

Using administrative data to improve policy making in the CEE

Policy brief

Though public authorities collect a lot of data on citizens and firms in post-socialist countries, administrative data are seldom used for research or to inform policy making. This policy brief explores the reasons and outlines recommendations for a more effective use of administrative data in generating evidence for policy making.

As a legacy of centralised planned economies, post-socialist European countries share a strong tradition of extensive data collection by the state. The administrative data collected by the tax authority, pension funds and other government bodies, though not in all cases, are often relatively high quality. Administrative data on individuals or firms have clear limitations but in many respects can be better than survey data. For example, administrative data can be more accurate, especially compared to surveys asking people to remember things in the past, or report their various forms of income. Administrative data can also provide large samples at a reasonable cost. Moreover, by linking two or more databases, we can produce such amount of information on the anonymised firm or individual in a complex data structure that would be infeasible to collect in a survey.

In Western Europe administrative data are regularly used to assess the impact of government interventions and also serve as a resource for the research community. The Nordic countries have a long track record in sharing admin data with researchers (see [Figlio et al 2015](#)) while others have started such initiatives relatively recently. For example, the German government set up a committee to explore the issue in 1999, established the [German Data Forum](#) in 2001, and opened two data service centres ([GML](#) and [IDSC](#)) in 2003. The UK set up [a task force](#) in late 2011 to examine the best ways to make administrative data available for research. Following its recommendation, the [Administrative Data Research Network](#) was founded in 2013 to develop the infrastructure for easy and ethical access. At the EU level, a directive was issued in 2003 to promote the secondary use of public sector information ([Directive 2003/98/EC](#)) and revised ten years later ([Directive 2013/37/EU](#)). The [Competence Centre on Microeconomic Evaluation](#) of the European Commission also actively promotes the use of admin data by governments of EU countries.

The use of administrative data often requires high technical skills in data cleaning and analysis. Most Central and Eastern European (CEE) countries have highly skilled researchers and statisticians who could meet this challenge and use admin data to support policy making.

So why don't we use this data resource?

One of the obstacles, ironically, lies in the laws on personal data protection endorsed when these countries reestablished democratic institutions in the early 1990s. Though researchers work with anonymised data, even that can be effectively blocked by strict data protection rules. In most CEE, data protection legislation is very restrictive, and often there are additional limitations on using information about ethnic background, health conditions or other sensitive characteristics. While in most countries there is legislation on citizens' right to access public information, this applies mainly to aggregate data and enforcement rules and institutions are weak. Lacking a clear legal framework or formal procedure, researchers' access to individual level administrative data is often determined in ad hoc and informal procedures.

Another common barrier is the lack of trust between academic and government organisations as well as within the government. In Hungary for example, a series of media scandals on the abuse of personal data by public officials during the 1990s increased public mistrust of how the government uses information about citizens, and generated fears within the public sector about the potential implications of using individual-level data. Data owners are also often concerned about losing their monopoly on publishing (or selling) their data or that external users may discover some shortcomings in data quality. Stakeholders may also doubt the need for improving access if they are unaware of new developments in statistical methods and the potential in using individual-level data.

Lastly, the relatively low efficiency of governance can play a role, as it implies that governments rarely commission proper (i.e. scientifically reliable) impact evaluations on their policies. Limited media attention or over-simplified public discourse on government accountability can further reduce demand for sophisticated analysis.

There are some promising new developments however

Some CEE countries are beginning to respond to internal demand or the EU directives. A new source of motivation has come from DG Employment of the EU Commission, which has recently introduced a rule that bidders to European Structural Funds should commit to producing at least one evaluation based on microdata and using counterfactual methods.

Open data initiatives have accelerated in [Croatia](#), the [Czech Republic](#), [Slovakia](#) and [Slovenia](#). Though these mainly focus on aggregate data, they can be an important first step in building a user community. Poland has established a Ministry of Digitisation that [promotes](#) the secondary use of admin data. Poland also regularly [uses social insurance data](#) to track the labour market career of university graduates as well as registered unemployment data to evaluate active labour market policies. Hungary has had an edge so far by adopting Act 101 in 2007, which eliminated the main barrier in the preceding legislation by establishing a legal basis for data owners to process personal data for the purposes of anonymisation. This has allowed significant progress in linking administrative databases and developing a [databank](#) at the [Institute of Economics of HAS](#), available to researchers.

Researchers can contribute to these positive developments in several ways

They can increase public awareness and understanding of the value of research based on administrative data by presenting results in an eye-catching and easy-to-understand, accessible way. Such efforts can be more successful if researchers engage with journalists and cooperate with each other in managing media relations. Researchers should also be more active in giving feedback on policy evaluations produced by the government: praising any attempt to commission external evaluations but highlighting possible weaknesses where the evaluation method was inadequate. Joint efforts to introduce courses in evaluation methods into the curricula of public administration training programmes would also help. These steps can help build up trust and understanding among stakeholders and prepare the ground for initiating changes in the legal framework. In the meantime, the research community needs to engage with other stakeholders in monitoring the adoption of the new [EU personal data protection regulation](#) to ensure that research and policy evaluation is endorsed as a legitimate cause for processing, archiving, and analyzing sensitive data.

These lessons emerged from discussions at a workshop held in Budapest on 21-22 September 2017. Participants came from the Visegrad countries, Albania, Bosnia-Herzegovina, Croatia, Romania, Serbia and Slovenia. The workshop was hosted by the [Hungarian Society of Economists](#), supported by the [Visegrad Fund](#) and organised in cooperation with the [Budapest Institute](#), [CELSI](#), [CERGE-EI](#) and [IBS](#).

July 2018

Ágota Scharle © Budapest Institute for Policy Analysis ([CC BY-NC](#))