The Potential of Administrative Microdata for Better Policy-Making in Europe

Sven Langedijk, Ian Vollbracht, and Paolo Paruolo

1 Introduction

This chapter looks at the future of evidence-based policy-making in Europe. It takes a bird's-eye view and it uses a stylised approach in tackling the question 'What might policy-making and policy evaluation look like in a hypothetical world of perfect availability of administrative microdata?' It reviews possible answers to this question, as well as related benefits and pitfalls.

Evidence-based policy has, since the 1940s and 1950s, been associated with the rise of empirical social sciences such as sociology, economics, political science and social psychology (Head 2015). By the 1970s, leading social scientists increasingly advocated the importance of rigorous behavioural and experimental methods. Analysis of quantitative data from social experiments was advocated, as was the application of advanced analytical methods to data collected by passive observation; these methods later evolved into the set of tools currently known as quasi-experimental methods. By the beginning of the new millennium, the movement had become known as the 'evidence-based policy' movement (Panhans and Singleton 2017).

While this trend was heterogeneous across regions of the world, in recent decades many government-spending programmes have increasingly been evaluated using quantitative methods. The objective is to better determine what works, and to enhance the efficiency and effectiveness of the programmes in question. These evaluations have often been based on programme data and specifically targeted surveys.

S. Langedijk (✉) · I. Vollbracht · P. Paruolo
Joint Research Centre, Ispra, VA, Italy
e-mail: sven.langedijk@ec.europa.eu; ian.vollbracht@ec.europa.eu; paolo.paruolo@ec.europa.eu

In a typical experimental evaluation design, a baseline survey is run before the policy intervention, and a follow-up survey is repeated to measure the outcomes for the treated and the control groups. The usual limitations of attrition bias, enumerator bias, reporting bias,[1] etc. apply to survey data; while a few of these sources of bias may remain when using administrative data (see Feeney et al. 2015), recourse to such data can greatly reduce them.

In the European Union (EU), the European Commission has embraced the idea of evaluation for spending programmes in the first decade of the new millennium (Stern 2009). The EU began by setting up evaluation requirements for centrally managed spending programmes, and joint evaluation schemes with authorities in the member states for spending programmes managed and executed at national or regional levels. Examples of the latter are the European Social Fund and the European Regional Development Fund.

The Commission has gradually expanded the scope and coverage of its policy evaluations to regulatory and legislative policies. The Juncker Commission made better regulation one of its core goals, and in 2015 published new guidelines for ex post policy evaluation, with the aim of improving policies, reducing administrative burdens, strengthening accountable government and supporting strategic decision-making (Mastenbroek et al. 2015).

Policy evaluations at first had limited quantification, relying mostly on data aggregated at country level, sometimes combined with targeted ex post surveys of beneficiaries and stakeholders. This often limited the ability to make causal evaluation of what worked and what did not.

Although evaluation techniques and the potential comprehensiveness of data continue to improve, even today, data (un)availability remains the key limiting factor preventing the widespread use of more rigorous approaches. At the same time, public authorities are sitting on a treasure trove of administrative data collected and used for other purposes, such as social security, taxation, education, communal administration, cadastres and company registration.

If these data could be effectively reused, then econometric and statistical techniques would allow disaggregation of the analysis along many dimensions, such as geographical, social and firm level. For socio-economic policies, enhanced data availability could allow policy evaluation centred on the life course of citizens (Gluckman 2017). Detailed and accurate evidence on 'what works' would allow a step change in the quality of public services and legislation.

The rest of the chapter is organised as follows. Section 2 considers current trends in evaluation, data gathering and storage, automation of data collection processes, and increased processing and storage capacity. At the end of the section, a number of simplifying assumptions on the extrapolation of current trends are made. Section

---

[1]There may be incentives for either the subject or the collector of data (or both) to misreport data. For example, an individual may be incentivised to underreport income in an application for social welfare services.

3 illustrates the stylised aspects of a world with perfectly functioning access to microdata for policy research. This abstraction is both useful and valid in relation to making relevant choices for future developments; it can help to get a conceptual grip on the direction in which current trends allow policy-makers to steer evidence-based policy-making. Section 4 looks at the potential pitfalls, while Sect. 5 presents concluding remarks.

2 Trends in Data and Policy and Programme Evaluation

This section describes current trends in microdata availability that may steer future developments in evaluation. The last part of the section presents a set of simplifying assumptions on the continuation of these trends; these assumptions are maintained in the rest of the chapter.

2.1 *Increased Availability of (Micro)data*

The world is producing increasing amounts of data. A key indicator of increased data production and use by private individuals is, for instance, the trend in global IP (internet) traffic. The industry predicts that this will increase threefold between 2016 and 2021, with the number of devices connected to IP networks at three times the global population by 2021. The Internet of Everything phenomenon, in which people, processes, data and things connect to the Internet and each other, is predicted to show impressive growth; globally, machine-to-machine connections are expected to increase by 140% over the period 2016–2021 (Cisco 2017).

