Variable Selection and Data Quality Challenges in Impact Assessments

Monica ROMAN¹,
Liliana-Olivia LUCACIU²

¹ PhD, University of Economic Studies,
Bucharest, Romania,

monica.roman@csie.ase.ro

² PhD candidate, University of Economic
Studies, Bucharest, Romania,

lilianalucaciu@live.com

Abstract: The research is focused on the role of two related key concepts, namely variables and data, in the impact evaluations of public projects. A difficult task of the evaluators and researchers is to select the appropriate variables to ensure the best model of reality and satisfy the evaluation methods' needs. Therefore, the paper aims to look at the current knowledge and discuss how variables and data could be best used to connect the evaluation models, the particularities of the intervention with the potential of the advanced quantitative assessment methods.

The results emphasise that evaluations operate with data with different levels of granularity, as required by the intervention logic. Structuring data in clusters and categories, performing evaluability assessments are useful in assessing data quality and limitations and improving them. In line with the existing literature, we demonstrate that data accessibility is a key constraint and imposes adjustment of the desired evaluation model to a feasible one.

While Big Data and Open Data systems significantly improved data quality in evaluations in recent years, blockchain, as a ledger technology with default features related to decentralisation and security, is expected to bring large benefits to evaluation. For evaluators and policymakers, blockchain potential is an area of further research looking for additional advantages that could enhance the use of quantitative methods.

Keywords: *impact evaluation, counterfactual impact evaluation, variables, data quality, big data, blockchain.*

How to cite: Roman, M., & Lucaciu, L.-O. (2021). Variable Selection and Data Quality Challenges in Impact Assessments. *Postmodern Openings*, 12(3Sup1), 01-20. <https://doi.org/10.18662/po/12.3Sup1/348>

1. Introduction

Public policies involve complex phenomena and abstract concepts, making the task of the evaluators and researchers difficult in transposing them into measurement tools for results and effects.

The conceptualisation and operationalisation of the evaluation aim at modelling the reality, the complex phenomena related to the interventions.

The variables and the indicators enable analysts and evaluators to describe and analyse the complex reality of the intervention. Particular attention is paid to indicators that play a central role in the analysis and evaluation, mainly in the evaluations based on quantitative methods. The selection and definition of indicators must be logical and relevant to the intervention, coherent with the theory and practice in the sector.

The "how we select" is equally important as the "what we select". Ensuring a high-quality process is a key capability of a good evaluator to identify an optimum number and most appropriate variables and, further on, the most appropriate indicators for accurate measurements.

The paper is focused on the impact evaluations of the European Structural Investment Funds (ESIF) and two key related concepts, namely variables and data. The paper's main objective is to argue the crucial role of variable selection and data quality in public policies evaluation. We consider the following dimensions of the analysis: the various taxonomies of variables and corresponding indicators, what specific variables should be selected, why they are adequate for the evaluation methods, and how data quality may affect the results of the evaluations. The research highlights the complexity of the concepts, their features of interest for the counterfactual impact evaluations and the methodological limitations they generate in practice.

The need to clearly understand the challenges of the impact evaluation is treated in the most recent papers (White, 2014), in the context of a significant increase in the number of impact evaluations worldwide. White (2014) argues that there is a need for a "further increase in both quantity and quality of impact evaluations" (White, 2014, p. 18). Befani et al. (2014) identify some challenges in the post-2015 evaluation context related to "Increasingly ambitious development goals, multiple layers of governance and lines of accountability". All these should be tackled by considering a quantitative evaluation design, however, with "adequate causal inference frameworks" (p. 1), while evaluations should be more impact-oriented. For answering the increasing complexity of multi-project evaluations, Williams (2015) suggests a systemic approach, relying on three pillars: interrelations, perspectives, and boundaries. Several papers consider the challenges related to using statistical

methods, such as randomised experiments or surveys (Vaessen, 2011), while others argue the advantages of using counterfactual impact evaluation in specific fields, such as environment (Ferraro, 2009). The issue of variable selection and data quality is approached in several papers and textbooks on impact evaluation, more specifically in counterfactual impact evaluation. Hujer et al. (2004) show the data limitations related to impact evaluation. However, this issue is not extensively covered by the existing literature or in a context associated with ESIF.

Considering our findings, the paper contributes to the evaluation literature in two ways. Firstly, we provide an updated perspective on the challenges associated with two critical concepts in the evaluation process, namely variable selection and data quality, using ESIF as a case study. Secondly, the paper has the original contribution of developing a potential solution to the identified limitations. In this regard, we consider data accessibility and benefits that evaluation could enjoy from Big Data systems and blockchain technology extension. Access to data is a largely recognised constraint in impact assessments; therefore, how this constraint is addressed by practitioners and what we can expect from Big Data and blockchain is an area of interest in this subject.

