Automated impact assessment - How digitizing government enables rapid and tailor-made policy responses.

Mariia Maksimova
University of Tartu
Tartu, Estonia
mariia.maksimova@ut.ee

Art Alishani
University of Tartu
Tartu, Estonia
art.alishani@ut.ee

Mihkel Solvak
University of Tartu
Tartu, Estonia
mihkel.solvak@ut.ee

Robert Krimmer
University of Tartu
Tartu, Estonia
robert.krimmer@ut.ee

ABSTRACT

As interest in the digital transformation of public administration grows, the main challenge remains to improve government governance systems and integrate a wider range of evidence into decision-making processes. The successful digitalization and application of such approaches improves the quality, responsiveness and flexibility of public administrations. The digialtization of processes has made it possible to use micro-level data to assess the impact of a policy or program and apply the feedback to improve the design and delivery of public services. Evidence-based policy-making evaluates programs based on their visible impacts. Large-scale data collected through digitized governance, coupled with econometric impact assessment, provides an ideal working toolkit for this. However, the current situation of European governments is one of slow adoption, as they are often slow to respond to new challenges. This is due to the static one-off impact assessment approaches used, the results of which quickly become outdated. With further digitalization, improvement of systems, and a rapidly changing situation, there is a need to speed up institutions' ability to quickly draw working solutions to offset the effects of unexpected events in society and economy and react without delays if policy effects dissipate. This paper demonstrates how a high level of digitalization in government allows addressing such issues by automating causal impact assessment and making it a continuous part of the service delivery. The use case is an automated system for assessing active labour market policies in Estonia using individual-level data from government digital registers. Building on this, it shows how impact assessment automation depends on automatically generated data, only available due to the digitalization of other public services, and how versatile it is when it comes to proving casual evidence in a suddenly changing environment.

dg.o 2022, June 15–17, 2022, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-9749-0/22/06…$15.00
https://doi.org/10.1145/3543434.3543653

CCS CONCEPTS

• **Applied computing** → **E-government**.

KEYWORDS

impact assessment, digitalization, proactivity

ACM Reference Format:
Mariia Maksimova, Art Alishani, Mihkel Solvak, and Robert Krimmer. 2022. Automated impact assessment - How digitizing government enables rapid and tailor-made policy responses.. In *DG.O 2022: The 23rd Annual International Conference on Digital Government Research (dg.o 2022), June 15–17, 2022, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3543434.3543653

1 INTRODUCTION

Public authorities face an increasing demand to provide citizens with quality services that are remotely accessible and available on any device. To effectively meet growing demands, governments need to embed new information and communication technologies (ICT) to facilitate interactions with end-users. The potential of ICT to increase the effectiveness of public administrations lies not only in making interactions more efficient, transparent and convenient [52, 63], but also in opening new ways for data exchange between different administrative areas, introducing common standards and making digital public services more human-centered [17, 51, 64]. Thus, the increasing availability of public sector data is helping officials to implement promising strategies that lead to new ways of managing operations, promoting the (re)use of public sector data and increasing efficiencies at various operational levels. The vast amount of historical and personal data that public administrations have access to also creates space for the use of increasingly advanced analytics, leading to the creation of new public values and the introduction of new data assessment mechanisms that support evidence-based policy-making and reduce administrative latency [24, 40].

The integration of data from different sources plays a crucial role in the implementation of digital government strategies that improve such areas as public effectiveness, operational efficiency, data management, and service delivery. Governments also have vast datasets of continuously updated personal and historical data, enabling the use of innovative analytics to generate new insights and inform public decision-makers. In addition, the growing ICT capabilities of public administrations are drawing attention to new

areas and ways in which relevant data mechanisms can increase policy responsiveness, and improve impact assessments and the accuracy of interventions. However, this is not an easy transition, and it does not come without its challenges. Current research identifies barriers that cut across different domains such as legal, management, security, ownership, quality, ethical, and privacy concerns [3, 21, 24, 40, 41].

Additionally, the use of big data arrays to evaluate ongoing actions and predict possible outcomes is a topic that is attracting attention from both business and government agencies. On the public administration side, the benefits of using data analytics to solve collective problems in various fields are noted, including city management [19, 22, 49, 59] or using social media data to support evidence-based policies [45]. In order to achieve high-quality results, public administrations are expected to develop their ICT management skills further and implement new technologies that support decision-making. Systems and tools based on previous generation technologies cannot always cope with such assigned tasks. As such, there is a growing demand for modern solutions that build on past successes and integrate new approaches that build on data to address pressing problems. Therefore, support public administrations in the delivery of digital transformation programs that go beyond electronic versions of current complex paper-based practices. As a result, it could be argued that developing methods that use open and real-time data to assess public decisions can provide a variety of benefits, including 1) optimizing resource spending through the development of efficient policies, and 2) improving budget spending and distribution of public goods, 3) improving the effectiveness of existing spending. This minimizes risks and takes opportunity costs into account. The labour market is one of the areas that can benefit significantly from such a development, as it is a vulnerable area that requires quick responses in situations where the nature and scale of challenges are difficult to predict.

The purpose of this paper is to showcase how the variety of micro-level data enables the creation of automated impact assessment tools. For this purpose, this study presents the case of the MALLE framework of active labour market policy impact assessment used by the Estonian Unemployment Insurance Fund (EUIF). By performing this research, the arguments put forward how higher-level digitalization in the public sector allows the development of precise tools and facilitates tailored policy responses. The rest of this study is organized as follows. Section 2 presents the main approaches to impact assessment. Section 3 discusses the digitization and automation of impact assessment and outlines the Estonian architecture supporting MALLE. In section 4, the case study is presented. Section 5 discusses the opportunities that public administrations could open up by reducing process latency and delivering proactive digital public services. Section 6 outlines the conclusions of the study.

