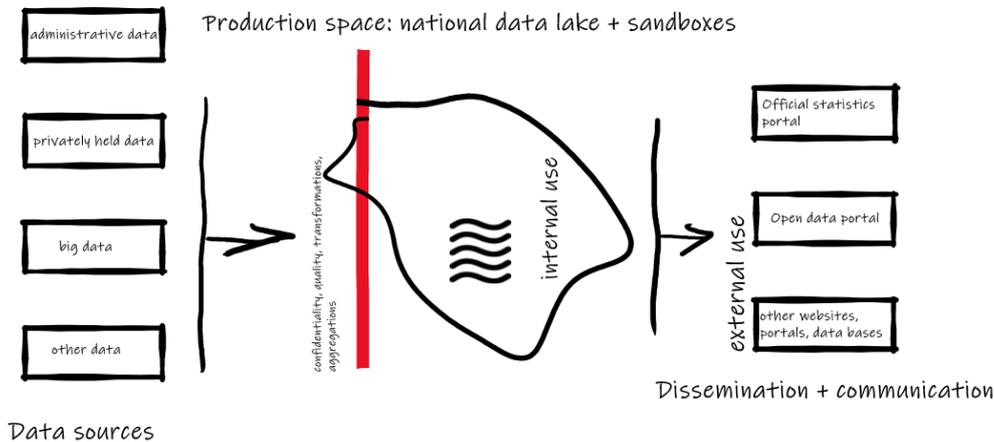
- Centralized data opening, although technologically very fast (e.g. automated Covid data opening on the national Open Data Portal has already been implemented). SL plans to open, in coordination with data owners, all State Data Governance System data resources that are likely to be necessary to the public. This requires primary data, as this is the only way to ensure maximum data enablement. However, there is a risk that, in the absence of a legal obligation, data owners will not be inclined to share the data they own.
- The high fragmentation of public sector data in different systems and registers means that the receipt of data in the State Data Governance System, which is technologically fast to implement, may be artificially delayed by systems or register managers due to lack of interest in providing primary data. This is already evident now that data acquisition contracts are being negotiated for a relatively long time, data is being refused in the most efficient ways, and so on. Another argument is separate national legislation prohibiting the re-use of data for purposes other than those specified.

III. Solution and its implementation

8. What is the formula of the SL solution for becoming main player in state data management? It is very simple – to create a common data space (or national data lake) to consolidate and enable important state data. This means to create a State Data Governance System on the basis of a multifunctional and capable data management platform, which would enable quick and efficient access to different data sources, with intention to provide services for state, science and business. And the role of SL extended by the Law on Official Statistics creates legal preconditions for achieving this goal.