The research relies on a qualitative methodology. We use documentary research (Silverman, 2011), and we have considered two types of documents: institutional documents provided by the European Commission and national bodies that regulate the field of ESIF evaluation and research papers and guides focused on the practice of evaluation. The time span covers the last two programming periods in the EU, namely 2007-2013 and 2014-2020. In each section, we have presented the theoretical concepts in detail, complemented with practical examples related to ESIF evaluations, to identify the role and challenges in the evaluation use of data and variables.

In the following sections, the paper presents the main categories of variables and corresponding indicators and their roles and associated challenges for the evaluation process (Section 2), the typology and roles of data (Section 3), and the constraints and potential solutions (Section 4). The final section presents the main research conclusions.

2. Variables and indicators selection, definitions, and quality assessment

The most recent evidence from the practice of evaluation highlights the challenges related to variable selection in ESIF. This issue is critical in

the context of complex and complicated development programs, like those implemented in the developing countries over the last two EU programming periods.

Recent evaluations of Cohesion Policy programs in Romania (Lucaciu et al., 2020; Lucaciu et al., 2021a; Lucaciu et al., 2021b) reveal examples of the challenges of conceptualising and operationalising an impact evaluation. The challenge comes from understanding the different dimensions of a phenomenon, focusing on the relevant ones, and transposing them in the variable that "varies" with the intervention.

One example drawn from practice is the social marginalisation phenomenon, with multiple dimensions. We need to understand and correctly define the dimensions addressed by the intervention and then select the appropriate variables and measurements. In this case, the well-targeted intervention may address only employment and not other dimensions such as access to services, level of education, etc. For the employment dimension, the appropriate variables may be employment status or continuity in employment, reflecting in addition employability (Lucaciu et al., 2021a).

Participation in education offers other examples with several options for the variables and indicators selection upon the change aimed by the intervention. Thus, increasing involvement in education could be preventive and capture with appropriate variables the reduction of the dropout or remedial capturing returning to the education of early school leavers.

Similar challenges are found in business support programmes. The intervention aims at increasing the firms' viability, competitiveness, and related to the last one, innovation and strengthening the market position. Viability or competitiveness may be challenging to identify the appropriate variables. For viability, one may use the variable "survival after two years of support", or profitability, or a combination of them. For competitiveness, the dimensions used in the largely accepted definitions should be used. As in social marginalisation, the model should focus on the dimensions addressed by the intervention because they are expected to vary, and the variation could not be observable at a higher level, involving several dimensions.

For a clear understanding of the role of each variable in the evaluation models and a good variable selection, one needs to understand the features of the variables. These features are presented further, together with a model for assessing the quality of the evaluation indicators.

Berman & Wang (2012) use the term "variables attributes" in the process of operationalisation. The attributes of variables are those specific characteristics that determine how variation occurs. The variables are

associated with attributes coherent with the dimensions of the intervention that allow the design of the evaluation model.

The primary filter in selecting variables and indicators for the evaluation model is "data availability". Practice confirms that data availability makes the difference between what we ideally would measure and what is feasible. Finally, the indicator comes in the process as the essential tool to measure the effects of interest, representing the desired change of the intervention (Gertler et al., 2016)

The shift from theory to practice leads the research in one of the most extensive "evaluation arenas", the European Structural and Investment Funds (ESIF) of the Cohesion Policy. European Commission (EC) guidance documents place less emphasis on analysing phenomena, identifying dimensions and variables, and referring directly to measurement indicators and their quality requirements, thus leaving the evaluators' responsibility to go through a systematic and structured operationalisation process. Although programs have already defined indicators for monitoring and evaluating programs, we rarely find impact indicators or indicators for long-term effects, and their adequacy for the evaluation process needs to be verified. To answer these limitations, Lucaciu (2018) proposes a five-step process to set the variables and indicators needed in an ESIF intervention evaluation.

- Clarification of the logic of the intervention, of the causal links from the identified problem to the change that will reflect the solution of the problem, from resources to actions, achievements, and impacts.
- Identification of the relevant dimensions for the proposed intervention.
 - Defining the variables and attributes relevant to the intervention.
 - Establishing the measurement mode, identifying the indicators.
 - Analysis of data sources and information for the identified indicators.

Berman & Wang (2012) highlight the critical elements of the variables analysis and the need to ensure they are relevant to the funding program and public intervention. The descriptive analysis of the variables must provide the essential information about the relevant phenomena of the financing program: the characteristics, the location, the time, and an exploration of the relations between the variables, differentiating the causal relations from the association ones. A good understanding of the aspects mentioned above is beneficial in quantitative evaluations, such as the econometric models' design, endogeneity testing, and inclusion of the instrumental variables in the model if needed.