2 BACKGROUND

2.1 Impact assessment and policy change

New emerging technologies are finding many applications among public administrations. This socio-technical transformation is creating new patterns of stakeholder interactions, and public policy mirrors these patterns, reflects their needs, and becomes a means of their implementation. In this way, state regulation is becoming a tool aimed at streamlining and ensuring sustainability through established and fixed norms and rules. However, government policies can have social, economic, and environmental consequences that lead to deviations from established norms and accepted programs and ultimately reduce the effectiveness of the implemented government regulatory system. Therefore, there is a need to introduce an impact assessment mechanism to determine the adequacy and effectiveness of states.

Impact assessment is a preliminary method of assessing the impact of a policy based on a variety of indicators in areas that the policy will impact upon implementation [20, 25]. In this regard, the effectiveness of the implemented policies can be assessed by using macroeconomic indicators and sociological surveys [46, 61], as well as examining the performance indicators of the government agencies [16, 50] directly involved in their implementation. The benefits of implementing impact assessments in public administration include the ability to identify social, economic and environmental benefits and costs. In doing so, it promotes more balanced decision-making at the state level. It also allows for a systematic expansion of the evidence base to support policy proposals, which can improve the ability to consider a range of implications of policy options even in the early stages of their development.

Impact assessment is used both at national level [2, 26, 42] to develop new strategies and at international organisations [38, 62]. This approach has established itself as common practice in numerous countries, as evidenced by the practice of its international application. In order to make the most effective decision, the data available must be used to produce the most significant results. In the past, effectiveness assessment was primarily a means of assessing the results of policies upon their implementation at the state level, but now, with more detailed information with higher granularity at an individual level and access to historical data, models can be developed that predict the outcome at lower levels of influence. Consequently, it is not limited to the assessment of public policies, but also to the assessment of individual indicators at the micro level. A high granularity of data is an important part of the evidence for policy as it allows to find more details about it [14]. This cannot be achieved without a sufficient degree of digitalization in public administration and automation, which would reduce the time required to process individual inquiries.

2.2 Digitalization and automating impact assessment

Global competition, knowledge economy and information society are among the prominent trends shaping today's society and in this regard, digital transformation is an integral part of today's life. The changes in modern society are mainly related to the changing role of information in the world. In the past, the term information society was primarily used, a society in which the phenomenon of information determines the social structure [8]. The technical means of communication, especially telecommunications, play a decisive role in the development of the information society, with the structure-forming role of information in the formation and development of society being the focus. A characteristic feature of the modern stage of development of society recognizes the leading

role in the development not so much information itself, but knowledge as its form, which led to the emergence of a new concept - the knowledge society [12]. While the notion of the information society is linked to the idea of technological innovation, the notion of the knowledge society touches on aspects of change in different areas of life and reflects the transition from using information for descriptive purposes to answering the question "how?" (See Figure 1).

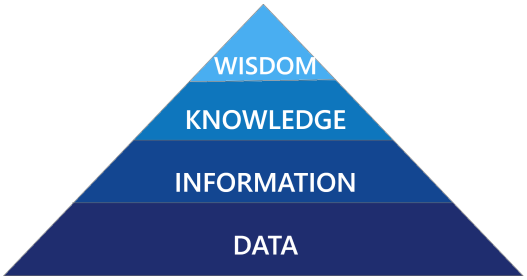

Figure 1: The DIKW hierarchy (Reprinted from: [1])

With digitalization, public administration can be made more transparent, accessible, and responsive by facilitating easy communication between the government and the public. As information sharing in the public sector becomes more open, decision-makers are taking into account citizens' needs, ideas and priorities, which can now be expressed through new channels or highlighted through their interactions with the government. It is an essential part of the digitalization of public administration to create a system that can both deliver public services in an electronic format and collect and store the data generated in the delivery of these services. By opening up possibilities to work with automatically generated data that is the by-product of this process, digitalization creates new possibilities for more responsive and flexible policies.

Therefore, digitization has led the state to collect a large amount of data on citizens, along with changes in indicators of socio-economic progress, which are increasing over time. The data is also accessible at the individual level of citizens' transactions, which makes it possible to assess the situation at a micro-level. With the ability to use micro-level data, it is possible to assess policy effectiveness more accurately and timely since it previously took 1-2 years to obtain stakeholder feedback and impact assessment [30]. Assessing policies at the micro-level is now a routine task for economic researchers and public service providers. In the last few decades, a large number of published individual evaluations have been carried out, summarized in several meta-analyses covering hundreds of studies (see on labour economics Card et al. [10], Card et al. [11]). Such an approach used to seem almost impossible due to the lack of data for full analysis and computing power, but now it is becoming part of the daily decision-making process.

Accordingly, different methods and technologies can be used in evaluations depending on the objective. When analyzing labour policy, they use either randomized designs [5], in which people or companies are randomly assigned to policy programmers, or more commonly, quasi-experimental designs [6], in which participants are compared with similar non-participants using statistical methods. It should be noted that randomized experiments and evaluations based on certain changes in the law (so-called natural experiments) cannot be automated because they require ad hoc-evaluations. Today, some countries are already introducing automated assessment tools for work actions. In Germany, TrEffeR (Treatment Effects and Prediction) has been introduced, a model for evaluating ongoing active labour market measures and identifying the effects on an individual basis, as well as for different subgroups and local areas European Commission [18], Stephan et al. [57]. The Public Employment Service (PES) of the Flemish region of Belgium (VDAB - Vlaamse Dienst voor Arbeidsbemiddeling en Beroepslopleiding) has implemented a profiling model that estimates the likelihood of becoming long-term unemployed [15]. Thus, easy access to registry data, improved computer speeds, and more flexible econometric and statistical tools have helped automate quasi-experimental approaches so that these evaluations can become routine.

