Towards modern user-oriented dissemination – chances and challenges for statistics

Magdalena Ambroch

Statistics Poland

Abstract

Robust information has become the most desired good in the modern world. In the flood of easily available data, official statistics is still distinguished by high quality, transparency and credibility. However, without a new approach to data dissemination and communication, it may be difficult to maintain the leading position in this area. The presentation will show how Statistics Poland tries to modernise its practice to keep the interest of data users and meet their diverse needs. Statistics Poland provides more and more user-oriented products, like profiled databases (including newly launched Knowledge Databases), digital tools and publications. Our experience shows how important it is to determine a strategic plan for modernising dissemination, including the need to build the understanding and openness to change within the staff.

Keywords: innovation, dissemination, change in the organization

1. Introduction

Polish official statistics enjoys high credibility among the recipients, reinforced by over hundred-year history and wide information offer. Each year, several tens of thousands of variables that characterised a number of research areas are produced from around 260 statistical surveys. In recent years, new information resources based on experimental research methods and non-statistical data sources, supplementary to official statistics, has been introduced. The challenges facing Statistics Poland are related to dissemination and communication of such data in a suitable, legible and attractive way for users.

The needs of our recipients are very diverse, depending on their knowledge, goals for which they search for data, customs and information technology skills. The recipients require both ready-made information and conclusions as well as detailed long time series for their own analyses. Along with technological advances, there has been a growing interest in open data and modern interactive forms of data presentation (infographics and applications). Simultaneously, there has still been a certain demand for traditional publications, especially from persons less familiar with new technologies. Legal regulations concerning, among others, data openness and digital accessibility of information for people with special needs provide additional context for users' needs.

2. User-oriented offer

Statistics Poland has been continuously modernising and diversifying its offer in order to meet all the expectations.

- The main means of data dissemination is databases, which can be used via application programming interface (API):
  - As early as the 90s, the first database with statistics at low levels of territorial division was launched (Local Data Bank). In the following years, further general or particular database products were developed, among others:
  - Knowledge Databases, main source of data for Poland, including High Value Data,
  - systems for specialized analyses, developed on the government and self-government demand: Strateg (Cohesion policy monitoring system), SMUP (Public Services Monitoring System),
  - National Reporting Platform – SDG, for monitoring 2030 Agenda.

  The catalogue of data bases is supplied by products for data visualisation, like Geostatistics Portal presenting spatial statistical data.
- The second pillar of the offer is publication. The emphasis is put on analytical reports, news releases and occasional publications aimed to promote statistics. All publications are available online free of charge.
- The offer is being expanded by some innovative and interactive products, like digital publications (for instance annual SDG reports released since 2020) or application Data for indexation for entrepreneurs (it allows to calculate different variants of the value of the corrective multiplier for different time periods, used for the indexation of quotas resulting from different types of contracts).
- Additional to products presenting official statistics are tools for interactive graphical presentation of experimental data which require a non-standard approach, for example:
  - a tool for presenting data based on satellite remote sensing for SDG indicators monitoring access to roads in rural areas, access to open public areas and management of urban land.
  - TranStat – an application monitoring road and maritime transport using huge data sets (Big Data) obtained in real time.

3. Challenge – need for change

The variety of products offered makes that everyone can find something for themselves. However, users sometimes indicate that they feel confused in this multitude. Managing so many products is also challenging for the institution. The databases offered by Statistics

Poland were created at different times and for different needs. Their technologies and functions were the most up-to-date at the time, but they differ between databases. Products of a smaller scale (digital publications, interactive data presentation applications) are often created as initiatives of small teams who (usually without additional financial resources) experiment with various open-source tools. As a result, data dissemination products are not integrated with each other, and they draw from different sources and are based on different data formats. Data in individual products are updated at different times and using different procedures. This is troublesome for recipients (values in different products can vary) and very burdensome for the staff (having to prepare similar data ranges in different formats to supply multiple products.

4. Towards modern dissemination

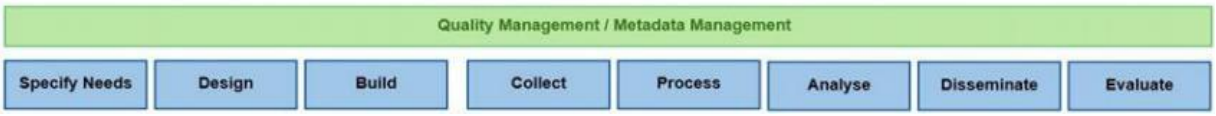
4.1 New information portal

Understanding the difficulties of data users in finding the desired information, the construction of a new, more intuitive, information portal of Statistics Poland has been started. It will have a thematic structure, organising statistical products and services and directing the recipients to portals integrated with the homepage, dedicated, among others, to databases, publications, methodology, surveys (for respondents), innovation in statistics, education and scientific activities. Logged visitors will be able to collect the information and resources they are interested in on their own digital shelf. The new portal will be launched at the end of this year. At the moment recipients can test a beta version (currently available only in Polish – new.stat.gov.pl), and submit their comments via a form. Their insights may be useful for further works as they can be used to modify approach and proposed solutions.