A key theme in the statistical and evaluation literature that needs to be considered from a practical perspective, is the distinction between independent and dependent variables, considered a “foundation in research” (Berman & Wang, 2012). The literature review presents different classifications of the intervention variables. The most relevant and often referred to (Gertler et al., 2016; Berman & Wang, 2012) is coherent with the intervention logic model and consists of four categories:

- Explanatory variables, also known as independent variables, are those entering the results chain, are not influenced by other chain variables, and explain the outcome.
- Outcome variables, known as dependent variables, are the variables to be explained and correspond to an outcome, an effect of the intervention.
- Control variables are research variables. They are neither explanatory nor outcome variables, they are used to test hypotheses.
- Confounding variables are also control variables but have the particularity that they can explain the effect without being part of the chain of intervention results.

Other variable classification models focus on how they intervene in the results chain, such as moderators, with direct influence in the chain or mediators, omitted from the results chain, but intermediate the influence of the independent variables on the dependent ones. The synthetic image of how various phenomena and their associated variables occur in the chain of production of results and effects is given in Figure 1.

Source: Field-Fote (2019). p. 84.

Figure 1. Types of variables and their links with the intervention logic

The policymakers are in general interested in the outcomes at beneficiaries' level and the induced effects, in the broader environment at the territorial level, both reflected in the variable selection (Bondonio, 2009 cited by Lucaciu 2018, p. 70).

"Three categories of variables in the causal chain of the intervention:

- Type A: Variables on available and allocated funds.
- Type B: Variables on changing the behaviour of assisted firms regarding investments and jobs created and maintained.
- Type C: Variables on the induced effects of the program on the economic and social target environment".

A planned approach to variables and indicators selection and definition should consider four sources of information and ideas (i) previous studies; (ii) usual known dimensions; (iii) dimensions defined by the programme (intervention) (iv) practical constraints generated, for example, by data availability (Berman & Wang, 2012).

The analysis, selection and definition of the variables is an example of connecting quantitative with qualitative analyses. The quantification and estimation of the relevant effects of the intervention depend on the evaluators' ability to select the variables and corresponding indicators so that they can measure the dimension of interest. The theory of change proves the relevance through qualitative analysis. Gertler et al. (2016) confirm this by mentioning that adaptation to context is as important as its method.

The issue of causality and relation is also an area of coherence of the qualitative and quantitative analysis and requires mixed competencies in public policy evaluation and quantitative statistical methods.

The statistical theory argues that "Correlation is not causation" (Gertler et al., 2016). Conn (2017, p. 731), quoting the renowned economist Thomas Sowell said: "One of the first things taught in the introductory statistical textbooks is, correlation is not causality. It is also one of the first things forgotten."

Two conditions must be met to confirm the causality relation between two variables:

- To have a plausible explanation of the connection between them (this will become the hypothesis) and results from the qualitative analysis of the logic of the intervention/chain of results;
- To empirically test this connection by statistical methods.

What are the most appropriate variables, how many are the challenging questions the evaluator and analyst must answer in applying quantitative methods for impact assessment? This double question is relevant when the number of variables is part of a trade-off for models'

optimisation. For example, such situations are found in Data Envelopment Analysis (DEA) (Roman & Gotiu, 2017) or in the more generic case of econometric models that require an optimum number of variables as a trade-off between bias and efficiency (Popescu & Roman, 2018). Omission of some variables can seriously affect the accuracy of the impact estimate, as well as their erroneous selection, while a large number of variables increase the effort, not necessarily adding value to the results.

The answers to the double question may differ depending on the analysis level, either at the immediate level of the intervention or at the final level. A known error is the use of proximal effects of phenomena relevant to the generation of final effects (Scott-Phillips, et al., 2011). Avoiding this error is possible (1) by clarifying the evaluation question, which is the level where policy users want to know the impact and reflecting it correctly in the chain of results (2) by the relevance of the intervention to the various levels of effects. An intervention with limited resources may have an observable impact at the local level but may not produce one at a national level.

The indicators quality defined and used in the ESIF programs is essential for the evaluation results. The EC is constantly concerned with monitoring and evaluation systems, and part of them with the quality of indicators. We find extensive publications in the form of manuals and guides of the Directorates-General of the EC applicable to the funding programs they manage (European Commission, 2014; European Commission, 2019; Morris et al., 2013)

Similarly to the theory of change, the terminology related to indicators varies in different policy areas, sometimes creating confusion. The indicators used in the ESIF 2014-2020 are classified on three levels of the intervention logic as follows:

- Outcome indicators – refer to variables that provide information about certain aspects of the results and reflect what is intended to change, not limited to the effect of the funded intervention, but also considering the effects of other interventions and factors.
- Output indicators - refer to variables that provide information on the direct products of the programs; they derive from logical intervention and express the actions that generate outputs.
- Input indicators – refer to the public support provided, the funding, and other inputs, if the case.