2.3 Estonian digital government context

Public sector organizations can access new data sources to deliver proactive digital public services when converting analog processes to digital formats. However, current practices governing the collection and storage of user data are relatively closed and challenging due to various organizational boundaries, data silos, and legal obstacles. Access to separate data storage might work to provide individual services, but it is not sufficient to develop a comprehensive picture of citizen events and provide proactive services. As a result, the need for functioning systems that support data exchange between actors involved in services is evident and crucial for the development of well-functioning digital government ecosystems and sufficient complementary services that increase the value of the system as a whole.

Estonia's e-government is based on four essential infrastructure elements [55]. The first is the electronic identification system (eID) for citizens, for which several forms are available. The most popular is the use of a physical chip card, which also acts as an identification card. Another variant is the SIM card (Mobile-ID), which enables the use of a mobile phone for secure access to electronic services and signing documents. As of January 2022, about 1/6 of the holders of different identification cards have a mobile ID (https://www.id.ee/). The following variant of electronic identification system is Smart-ID, a cloud solution consisting of an application that allows user identification even if the device on which it is used has no SIM card. About 60% of the holders of various identification cards also have a smart ID (https://www.id.ee/). Although there are many methods of electronic identification, all systems are based on the principle of secure identification for accessing electronic services or signing documents.

The second element is a centrally managed distributed data exchange layer called X-Road, which enables data exchange between different actors. It is a one-stop-shop solution to access all data generated by the public sector in the country. This X-Road has been in use since 2001 and was originally designed to access various national databases. However, as it develops, it now allows large data sets to be modified and transferred and allows users to perform searches on them. As of January 1, 2022, 672 institutions and enterprises and 1,494 information systems are already integrated into

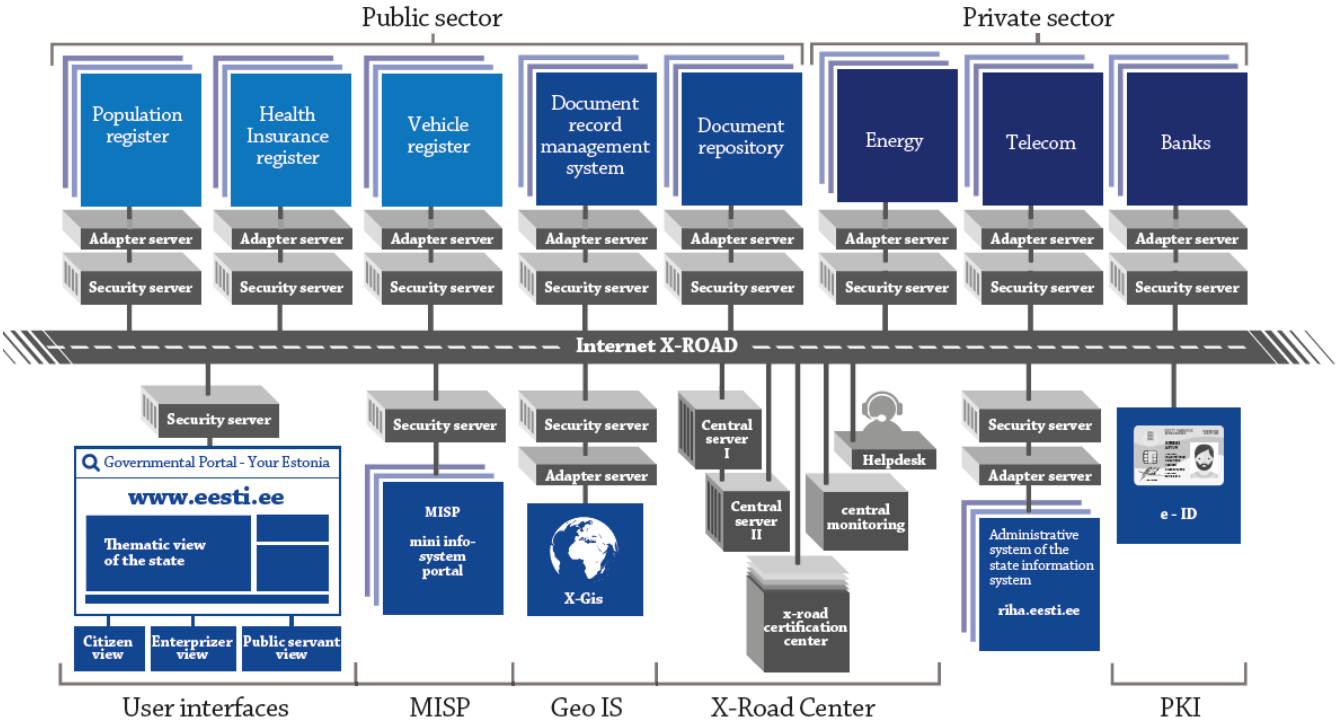

Figure 2: Schematic picture of X-road (Reprinted from: [56])

it (https://www.x-tee.ee/factsheets/EE/#eng). Approximately 200 million service requests are generated each month, of which only 3% are requests from citizens. The X-Road as such can be seen as an essential tool that enables daily data exchange and communication between different actors. The standardization of the process made it possible to reduce citizen involvement in service delivery, simplify information sharing, avoid duplication of data across multiple government agencies, and create a secure environment for these processes.

Regarding the X-Road system architecture (See Figure 2), most call requests are logged centrally by the X-Road Center in the central compliance monitoring log. External users cannot access data exchanged between parties connected to X-Road. The identities of the parties, mainly service portals and base registries, are also stored centrally. The central log stores limited information about service consumption for traffic monitoring and compliance purposes. This includes information about who sent the service request, to which service provider and when the response to the request was received. Also, as part of establishing a centralized log, requests are consolidated into a single record.