Figure 2

Production space: national data lake and sandboxes



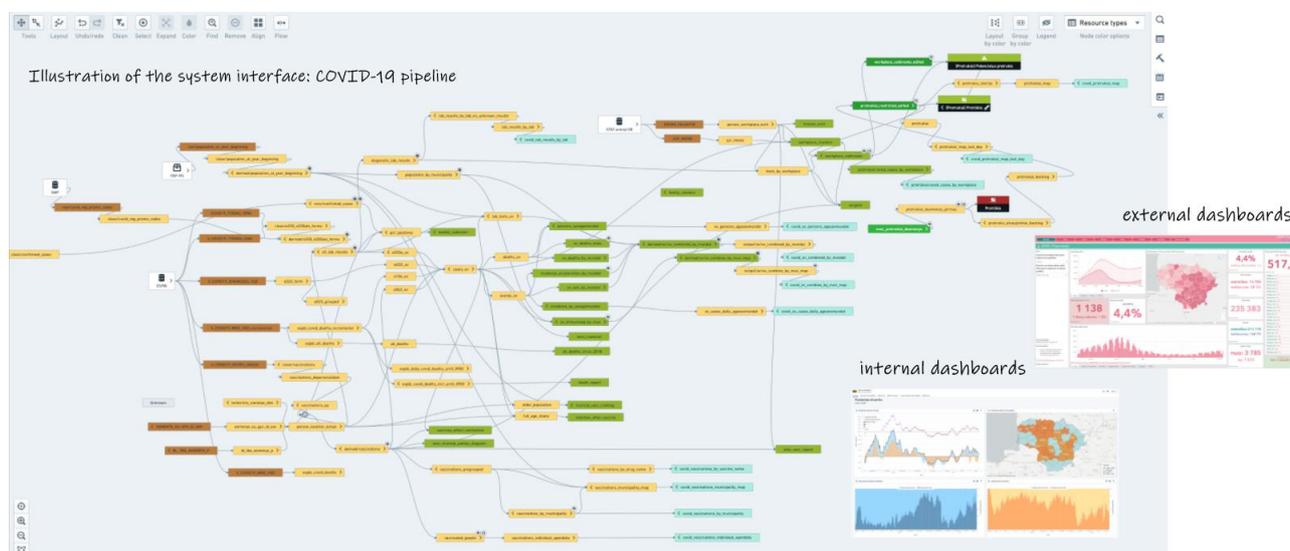
A. Covid data management

9. The main source of operational Covid data in Lithuania is the *e-health* system managed by the Ministry of Health and operated by Centre of Registers, which is mainly intended for the collection and storage of patient data by healthcare institutions and physicians, and for the provision of treatment-related e-services. This means that the system is not suitable for data analysis, so in the face of the Covid pandemic, the data collected in *e-health* system are not possible to be used in practice. Currently, *e-health* system provides daily primary data related with Covid diagnostics, morbidity, mortality, and so on. The following is a brief overview of the main Covid data management processes: morbidity and mortality, vaccination and outbreak management.

- *Morbidity and mortality data management process.* Figure 3 below (and this short [video](#)) presents examples of the system interface: the full path of Covid data on testing, confirmed cases, deaths to the final product, internal and external dashboards

and reports. The additional need for the new indicators is realized without destroying the overall data sequence. For example, if the Ministry of Health or other state body need to submit new reports to international organizations, this is realized in the State Data Governance System very quickly.

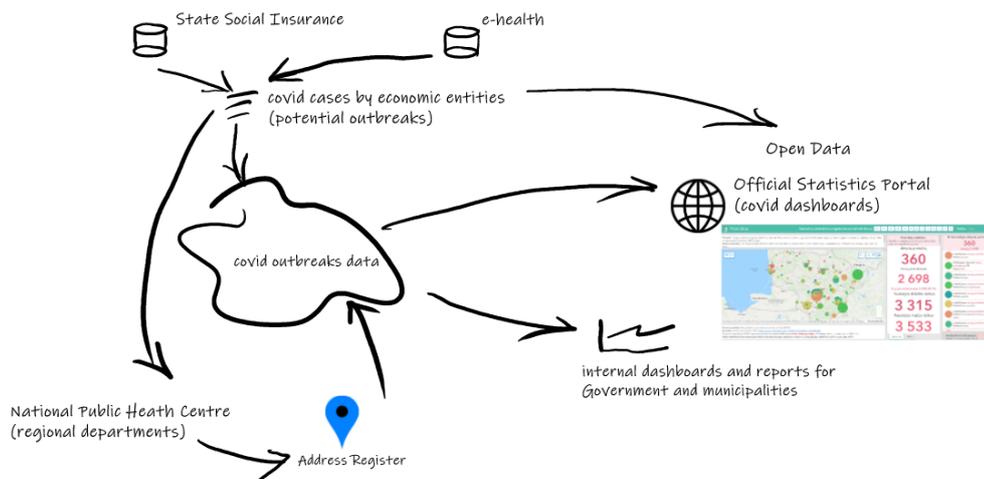
Figure 3
Illustration of the system interface: Covid-19 pipeline, internal dashboard and external dashboard.



- *Implementation of vaccination logistics and data management process.* The vaccination data module in the State Data Governance System includes not only data management but also logistics of vaccines. This is a successful example of collaboration with the Ministry of Health – when, in a sense, the platform managed by SL is “leased” for additional tasks. The Ministry of Health enters information on arrivals of vaccines into the State Data Governance System and these are automatically distributed to the vaccination centres according to the established algorithm. The distribution takes into account the size and composition of the population by age and the priority groups set by the Government.
- *Outbreaks data map.* Covid outbreaks monitoring is one of the most challenging tasks in controlling of pandemic. Prior to automating outbreaks data management in the State Data Governance System, the regional units of the National Public Health Centre used their own internal database, which, like *e-health* system, has no analytical component and any interface with other systems is complicated to implement due to outdated technology. In addition, the Government and the public were not provided with clear information on outbreaks and their management. The outbreak management process differs from others in the way how data is collected – data is uploaded by regional unit staff, and the State Data Governance System allows integrated analysis of the situation and data dissemination.

Figure 4
Operational data of Covid-19 outbreaks

operational data of covid-19 outbreaks



- *Hospitals indicators.* Hospital data collection is organized in a similar way to outbreaks. Regional health facilities provide daily data on the new Covid patients, Covid treatment beds, and the workability of physicians and nurses.