A specific category of indicators is defined in the context of ESIF, i.e., "Common indicators". They refer to a selection of indicators that the EC recommended to the EU member states to allow aggregated achievements at the EU level. This approach is strengthened with regulatory

provisions in the Monitoring and evaluation of European Cohesion Policy (European Commission, 2014) and the programming guidelines, including the following requirement:

- Include at least one outcome indicator for each specific objective (SO).
- There should be no more than two result indicators on the OS.
- Outcome indicators for expected results should include an adequately quantified baseline and target.
- The output indicators include quantified targets; it is expected to contribute to the results and be defined for each investment priority, reflecting the actions covered.

The European Commission (2014) formulates four criteria as a requirement for assessing program indicators. These requirements can be used to create tailored tools for analysing the quality of indicators, such as the example designed for ESIF programmes in Romania, provided below.

- Responsibility and relevance: the outcome indicator must respond to policy and capture the essence of the outcome, which means that the value of each outcome indicator is influenced by actions and outputs and reflects the expected policy-sensitive change.
- Clarity / Normative: the indicator has a clear and accepted interpretation.
- Robustness: the indicator is not influenced by extreme values, which means that the indicator reflects the intended change and is not sensitive to large variations generated by a small number of entities or inputs.
- Statistical validation: the result indicator is based on information collected from all entities/participants targeted by the intervention (statistical population) or a representative sample.
- Data for the indicator are collected regularly; the data are available when needed for evaluation purposes.

The SMART test is also used in planning and programming practice (European Commission, 2019). It includes the criteria: specificity, measurability, the extent to which targets can be achieved, realism (data can be collected, monitored, aggregated, and interpreted), setting the time horizon for measuring indicators.

In public policies, real-world indicators are not limited to an intervention, a program, or a policy. The coherence of indicators across policies provides benefits to all policymakers in a complex public policy environment with intersectoral programs at multiple territorial and governance levels. SMART City local policies are one of the relevant

examples, where actors participate in numerous programs funded by different funds, projects whose results are expected to contribute to the objectives of SMART City. It is thus useful to connect the system of performance indicators of smart cities (Carli et al., 2013) to the indicator systems of programs financed from structural funds, in Romania as in other countries, the main source of financing for many programs. Lucaciu (2018) identifies common performance indicators for SMART cities strategies and ESIF highlighting common indicators at the level of result and impact, sometimes at different levels of disaggregation.

To conclude, this section emphasised the need for a proper variable selection as a challenge of the evaluation process. This challenge may be overcome by considering all the variable characteristics, their various taxonomies and roles, and their adequacy for evaluation.

3. Data used in the evaluation and the data quality issue

Behind variables, which are abstract concepts, we find data. In a more quantifiable environment, data impacts the evaluation process through their quantity and availability or quality. ESIF impact assessment operates with variables concerning various units of analysis, such as persons or entities influenced by the intervention and variables relating to different territorial levels or large communities and groups.

For this reason, we find in evaluations data with different levels of granularity (Crato & Paruolo, 2019). Microdata are used for measuring the effects at the unit of analysis, while aggregated data are used for induced effects at the policy target population level or data for context analysis and influencing factors the generic covariates in the evaluation models.

Evaluation methodologies use data classifications relative to their usefulness in the evaluation process. One classification, useful for organising data collection, is structured on five clusters: data from individuals, data from groups, observations, physical measurements and revisiting existing records.

The clusters include a broad range of collection methods, ranging from traditional individual and group interviews, focus groups, panels, surveys, biophysical and geographical measurements, to more modern data collection methods using smartphones, feedback using tools based on information and communication technology (ICT). While the first four clusters refer to primary data collected for evaluation purposes, the last cluster relates to data collected for a purpose other than evaluation and can be extracted and processed to meet the evaluation needs.

The European Commission (2021) uses in the monitoring of the evaluation of the interventions financed by the Cohesion Policy a classification that highlights four types of data, namely:

- Primary data collected directly by evaluators through collection methods dedicated to satisfying the need for data for evaluation.
- Statistical data produced by "statistical data producers".
- Administrative data - data collected by government entities during the current activity for administrative purposes; monitoring data are administrative data specific to the evaluated program, collected to serve the management of the intervention.