The standardization of the data exchange process and the centralization of data storage do not directly affect the owners of the data. Owners are also responsible for providing data, ensuring accuracy, updating and keeping data up-to-date. The integration of such a system does not bring any added value if the activities of each public administration agency involved in the process are not digitalized, since such a system is only a tool and a step on the way of digitalization. Thus, X-Road provides constant and uninterrupted access to government databases, but is not responsible for their quality and assumes no responsibility for their owners. The standardization of the data exchange process and the centralization of data storage do not directly affect the data owners. The latter are responsible for providing the data, ensuring its accuracy and keeping it up to date.

The RIHA (Republic of Estonia Information System Authority) repository system is a third element that acts as an e-service catalogue/registration for the state information system (https://www.riha.ee/Avaleht). At the same time, however, it is also an environment in which various procedural and administrative interactions take place. By law, registration with RIHA is mandatory for databases and information systems, and all databases are responsible for keeping the description of their content up to date. This is due to the principle of data storage adopted in Estonia, which is called "once only" [28]. It is that the information should not be duplicated anywhere else in the ecosystem. It helps reduce the cost of collecting and storing information and simplifies the system of interaction between authorities offering various electronic services. In addition, the system offers various reusable components such as XML assets and classifications that make it easier for new actors to connect to the system. The purpose of RIHA is to create a transparent, balanced and effective management of the information systems available in the country.

The fourth and last component is the national portal, which is a user interface for receiving electronic services (https://www.eesti.ee/et). All four elements together form a developed digital system of e-government functionality. In order to better understand how such digitalization allows rapid tailor-made policy responses, this paper proposes to examine the use-case of the active labour market policy impact assessment, which uses X-Road to get the data on the micro-level and use it for assessment model. By identifying actual usage-case and pointing out the benefits of this approach, we clarify the main argument of this paper that without high digitalization and real-time access to the data scattered across different government agencies, it becomes very difficult to react and respond to rapid changes. It is essential in current times due to the shift towards evidence-based policies and citizens' demand for more efficient, proactive, and tailored actions of the public administration.

3 ACTIVE LABOUR MARKET POLICY IMPACT ASSESSMENT - MALLE FRAMEWORK

The situation in the labour market is a critical area for government activity. It is of crucial importance as it influences economic development, citizens' living standards and the development and specialization of different professional and social skills [44]. While states continue to strive to enhance viability and increase the well-being of citizens, it has become increasingly difficult to achieve these goals without having programs in place that provide targeted support and respond to recent trends and situations. Therefore, it is of great concern to policy makers that labour market policies put in place should be properly evaluated to assess their impact and relative contribution. This paper provides a summary overview of the automated labour market policy impact assessment tool used in Estonia.

The quantitative evaluation of active labour market policy measures is a demanding and essential task that labour economists both in academia and in the public employment service regularly meet. Previously, the Estonian Unemployment Insurance Fund (EUIF) conducted evaluations of services on an ad hoc basis, about a study every two years, or commissioned external researchers to carry out analysis. However, with increased computing power and easier access to micro-level registry data within the EUIF, counterfactual impact assessment can now be done much more quickly and automatically. This reduces process latencies and enables a monthly iteration of the evaluations if required. To illustrate how EUIF labour market policy measures are automated, this paper provides an overview of what data is used and how it is combined, how the assessment is done and how the results are used.

In Estonia, micro-econometric evaluation of active labour market programs dates back to the early 2000s. The studies include the evaluation of labour market training programs [4, 13, 32, 34, 58], wage subsidies [4], and business start-up grants [60]. In addition, there are evaluations of labour market programs provided by other public institutions [35]. All recent studies were based on linked administrative micro-data from the Estonian Unemployment Insurance Fund and Tax and Customs Board. The studies used various matching methods, such as propensity score matching or exact matching. The majority of recent impact evaluations of active labour market policies in Estonia have been implemented by Estonian Unemployment Insurance Fund [32, 58, 60]. The lack of evaluation studies by academics is mainly due to the strict data protection regulations, which make it difficult for researchers outside the EUIF to access the micro data. More freely access to anonymized logs of service usage is possible, but it is quite difficult to trace a request for a service by an individual through automated requests between systems. In addition, the anonymization of logs makes further joining of records from different systems impossible.

Since 2021, the first version of the automated tool for active labour market policy evaluation - MALLE (stands for MAchine Learning in Labour Economics) (See Figure 3) was implemented in EUIF. It is an internal model that was jointly developed by analysts of EUIF and researchers of the University of Tartu with the support of private IT companies. The purpose of the tool is to speed up and automate labour market evaluations. The tool consists of three parts: 1) command files that automatically access data from EUIF databases and preprocess the data for evaluation, 2) command files written in R that perform statistical matching and write down matched micro level data, and 3) files for interface that visualise evaluation results either in Tableau or R Shiny and do necessary on-line calculations, such as impact effects in subgroups and approximate confidence intervals. The tool uses propensity score matching [9, 33] with some strict constraints, optionally combined with regression adjustment in the second step. Statistical matching combined with regression adjustment is one of the common approaches to assessing the impact of labour policies. It enables extensive background data and previous labour market experiences to be taken into account in order to make participants and non-participants comparable. Since effects in subgroups also have to be estimated after matching, which can lead to unbalanced treatment and control groups, a regression model is applied to reduce the effect of unbalanced covariates.