10. All key Covid data “travel” each morning to a [daily report](#), which is a key data source for the media and the public. At the same time, data are updated in the special [Covid dashboards](#) and in the [SL open data](#) and national [Open data portal](#) as well.

B. Preparation for the production of official statistics

11. As can be predicted from the above statements, the objectives of the State Data Governance System are briefly named as follows: data consolidation, information production, data exchange, and analytical space services. The current scope (first half of the 2021) of the State Data Governance System is 28 data owners and 53 of their registers or information systems. This covers almost all the current needs for official statistics, as well as a significant part of the state strategic indicators to be developed. It should be noted that the aim is to consolidate the primary data. This will allow the centralized production of a new generation statistics – much more operational and detailed.

12. Priority tasks in this period: 1) data ingestion, 2) development and application of anonymization / encryption model ensuring data security, 3) development of user rights system ensuring data security, 4) transfer of current algorithms for calculation of official statistics indicators.

IV. Influence on the organization: changes in image and reputation

13. Following the takeover of Covid data management and dissemination by SL, major changes in the SL brand and reputation were noticed. SL measure the value of its output in society by calculating two indices: the usability of statistics and consumer satisfaction. The usability of statistics includes the so-called objective evaluation indicators: usability of the Official Statistics Portal, Database of Indicators and social networks, quotation of SL and its production in the media, individual requests of users. Accessibility of SL brand and production in main national media grew from 466.8 million to 3.8 billion, in most popular social network Facebook – from 420 thousand to 1.0 million, user satisfaction level – from 72.4 to 75.5.

14. In some cases, SL recorded a growth of hundreds of times. Obviously, this was due to the publication of Covid operational data. However, it should be noted that not only Covid

statistics but also statistics in other areas have attracted considerable interest. This means that SL and its activities have become much more visible and well-known in public. It should be noted that Covid also inspired a certain breakthrough in the use of open data. For example, open data of vaccination has been viewed 500,000 times since mid-February in the SL open data portal, and on average open data files receive up to 7,000 views per day.

15. Regarding subtler internal changes in the organization, it is necessary to mention that the emergence of a completely new idea in common has, in a sense, given a “new breath” to SL, a relatively large bureaucratic organization in Lithuania. Along with the State Data Governance System project, almost 30 new highly qualified specialists came to SL and brought a different perspective, work culture and habits. Along with structural changes in the organization, work environment changes during a pandemic as well, this inevitably works and encourages change: a snowball effect seems to be starting to work.

V. Conclusions

16. The main insight of this extremely dynamic time and extraordinary experience is that becoming a state data steward is not a technological problem. This is primarily a legal issue. Statistical authorities need a legal mandate to expand their roles to become data stewards and to empower their own competencies and data. At present, data opening, sharing and re-using are difficult to implement due to legal provisions restricting the activities of statistical authorities. And this is very important: if statistical authorities, as the ones with the greatest competence to manage data, do not expand their role not only in producing official statistics but also in sharing their resources, they are going to become redundant as new players will enter the data market with the Data Act.

VI. Future plans: centralized production of public sector open data

17. The objectives of SL in 2022–2026 are related to the reform of the opening of state data. The aim of the reform is to ensure the availability of reliable public sector data, the possibility to share it between economic sectors, to ensure the possibility of re-use of data and to create preconditions for public sector institutions to make data-based decisions and for businesses to create digital innovations. The funds of The Recovery and Resilience Facility are planned to be used for this purpose.

18. In the context of data reform, the State Data Governance System is perceived as a national data lake, consolidating important state data, which in turn, using standardized anonymization, confidentiality and opening methods, is used for a wide range of information production: open data, official, experimental and operational statistics, strategic state development monitoring, special governmental tasks, sectoral data spaces, research, business innovation.

19. The reform aims to expand the amount of data published in the State Data Governance System by using 376 state information resources, to create preconditions for re-use of data in all state information systems and registers, and to open high-value data to the public, business and science. The fact that the entire data ecosystem is implemented in the State Data Governance System created on the basis of a ready-made (big) data management platform determines that it is able to implement other services in addition to the main purpose: data opening, analytical spaces (sandboxes), big data retrieval from business or other sources and various exchanges with each other and with the outside world.

20. Re-use of data held by public sector bodies creates added value for the benefit of re-users, end-users and society at large. At the same time, in many cases for the benefit of the public sector body itself, as it promotes transparency and accountability and receives feedback from users and end-users, the relevant public sector body is encouraged to ensure the quality of the information collected and the performance of its public tasks.