Correlating the two classifications, we find that the primary data correspond to the first four clusters: data from individuals, groups, observations and measurements, statistical, and administrative data, including monitoring data correspond to the fifth cluster, secondary data, already recorded.

As noted by Elia et al. (2015, p. 7), "The main challenge encountered to carry out CIE1 analysis was the availability and access to a suitable data set on the control group". The authors point out limitations on the conclusions of the evaluations analysed due to the quality of the data. The limitations include the use of a limited set of variables that fail to eliminate selection bias, external validity is limited because the group treated corresponds to the target population of the policy, insufficient sample size data for important variables.

"Administrative data can suffer qualitative challenges that need to be aware and solved. Data are sometimes (1) missing or incomplete, (2) inaccurately reported, (3) subject to definitions that have changed over time and therefore cannot be compared, (4) irrelevant to specific events or customers or cannot be disaggregated in necessary ways, (5) confidential, unavailable for analysis, or (6) insufficiently available in electronic format." (Berman and Wang, 2012, p. 82)

In general terms, data quality reflects to what extent a set of inherent characteristics of an object fulfils requirements. A simple definition of quality is "fit for use" or "fit for purpose". (United Nations, 2019, p. 3). From the users' perspective, the concept of "fitness for use" means adequacy of data for the consumers' use.

For data quality assessment purposes, we found a helpful definition of Wang & Strong (1996, p. 6): "the dimension of quality is defined as a set of data quality attributes that represent a single aspect or construct of data

¹ Counterfactual impact evaluation

quality". The same cited authors proposed a general conceptual framework structured in four categories, each with three to seven dimensions, as seen in the table below.

Table 1 Data quality categories and dimensions

Categories	Accuracy	Relevance	Representation	Accessibility
Quality dimensions	Trust	Added value	Interpretability	Accessibility
	Accuracy	Relevance	Understanding	Cost-
	Objectivity	Timeliness	ease	effectiveness
	Completion	Operation ease	Representation	Access
	Traceability	Adequacy of	consistency	security
	Reputation	data volume	Conciseness	
	Variety of sources	Flexibility		

Source: Wang & Strong, 2017, p. 16

Similar dimensions of data quality, such as accuracy, timeliness, accessibility, relevance, completeness, consistency, reliability, are found in methodologies customised to sectoral policies (World Health Organization, 2017). They are grouped into more comprehensive dimensions, such as would be the following.

- Completeness and timeliness of data.
- Internal consistency of the reported data aiming at the coherence of the data for the same indicator in a time interval.
- Coherence of the data across indicators.
- Consistency of data reported with original data indicating the degree of accuracy.
- External consistency of data refers to the consistency of data from different sources.
- Adequacy of population data used for derived indicators.

4. Main constraints in the ESIF evaluations due to data accessibility and perspectives

A major difficulty is reported in research, evaluations and studies, the operation of data falling under the regulations to protect personal data, limiting accessibility to available data. Regulation 2016/679 on the protection of individuals concerning the processing of personal data and on the free movement (General Data Protection Regulation, GDPR), Directive 95/46/EC after long debates changed the data protection landscape. The implementation of the GDPR highlights at least three tacit innovative

elements: extended jurisdiction over all entities processing personal data of EU residents, stringent obligations and responsibilities of processors and controllers of personal data and remedial measures, liability and penalties for breaches of the Regulation (Crato & Paruolo, 2019). The data collection and processing for evaluations were deeply affected, including primary data and the use of secondary data. In the first case, final beneficiaries are reluctant to offer information and participate in surveys. In contrast, in the case of administrative data, more administrative work is required to comply with GDPR provisions. "The feasibility of merging administrative data, often held by different agencies, is limited, being subject to countries' and regions' data protection laws" (Elia et al., 2015, p. 7).

Trivellato (2019) shows that the importance of using microdata in scientific research in the social sciences and public policy evaluation is growing, and the attribute of "public good" increasingly characterises microdata. This opinion is based on the fact that the distinguishing feature is the collective use of individual data and on the replicability of data so that researchers must be able to resume analyses and challenge previous results by extending the research.

The evaluation model and, subsequently, the tools, including the statistical models, are influenced by the availability of data (Rettore & Trivellato, 2019). For both the beneficiaries and the non-beneficiaries, the quantitative methods need data on their characteristics, pre- and post-intervention variables. The administrative data collected for certain groups do not always meet the evaluation needs because their collection and storage pursued the objectives for which those administrative registers were created. Programmes sometimes operate with large data sets, involving hundreds of thousands of individuals who have been involved in programs, whose situation, before and after the intervention, needs to be modelled and translated into variables with accessible indicators and data.