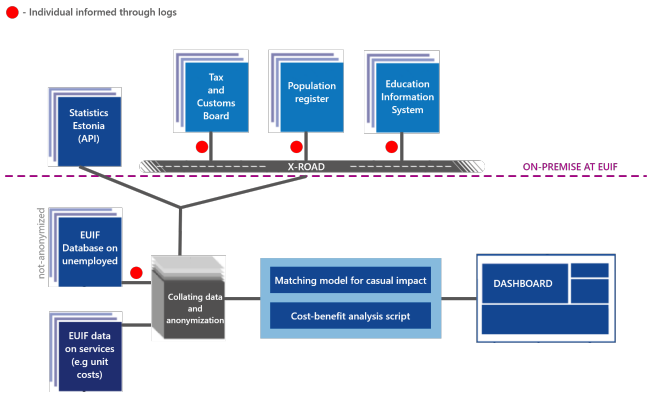

Figure 3: MALLE framework

The matching was based on the propensity score that was estimated using a logit model including various socio-economic variables and past labour market information. The variables matched exactly were: unemployment calendar month (control group had to be unemployed at the same calendar month when the labour market service was provided), observation period (both treatment

and control group observations had to be observed for the same time period 1-24 months), other services received during the same unemployment spell, a few measure-specific variables (e.g. ability to work in case of work-related rehabilitation, knowledge of Estonian in case of language courses). Model used 5 nearest neighbours (or more if there are controls with exactly the same value) to construct control and treatment groups. The matching procedure is done with R package "Matching". It estimates the effect of treatment variables on employed probability and gross monthly labour earnings up to two years after the beginning (or end in the case of wage subsidy) of active labour market measure.

The June 1, 2021 version of the model assesses the following labour market services: labour market training (separately vocational training, Estonian language training, computer training, general skills), work practice, wage subsidy, work-related rehabilitation. The variables used for matching include general information about the unemployed background such as age, gender, education, language, everything related to their previous employment experience and earnings, previous periods of unemployment and receipt of services or benefits, measures on work-related disabilities, location. All the data come automatically from the EUIF information system, which relies on other registries (Employment register, Tax and customs board register, Education register, population register), but a few items, e.g. working abroad or before the registers were created, are entered to the information system manually.

The technical implementation is carried out by combining R software and Tableau. R Server has access to Oracle database tables, data tables are then cleaned, merged and variables are defined. The matching is done with impact assessment, the average treatment effects on treated (ATT) on matched data is computed:

$$ATET = E_X\left[\left\{E\left(Y_1 \mid D=1, X^{Exact}, p(X^{Rest})\right) - E\left(Y_0 \mid D=0, X^{Exact}, p(X^{Rest})\right)\right\} \middle| D=1\right]$$

ATT is found from regression model:

$$Y = \beta_0 + \sum_{k=1}^{24} \beta_{1k} \times D \times T_k + \sum_{k=1}^{24} \beta_{2k} \times T_k + \beta_3 \times X + u$$

Matching also yields aggregated results with optional Abadie-Imbens standard errors and a micro-level dataset. After matching treated and controls are found, one can use either a Shiny application in R or Tableau to visualize the results. The application calculates the effect of the labour market policies based on the micro-data and calculates approximate standard errors (which currently ignore the uncertainty due to propensity score estimation and one-to-many matching) (See Fig.4). It also contains some basic information on the quality of the matching process (similarity between treated and non-treated groups) (See Fig.5).

The current model uses R (version 3.6.0), operated with RStudio (version 1.1.463), in a Linux machine (release 3.10.0-1062.1.2.el7.x86_64). The model is run on 16 cores (Intel(R) Xeon(R) Gold 6154 CPU @ 3.00GHz) with 200 GB RAM. The approximate time the model takes

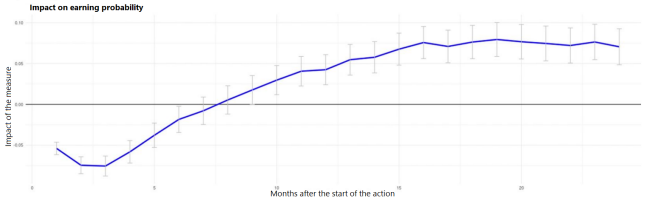

Figure 4: Impact of Estonian language courses on earning probability

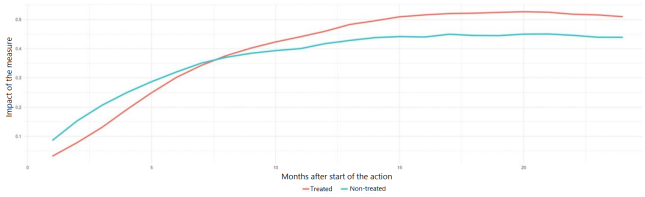

Figure 5: Impact of Estonian language courses on earning probability, treated vs non-treated groups

to get the data, do the matching and save micro-level datasets is about the following:

- Preparation of data from anonymised database – up to 20 minutes
- Matching for seven different treatments and storage of all micro data – about a day
- Loading of matched data into online tools 2-3 minutes
- Online calculations of effects in subgroups and approximate standard errors – a few minutes, depending on the size of the treatment group, number of subgroups.

4 DISCUSSION AND CONCLUSIONS

The role of ICT and data remains crucial in informing policy-making and enabling public administrations to experiment with new ways of developing modern digital public services that are inherently user-friendly, delivered in a timely manner and executed proactively in the background [29, 54]. It remains important to compare the stated goals with tangible results and to see what works the best in the policy-making cycle. This in turn requires integrated data from different sources and appropriate econometric methods and models to assess public administration actions and responses [14]. As the MALLE framework shows, an appropriate combination of data and econometric models can provide policymakers with significant support, complementary tools and decision-making bases [14]. Building on this, this study makes three observations.