4.2 Modernization of the system

The way statistics are presented and disseminated is of key importance for users. However, what can be seen on the outside, depends on the “kitchen” which is invisible to our recipients. Therefore the entire system of statistical production is undergoing modernization basing on Generic Business Statistical Process Model (GBSPM). The statistical process will be automatized by dedicated systems supporting each phase. In terms of dissemination, such a system is the Publication Data Repository (and Geospatial Data Repository), which contains outcome data, ready for dissemination (the repository can also be used directly by external users via API from this year onwards). At the moment only Knowledge Databases are linked to this Repository, but ultimately, all digital products presenting official statistics will be integrated and based on this common data source**.**

Figure 1: Generic Business Statistical Process Model



This integration is a very complex and long-term project. It requires the reconstruction of existing structures of the individual databases following the logic of metadata applied in the Knowledge Databases, which were made available in 2023 in a new version and in new structures.

Modernized Knowledge Databases have a new interface and functions tailored to the needs of both occasional and expert users. They consist of two parts:

- dashboards which are graphical guides to statistical areas. They are particularly useful for beginners who can find here basic indicators, interpretative tips and information on what kind of data they can look for deeper in the statistical resources;
- database tailored for the needs of more advanced users. Here, long data series are introduced, very detailed in terms of the classifications and code lists used. The user can modify the scope and layout of the reports and, once logged in, has the additional option of saving their searches and summaries.

Both parts are supplemented with a methodological context, which means that detailed metadata can be accessed from anywhere in the Knowledge Databases.

A unique solution introduced in the Knowledge Databases is the ability to combine data from various areas into a single summary table. This is possible thanks to a modern approach to data resources management. Data in the database are based on the concept of a publication variable. The publication variable is the basic unit of information in a database. It consists of the publication name and cross-section with measurement units and observation periods. It means, that each value in the database is described by metadata, so that the control of the database resources and their retrieval from the Publication Data Repository is entirely based on metadata and processed through the Statistical Metadata System. This concept will ultimately be used in other database products, allowing them to be simultaneously supplied with consistent values from the Publication Data Repository. In the future, it is also planned to link digital publications to this mechanism, which will automate the process of preparing them.

5. Dissemination involves everyone

The plans to integrate all dissemination products are very ambitious. Their success depends on the involvement of the whole institution and paying attention to the needs of the recipients, also by employees indirectly related to the communication of statistical information. It is

important for the statistical teams to understand that the way of data presentation is crucial to their understanding and effective use. Changing the traditional forms of disseminating statistical data can contribute to greater interest and involvement of users, which consequently will bring benefits to the institution, both internally and externally. The modernization of the forms of data dissemination is related to the preparation of data and products in a different way than before (for new structures). This will allow for some relief from the current work, as well as for acquiring new competences and skills.

In Statistics Poland, there are activities that enable, on one hand, to force change and, on the other, to encourage to them:

- top management decisions, which define the priority directions of work and indicate those dissemination forms and products on which the institution will focus,
- communicating employees the reasons for changes and the individual benefits associated with the new approach (e.g. targeted reduction of the burden of repetitive work, the ability to focus on development)
- activities from the bottom encouraging to look at statistical products from the perspective of their recipient. In Statistics Poland, teams responsible for designing the dissemination and communication products often use statistical data from various sources for their own work (e.g. preparation of macroeconomic and problem analysis). These teams, being both a supplier (data design) and a recipient (data user) understand the needs of external recipients very well. They try to share this perspective also with employees involved in preparing data for dissemination or with experts specializing in one research area. For this purpose, e.g. workshops on the use of databases and meetings promoting innovative digital products are organized. Their aim is to present how useful and attractive the result of their work can be when it is properly disseminated.

These activities are of ongoing nature and will therefore be complemented in the near future by a horizontal dissemination policy, setting strategic development directions and priorities for this important area of statistical activity.

Acknowledgment

Dissemination is one of the final phases of the statistical process, but in the era of digital information it becomes particularly important. Effective dissemination requires an integrated and systemic approach and the design of the entire process. It is important to build awareness that success in dissemination depends on the involvement of the entire institution, including those teams that do not participate directly in this phase of the statistical process.