The intervention in education, aiming at preventing early school leaving, provide support to large populations of children and students, the expected effect at the proximal level being the behaviour change to continue studies across educational levels and improve school performance. In contrast, induced effects are measured at the school or territorial level in terms of lower dropout rate or increased participation.

In the absence of monitoring systems able to record participation in education at the individual level, the evaluation could rely only on administrative data at the school territorial or sectoral level, capturing variables of interest. Children and young students' data raise many difficulties regarding personal data protection, hardly to be resolved in an

evaluation exercise. Therefore, well-designed systems must combine administrative data with more specific data on the beneficiaries and control groups for the counterfactual evaluations. Thus, ideally, microdata at the participant level pre-, during and post-intervention is needed. Complementarily or alternatively, if microdata at individual level are not available, the impact can be estimated at the level of schools or local communities (Lucaciu et al., 2021b).

The practice in evaluation shows the evaluation model is adjusted, from the desired model, given by the theoretical mechanisms of effects generation, to the feasible model, highlighting the limitations and their influence on the results of the analyses. The availability and accessibility of data have improved with the extension of Big Data and Open Data systems, access to public administrative data has improved significantly (Rettore & Trivellato, 2019).

Another example concerns people from vulnerable groups whose post-intervention situation is not always possible to follow without the responsibility of a public authority. Thus, the evaluation model adjusts, from the desired model, given by the theoretical mechanisms of producing the effects to the feasible model, highlighting the limitations and their influence on the outcomes. The availability and accessibility of data already have beneficial effects on quantitative assessments, access to public administrative data has improved significantly (Rettore & Trivellato, 2019).

A model of a quality monitoring and evaluation system applied in higher education (Li et al., 2017) represents a good basis for developing a system tailored to the needs of the impact assessment of education policies. The model (Figure 2) combines the characteristics of the monitoring and evaluation system with a Big Data processing platform, forming a multi-functional system including functions of data acquisition, processing and use of results. The data collected could be adjusted to the variables of interest of other interventions, while the reports generation could be tailored to the monitoring and evaluation needs.

If Big Data and their public access through Open Data platforms are a big step forward increasing accessibility to data needed in quantitative evaluation and research, blockchain technology opens a new perspective.

Interest in the blockchain is growing. Governments and private organisations are investing heavily in large Big Data and blockchain technologies due to their potential to solve real-world problems.

"Big Data" has generated a strong interest in various research and engineering fields in recent years. Despite the many advantages and applications, many challenges are to be solved for better quality, processing,

management, confidentiality, and data security. Blockchain, benefitting of inner advantages in terms of decentralisation and security, has a great potential to improve services and applications with large data volumes (Deepa et al., 2021).

Figure 2. The architecture of the Higher Education Quality Monitoring and Evaluation System

Source: Li et al., 2017, p. 339

The relevance of blockchain to public policy research needs is given by its nature as a technology of ledgers, given that administrative ledgers are the main sources of data for evaluation. Ledgers in blockchain technology are publicly available for everyone to inspect data blocks and transactions. The advantage is that users stay anonymous, identifying only using their own key as an address (Karafiloski & Mishev, 2017). In addition, blockchain offers advantages in terms of accessibility through decentralisation, operation with large volumes of data, real-time processing, confidentiality, and data security, for public policy evaluation, the advantages refer to surveys for collecting primary data from beneficiaries. Blockchain

technology can reduce data privacy and security concerns that affect survey response rates, especially online.

5. Conclusions

The variables and the indicators enable analysts and evaluators to describe and analyse the complex reality of the intervention. One difficult task of the evaluators is to select the appropriate variables, ensuring best modelling of reality and satisfying the needs of the evaluation methods, including quantitative necessary to answer the evaluation questions and provide robust and credible evidence.

A good understanding of the dimensions of the planned change enables the analysts to select the variables rigorously and define the indicators and measurement tools adequate for a large range of methods, including the quantitative ones.

The results of our research emphasise that evaluations operate with data with different levels of granularity, as required by the intervention logic. Therefore, structuring data in clusters and categories is useful in assessing data quality and their adequacy to the methodological needs, and tested models are useful for researchers and practitioners. A largely applicable tool for quality assessment is structured on four categories of dimensions: accuracy, relevance, representation, and accessibility. This may reflect the key constraints of accessibility and impose adjustments of the evaluation model from desirable to feasible. In line with the existing literature, we demonstrate that data accessibility is a key constraint for ESIF evaluation and impose adjustment of the desired evaluation model to a feasible one.

The study also reveals the need for evaluators and analysts to complement their skills and master the typology of variables, indicators, and data to find the best and feasible models. There are no prescriptive rules regarding how many variables or what specific variables to select in a model. The experience and effort of the analysts will help optimising trade-offs between the number or type of variables selected and the model's features.