Firstly, as a system based on automatically generated data, the advantages of the MALLE compared to the existing system in the EUIF are that it is fast, flexible and automated. For policymakers, it provides access to more up-to-date and authentic information. It enables more tailored policy responses and programs, and reduces process latencies by allowing assessments whenever they are needed. From conducting assessments every two years, the EUIF can execute it each month in an automated matter. This system also creates more space for human-centric and proactive services

that are tailored to users' life events, are context-sensitive, and arm officials with better evidence to support decision-making [36]. However, the current version has some weaknesses, such as simplified standard errors. Therefore, there is a trade-off between the accuracy of the estimates and the speed and flexibility to get the results. The automation allows routine evaluations to be performed quickly when using a quasi-experimental approach. The online tool enables flexible target group selection. Hypotheses can also be quickly put forward as to why some measures work better and how the effectiveness of measures has changed during the crises. In this way, it can be checked whether some measures work better for certain groups.

Second, MALLE also highlights how an established e-government architecture, particularly in the case of X-Road in Estonia, enables the extraction of administrative data to support the development of proactive digital public services and smart frameworks for impact assessment. Administrative data is often collected by public organizations and includes large datasets from various fields such as employment, environment, housing, public health, social security, education, etc. These datasets, while used to support public administrative functions, are typically representative and cover large population groups from citizens and businesses, are regularly updated, are reasonably accurate, and are occasionally available in real or near real time. As such, interest in administrative data is growing significantly given the value and insights it offers decision-makers when combined with other types of data, including big data [23, 31, 43, 47, 53].

Third, this study supports the argument that high levels of digitalization enable the creation of additional advanced tools that support decision-making and proactive digital public services. The successful implementation and deployment of digital government infrastructure, such as the case of X-road in Estonia, creates room for further developments that were not intended initially. In the case of MALLE, this system adds value not only by assuming what is happening in the labour market, but also by using micro-level data to monitor the situation more frequently and to support the shift towards an economy that increasingly builds on data to function in or close to real-time. In the field of public services provision, transactions at the user level allow identifying citizens' demand for certain services [39] and analysis on the group level can identify additional patterns and demands. That can be used both to optimize the processes and allow the state to recommend specific services in a proactive manner [27, 37, 65]. In addition, similar systems could also be developed for the private sector. Governments are increasingly collecting information about companies, consumer behaviour, and indicators regarding the general situation in various areas and industries. Therefore, it is becoming possible to use data to anticipate and monitor the situation of individual companies and also analyze the status of industries [7, 48]. Hence, develop systems that signal when organizations are at risk and formulate proactive policies to respond to various adverse events.

ACKNOWLEDGMENTS

The authors acknowledge the valuable input by our colleagues Andres Võrk (Conceptualization and methodology development) and Taavi Unt (Data curation and software) who were involved in the development of the MALLE framework and writing the initial draft of this paper. The work for this paper has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No.s 857622.

REFERENCES

[1] Russell L Ackoff. 1989. From data to wisdom. *Journal of applied systems analysis* 16, 1 (1989), 3–9.
[2] Camilla Adelle and Sabine Weiland. 2012. Policy assessment: the state of the art. *Impact assessment and project appraisal* 30, 1 (2012), 25–33.
[3] Anil K Aggarwal. 2019. Opportunities and challenges of big data in public sector. In *Web services: Concepts, methodologies, tools, and applications*. IGI Global, 1749–1761.
[4] Sten Anspal, Janno Järve, K Jääts, Epp Kallaste, K Maar, A Mägi, and A Toots. 2012. Palgatoetuse ja tööturukoolituse teenuste vahehindamine (Interim evaluation of the wage subsidies and labour market training). http://www2.sm.ee/esf2007/files/hindamine2010/2012_06_18%20T%C3%B6%C3%B6turumeetmete%20hindamise%20l%C3%B5ppraport%20%28CENTAR%20InterAct%29%20final.pdf
[5] Katherine Baicker, Amy Finkelstein, Jae Song, and Sarah Taubman. 2014. The impact of Medicaid on labor market activity and program participation: evidence from the Oregon Health Insurance Experiment. *American Economic Review* 104, 5 (2014), 322–28.
[6] Michael Baker, Jonathan Gruber, and Kevin Milligan. 2008. Universal child care, maternal labor supply, and family well-being. *Journal of political Economy* 116, 4 (2008), 709–745.
[7] Fernando Belfo, António Trigo, and Raquel Pérez Estébanez. 2015. Impact of ICT innovative momentum on real-time Accounting. *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy* 6, 2 (2015), 1–17.
[8] Daniel Bell. 1980. The social framework of the information society. *The microelectronics revolution* 545 (1980).
[9] Alex Bryson, Richard Dorsett, and Susan Purdon. 2002. The use of propensity score matching in the evaluation of active labour market policies. (2002).
[10] David Card, Jochen Kluve, and Andrea Weber. 2010. Active Labour Market Policy Evaluations: A Meta-Analysis. *The Economic Journal* 120, 548 (2010), F452–F477. https://doi.org/10.1111/j.1468-0297.2010.02387.x
[11] David Card, Jochen Kluve, and Andrea Weber. 2018. What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association* 16, 3 (June 2018), 894–931. https://doi.org/10.1093/jeea/jvx028
[12] Manuel Castells. 1996. The information age: Economy, society and culture (3 volumes). *Blackwell, Oxford* 1997 (1996), 1998.
[13] Centar. 2018. Eesti keelest erineva emakeelega täiskasvanute eesti keele õpe lõimumis- ja tööhõivepoliitikas: kvaliteet, mõju ja korraldus. https://www.sm.ee/sites/default/files/4._osa_eesti_keele_ope_tooturuteenusena_uus_logo.pdf
[14] Nuno Crato and Paolo Paruolo. 2019. The Power of Microdata: An Introduction. In *Data-Driven Policy Impact Evaluation*. Springer, Cham, 1–14.
[15] Lieselot Danneels and Stijn Viaene. 2015. Simple rules strategy to transform government: An ADR approach. *Government Information Quarterly* 32, 4 (2015), 516–525.
[16] Rex Deighton-Smith, Angelo Erbacci, and Céline Kauffmann. 2016. Promoting inclusive growth through better regulation: The role of regulatory impact assessment. (2016).
[17] Carol C Diamond, Farzad Mostashari, and Clay Shirky. 2009. Collecting and sharing data for population health: a new paradigm. *Health affairs* 28, 2 (2009), 454–466.
[18] European Commission. 2017. PES Practice Germany: TrEffeR (Treatment Effects and Prediction), author Torben Schewe.
[19] Michael Fitzgerald. 2016. Data-driven city management: A close look at Amsterdam's smart city initiative. (2016).
[20] Alberto Fonseca and Robert B Gibson. 2020. Testing an ex-ante framework for the evaluation of impact assessment laws: Lessons from Canada and Brazil. *Environmental Impact Assessment Review* 81 (2020), 106355.
[21] Cecilia Fredriksson, Farooq Mubarak, Marja Tuohimaa, and Ming Zhan. 2017. Big data in the public sector: A systematic literature review. *Scandinavian Journal of Public Administration* 21, 3 (2017), 39–62.
[22] Stephen Goldsmith and Susan Crawford. 2014. *The responsive city: Engaging communities through data-smart governance*. John Wiley & Sons.
[23] Katie Harron, Chris Dibben, James Boyd, Anders Hjern, Mahmoud Azimaee, Mauricio L Barreto, and Harvey Goldstein. 2017. Challenges in administrative data linkage for research. *Big data & society* 4, 2 (2017), 2053951717745678.
[24] Rasha G Hassan and Othman O Khalifa. 2016. E-Government-an Information Security Perspective. *International Journal of Computer Trends and Technology (IJCTT)* 36, 1 (2016), 1–9.