Limited access to data is a key constraint in extending the use of quantitative impact assessment methods. In this sense, we suggest the use of Big Data and Open Data systems. These have significantly improved the data and the evaluations' potential to use extensively quantitative methods, despite the constraints generated by personal data protection regulations and limited administrative data interconnections.

As a ledger technology with default features related to decentralisation and security, blockchain is expected to bring large benefits

to evaluation. For evaluators and policymakers, blockchain potential is an area of further research looking for additional advantages that could enhance the use of quantitative methods. Therefore, our results may be of use for overcoming the challenges raised in the practice of ESIF evaluation during the programming period 2021-2027.

References

- Befani, B., Barnett, C., & Stern, E. (2014). Introduction–Rethinking impact evaluation for development. *IDS Bulletin*, 45(6), 1-5.
<https://doi.org/10.1111/1759-5436.12108>
- Berman, E., & Wang, X. (2012). *Essential statistics for public managers and policy analysts*. Sage Publications.
- Carli, R., Dotoli, M., Pellegrino, R., & Ranieri, L. (2013). Measuring and Managing the Smartness of Cities: A Framework for Classifying Performance Indicators. *2013 IEEE International Conference On Systems, Man, And Cybernetics*. <https://doi.org/10.1109/smc.2013.223>
- Conn, V. (2017). Don't Rock the Analytical Boat: Correlation Is Not Causation. *Western Journal Of Nursing Research*, 39(6), 731-732.
<https://doi.org/10.1177/0193945917701090>
- Crato, N., & Paruolo, P. (2019). The Power of microdata. In N. Crato, & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation. How access to microdata is transforming policy design*. (p. 1-14). Springer. <http://doi.org/10.1007/978-3-319-78461-8>
- Deepa, N., Pham, Q., Nguyen, D., Bhattacharya, S., Prabadevi, B., & Gadekallu, T. et al. (2021). *A Survey on Blockchain for Big Data: Approaches, Opportunities, and Future Directions*. arXiv.org. <https://arxiv.org/abs/2009.00858>
- Elia, L., Santangelo, G., & Schnepf, S. (2015). *Synthesis report on the 'Pilot projects to carry out ESF related counterfactual impact evaluations*. Directorate-General for Employment, Social Affairs and Inclusion (European Commission), Joint Research Centre (European Commission)
<https://op.europa.eu/en/publication-detail/-/publication/3d8a3354-084d-11e6-b713-01aa75ed71a1/language-en>
- European Commission. (2014, January). *Monitoring and evaluation of European Cohesion Policy. European Regional Development Fund, European Social Fund, European Cohesion Fund. Guidance document on ex-ante evaluation*.
https://ec.europa.eu/regional_policy/sources/docoffic/2014/working/ex_ante_en.pdf
- European Commission. (2019). *Guidelines on linking planning/programming, monitoring and evaluation* (2nd ed.). European Commission.
<https://ec.europa.eu/neighbourhood->

- [enlargement/sites/default/files/dg_near_guidance_note_-_addressing_capacity_development_in_programming_me.pdf](#)
- European Commission. (2021). *Evaluation*.
https://ec.europa.eu/regional_policy/en/policy/evaluations/member-states/
- Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New directions for evaluation*, 2009(122), 75-84.
<https://doi.org/10.1002/ev.297>
- Field-Fote, E. (2019). Mediators and Moderators, Confounders and Covariates: Exploring the Variables That Illuminate or Obscure the “Active Ingredients” in Neurorehabilitation. *Journal Of Neurologic Physical Therapy*, 43(2), 83-84. <https://doi.org/10.1097/npt.0000000000000275>
- Gertler, P., Martinez, S., Premand, P., Rawlings, L., & Vermeersch, C. (2016). *Impact Evaluation in Practice*, (2nd ed.). World Bank.
- Hujer, R., Caliendo, M., & Radic, D. (2004). Methods and limitations of evaluation and impact research. *Found Eval Impact Res (CEDEFOP Ref Ser 58)*, 30, 131-90. https://www.cedefop.europa.eu/files/BgR1_Hujer.pdf
- Karafiloski, E., & Mishev, A. (2017). Blockchain solutions for big data challenges: A literature review. *IEEE EUROCON 2017 – 17th International Conference On Smart Technologies*. <https://doi.org/10.1109/eurocon.2017.8011213>
- Li, Y., Li, P., Zhu, F., & Wang, R. (2017). Design of higher education quality monitoring and evaluation platform based on big data. *12Th International Conference On Computer Science And Education (ICCSE)*.
<https://doi.org/10.1109/iccse.2017.8085513>
- Lucaciu, L. (2018). A Look at the Evaluation Framework for Smart Growth Programmes. *Revista Romaneasca Pentru Educatie Multidimensionala*, 10(3), 60-76. <https://doi.org/10.18662/rrem/63>
- Lucaciu, L., Roman, M., Rus, S., Florian, B., Vasile, M., & Mariş, C. (2020). *Evaluarea retrospectivă a intervențiilor POSDRU 2007-2013 în domeniul educației*. București: Ministerul Investițiilor și Proiectelor Europene.
- Lucaciu, L., Roman, M., Rus, S., Florian, B., Vasile, M., & Mariş, C. (2021a). *Prima evaluare a intervențiilor POCU 2014-2020 în domeniul incluziunii sociale* [The first evaluation of the POCU interventions 2014-2020 in the field of social inclusion]. București: Ministerul Investițiilor și Proiectelor Europene.
- Lucaciu, L., Roman, M., Savin, A., Ionescu, D., Bulai, A., Rus, S., Florian, B., Vasile, M., Maris, C., Buhaescu-Ciuca, S., Savin, A., Platon, G., Rosu, C., Perianu, E., Dalu, A.-M., Haj, M. C., Manoleli-Preda, D., Florescu, A.-E., Baci, A., Jurj, A., Florina, T. (2021b). *Prima evaluare a intervențiilor POCU 2014-2020 în domeniul educației* [The first evaluation of the POCU interventions 2014-2020 in the field of education]. Bucharest: Ministerul