[25] Katharina Helming, Katharina Diehl, Tom Kuhlman, Torbjörn Jansson, Peter H Verburg, Martha Bakker, Marta Perez-Soba, Laurence Jones, Pieter Johannes Verkerk, Paul Tabbush, et al. 2011. Ex ante impact assessment of policies affecting land use, Part B: application of the analytical framework. *Ecology and Society* 16, 1 (2011).

[26] Julia Hertin, John Turnpenny, Andrew Jordan, Mans Nilsson, Duncan Russel, and Björn Nykvist. 2009. Rationalising the policy mess? Ex ante policy assessment and the utilisation of knowledge in the policy process. *Environment and Planning A* 41, 5 (2009), 1185–1200.

[27] Vincent Homburg and Andres Dijkshoorn. 2012. Diffusion of Personalized E-Government Services among Dutch Municipalities: An Empirical Investigation and Explanation. *E-Government Services Design, Adoption, and Evaluation* (2012), 221.

[28] Robert Krimmer, Tarmo Kalvet, Maarja Toots, Aleksandrs Cepilovs, and Efthimios Tambouris. 2017. Exploring and demonstrating the once-only principle: a European perspective. In *Proceedings of the 18th annual international conference on digital government research*. 546–551.

[29] Peter Kuhn, Matthias Buchinger, and Dian Balta. 2021. How to Redesign Government Processes for Proactive Public Services?. In *International Conference on Electronic Government*. Springer, 29–40.

[30] Sven Langedijk, Ian Vollbracht, and Paolo Paruolo. 2019. The potential of administrative microdata for better policy-making in Europe. *Data-Driven Policy Impact Evaluation* (2019), 333.

[31] Graeme Laurie and Leslie Stevens. 2016. Developing a public interest mandate for the governance and use of administrative data in the United Kingdom. *Journal of Law and Society* 43, 3 (2016), 360–392.

[32] Anne Lauringson, Kristi Villsaar, Liis Tammik, and Teele Luhavee. 2011. Tööturukoolituse mõjuanalüüs (Counterfactual Impact Analyses of the labour market training). https://www.tootukassa.ee/sites/tootukassa.ee/files/TOOTURUKOOLITUSE_MOJUANALUUS.pdf

[33] Michael Lechner. 2002. Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics* 84, 2 (2002), 205–220.

[34] Reelika Leetmaa, Andres Võrk, Raul Eamets, and Kaja Sõstra. 2003. Aktiivse tööpoliitika tulemuslikkuse analüüs Eestis (Evaluation of the active labour market policy in Estonia).

[35] Reelika Leetmaa, Andres Võrk, Miko Kupts, and Laura Kirss. 2015. Counterfactual Impact Evaluation (CIE) of Estonian Adult Vocational Training Activity. http://www.praxis.ee/wp-content/uploads/2015/02/Praxis_CIE_Estonia_31Oct2015_EE.pdf

[36] Dennis Linders, Calvin Zhou-Peng Liao, and Cheng-Ming Wang. 2018. Proactive e-Governance: Flipping the service delivery model from pull to push in Taiwan. *Government Information Quarterly* 35, 4 (2018), S68–S76.

[37] Bin Liu, Tao Chen, Peng Jia, and Lin Wang. 2020. Effective public service delivery supported by time-decayed Bayesian personalized ranking. *Knowledge-Based Systems* 206 (2020), 106376.

[38] Fernando Loayza. 2012. Strategic Environmental Assessment in the World Bank. (2012).

[39] Mariia Maksimova, Mihkel Solvak, and Robert Krimmer. 2021. Data-Driven Personalized E-Government Services: Literature Review and Case Study. In *International Conference on Electronic Participation*. Springer, 151–165.

[40] Petar Milić, Kristijan Kuk, Turhan Civelek, Brankica Popović, and Stefan Kartunov. 2016. The importance of secure access to e-government services. *Communications* 26, 16 (2016), 1873–18893.

[41] Ricard Munné. 2016. Big data in the public sector. In *New Horizons for a Data-Driven Economy*. Springer, Cham, 195–208.

[42] Måns Nilsson, Andrew Jordan, John Turnpenny, Julia Hertin, Björn Nykvist, and Duncan Russel. 2008. The use and non-use of policy appraisal tools in public policy making: an analysis of three European countries and the European Union. *Policy Sciences* 41, 4 (2008), 335–355.

[43] Daniel Tumminelli O'brien, Robert J Sampson, and Christopher Winship. 2015. Ecometrics in the age of big data: Measuring and assessing "broken windows" using large-scale administrative records. *Sociological Methodology* 45, 1 (2015), 101–147.

[44] Truman G Packard and Claudio E Montenegro. 2017. Labor policy and digital technology use: indicative evidence from cross-country correlations. *World Bank Policy Research Working Paper* 8221 (2017).

[45] Panos Panagiotopoulos, Frances Bowen, and Phillip Brooker. 2017. The value of social media data: Integrating crowd capabilities in evidence-based policy. *Government Information Quarterly* 34, 4 (2017), 601–612.

[46] Alari Paulus, Mitja Čok, Francesco Figari, Péter Hegedüs, Nataša Kump, Orsolya Lelkes, Horacio Levy, Christine Lietz, Silja Lüpsik, Daniela Mantovani, et al. 2009. *The effects of taxes and benefits on income distribution in the enlarged EU*. Technical Report. EUROMOD working paper.

[47] Anthony Quinn, David Denney, Nick Hardwick, Rahul Jalil, and Rosie Meek. 2020. The feasibility and challenge of using administrative data: a case study of historical prisoner surveys. *International Journal of Social Research Methodology* (2020), 1–8.

[48] Mysore Ramaswamy, Richard Calvasina, Eugene Calvasina, Gerald Calvasina, and UT Cedar City. 2011. Impact of Information Technology Innovations on Financial Transaction Processing. *Issues in Information Systems* 12, 1 (2011), 36–43.

[49] M Mazhar Rathore, Awais Ahmad, Anand Paul, and Seungmin Rho. 2016. Urban planning and building smart cities based on the internet of things using big data analytics. *Computer networks* 101 (2016), 63–80.

[50] Chris JR Roche. 1999. *Impact assessment for development agencies: Learning to value change*. Oxfam.

[51] Poonam Salwan and Veerpaul Kaur Maan. 2021. E-Governance using Big Data. In *Innovations in Computer Science and Engineering*. Springer, 123–133.

[52] Robert Schware. 2005. *E-development: From excitement to effectiveness*. World Bank Publications.

[53] Anna Sexton, Elizabeth Shepherd, Oliver Duke-Williams, and Alexandra Eveleigh. 2018. The role and nature of consent in government administrative data. *Big Data & Society* 5, 2 (2018), 2053951718819560.

[54] Natalie Shlomo. 2019. Overview of data linkage methods for policy design and evaluation. In *Data-Driven Policy Impact Evaluation*. Springer, Cham, 47–65.

[55] Mihkel Solvak, Taavi Unt, Dmitri Rozgonjuk, Andres Võrk, Märten Veskimäe, and Kristjan Vassil. 2019. E-governance diffusion: Population level e-service adoption rates and usage patterns. *Telematics and Informatics* 36 (2019), 39–54.

[56] Mihkel Solvak and Kristjan Vassil. 2016. E-voting in Estonia: Technological diffusion and other developments over ten years (2005-2015). (2016).

[57] Gesine Stephan, Susanne Rässler, and Torben Schewe. 2006. Das TrEffeR-Projekt der Bundesagentur für Arbeit: die Wirkung von Maßnahmen aktiver Arbeitsmarktpolitik. *Zeitschrift für ArbeitsmarktForschung* 39, 3/4 (2006), 447–465. http://doku.iab.de/zaf/2006/2006_3-4_zaf_stephan_raessler_schewe.pdf

[58] Liis Tammik. 2015. Eesti Töötukassa pakutava tööturukoolituse mõju koolitusel osalenute hõivele ja palgale. Magistritöö. http://dspace.ut.ee/bitstream/handle/10062/47196/tammik_liis.pdf

[59] Anthony M Townsend. 2013. *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company.

[60] Kristi Villsaar, Liis Tammik, and etal. 2014. Ettevõtluse alustamise toetuse mõjuanalüüs. https://www.tootukassa.ee/sites/tootukassa.ee/files/EvATanalyys_30052014.pdf

[61] A Võrk, M Piirits, and E Jõgi. 2015. The Impact of Introduction of Funded Pension Schemes on Intragenerational Inequality in Estonia. *NETSPAR Academic Series, DP* 6 (2015), 2015–021.

[62] Fanny Vuorinen, Barbara C Pescadinha, Mikael Hildén, and Jacques Jansen. 2014. Impact Assessments at International Organisations: An Overview of Strategies and Processes. (2014).

[63] Darrell M West et al. 2004. *Global e-government, 2004*. Center for Public Policy, Brown University Providence, RI.

[64] Hilmar Westholm. 2005. Models of improving e-governance by back office reorganisation and integration. *Journal of Public Policy* 25, 1 (2005), 99–132.

[65] Caie Xu, Lisha Xu, Yingying Lu, Huan Xu, and Zhongliang Zhu. 2019. E-government recommendation algorithm based on probabilistic semantic cluster analysis in combination of improved collaborative filtering in big-data environment of government affairs. *Personal and Ubiquitous Computing* 23, 3 (2019), 475–485.