- Investițiilor și Proiectelor Europene. <http://www.anc.edu.ro/wp-content/uploads/2021/04/Primul-raport-de-evaluare-a-interven%C8%9Biilor-POCU-2014-2020-%C3%AEn-domeniul-educa%C8%9Biei-1.pdf>
- Morris, S., Tödtling-Schönhofer, H., & Wiseman, M. (2013). *Design and commissioning of counterfactual impact evaluations*. <https://op.europa.eu/en/publication-detail/-/publication/f879a9c1-4e50-4a7b-954c-9a88d1be369c>
- Popescu, M., & Roman, M. (2018). Vocational training and employability: Evaluation evidence from Romania. *Evaluation And Program Planning*, 67, 38-46. <https://doi.org/10.1016/j.evalprogplan.2017.11.001>
- Rettore, E., & Trivellato, U. (2019). The Use of Administrative Data to Evaluate the Impact of Active Labor Market Policies: The Case of the Italian Liste di Mobilità. In N. Crato & P. Paruolo, *Data-driven policy impact evaluation* (1st ed., pp. 165-182). Springer Open
- Roman, M., & Goțiu (Lucaci), L. (2017). Non-parametric methods applied in the efficiency analysis of European structural funding in Romania. *Journal Of Applied Quantitative Methods*, 18(2). https://mpra.ub.uni-muenchen.de/80548/1/MPPA_paper_80548.pdf
- Scott-Phillips, T., Dickins, T., & West, S. (2011). Evolutionary Theory and the Ultimate–Proximate Distinction in the Human Behavioral Sciences. *Perspectives On Psychological Science*, 6(1), 38-47. <https://doi.org/10.1177/1745691610393528>
- Silverman, D. (2011). *Interpreting qualitative data. A Guide to the Principles of Qualitative Research*. Sage
- Trivellato, U. (2019). Microdata for Social Sciences and Policy Evaluation as a Public Good. In N. Crato & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation* (pp. 27-45). Springer Open.
- United Nations. (2019). *National Quality Assurance Frameworks Manual for Official Statistics*. United Nations Department of Economic and Social Affairs Statistics Division. <https://unstats.un.org/unsd/methodology/dataquality/references/1902216-UNNQAFManual-WEB.pdf>
- Vaessen, J. (2011). Challenges in impact evaluation of development interventions: randomised experiments and complexity. *Evaluating the complex: attribution, contribution, and beyond. comparative policy evaluation*, 18, 283-313. <https://medialibrary.uantwerpen.be/oldcontent/container2143/files/Publications/DP/2010/01-Vaessen.pdf>
- Wang, R., & Strong, D. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal Of Management Information Systems*, 12(4), 5-33. <https://doi.org/10.1080/07421222.1996.11518099>

- White, H. (2014). Current challenges in impact evaluation. *The European Journal of Development Research*, 26(1), 18-30. <https://doi.org/10.1057/ejdr.2013.45>
- Williams, B. (2015). Prosaic or profound? The adoption of systems ideas by impact evaluation. *IDS Bulletin*, 46(1), 7-16. <https://doi.org/10.1111/1759-5436.12117>
- World Health Organization. (2017). *Data quality review: module 1: framework and metrics*. <https://apps.who.int/iris/handle/10665/259224>