There have been large strides forward in the past two decades with regard to computing power, data storage capacity, analytical techniques and algorithm development. These trends, together with a massive increase in the use of devices connected to the Internet by private citizens, have allowed big tech companies such as Amazon and Google to expand at a dramatic rate. While there is little doubt about the benefits that these innovations have delivered in terms of choice, speed and access to information, citizens' concerns about data privacy and security have, in parallel, become much more visible issues in public policy discourse.[2]

In the governmental sphere, the use (for evaluation and policy research in particular) and the potential interlinkage of administrative data held by public institutions have moved forward at a much slower rate.[3] This may reflect some combination of inertia and data security and privacy concerns. However, some steps forward have been achieved, as illustrated below.

---

[2]The recent EU General Data Protection Regulation, see https://www.eugdpr.org, is addressing the issue of protection of personal information.

[3]Discussion of the work of signal intelligence agencies is beyond the scope of this chapter.

2.2 Administrative Data Linkage Centres

Over the past decade, centres for linking and granting access to microdata to researchers have been set up in a number of countries. The Jameel Poverty Action Lab (J-PAL) North America, based at the Massachusetts Institute of Technology, has compiled a catalogue of administrative datasets available in the United States.[4] In Europe, most countries provide (limited) access to microdata, with some states providing linking services. Statistics Netherlands,[5] for instance, provides linking services to researchers in the Netherlands and other EU countries.

In the United Kingdom, the government's Economic and Social Research Council funded the Administrative Data Research Network (ADRN), with an initial funding period spanning 2013–2018.[6] Other examples of data access and linkage centres are at Germany's *Institut für Arbeitsmarkt-und Berufsforschung* (Institute for Employment Research) (IAB) (see Chap. 7) and New Zealand's Integrated Data Infrastructure, see Gendall et al. (2018).

2.3 Trends in Economic Publications

The current increased availability of microdata is moving the focus of economic research, which has become increasingly centred on data analysis (Hamermesh 2013). A recent analysis of fields and styles in publication in economics by Angrist et al. (2017) documents a shift in publications and citations towards empirical work, where the empirical citation share is now at around 50%, when compared with the two alternative categories of 'economic theory' and 'econometrics'.

Angrist and Pischke (2017) call for a parallel revision of undergraduate teaching in econometrics, which they argue should be more focused on causal questions and empirical examples, should have a controlled-experiment statistical framework as reference and an emphasis on quasi-experimental tools.

Panhans and Singleton (2017) document the rise of empirical papers and of quasi-experimental methods in economics. They track the citations of quasi-experimental methods in the major economic journals. A similar interrogation of Google Scholar for the words 'counterfactual' or 'counterfactual impact evaluation'

---

[4]See https://www.povertyactionlab.org/admindatacatalog. All links and Internet searches in this chapter were accessed or processed in November 2017.

[5]For microdata at Statistics Netherlands, see https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research.

[6]The ADRN (https://adrn.ac.uk/) consists of: a coordinating body (the Administrative Data Service); four Administrative Data Research Centres, one in each country in the United Kingdom; national statistics authorities, government departments and agencies (the data providers), the Economic and Social Research Council (the funders), and the UK Statistics Authority (which leads the network's board).
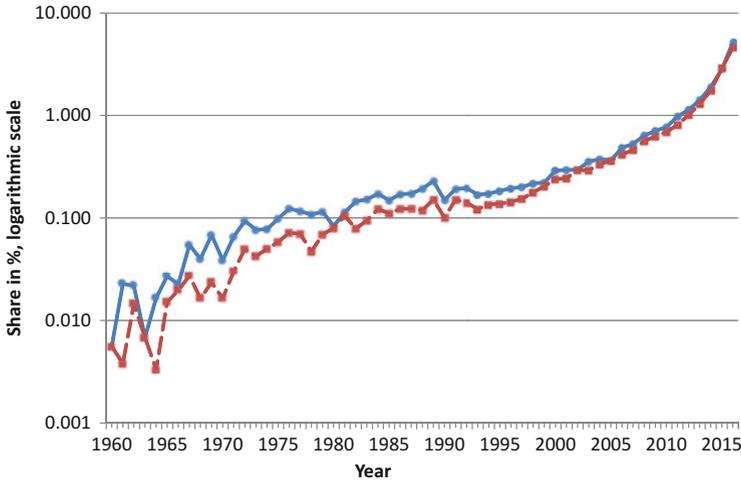
Fig. 1 Percentage citations of 'counterfactual' and of 'counterfactual impact evaluation' in economics. The percentages shown on the vertical axis are 100 *a/c* (solid line) and 100 *b/c* (dashed line), where *a*, *b* and *c* are the number of hits for the queries 'counterfactual AND economics', 'counterfactual impact evaluation AND economics' and 'economics', respectively

with 'economics' illustrates the increased incidence of counterfactual methods within economics (Fig. 1).

There is still a debate internal to economics on how reliable quasi-experimental methods are with respect to controlled experiments, which use randomisation (Bertrand et al. 2004). This is partly reflected in the idea of classifying studies according to 'strength of evidence' using the Maryland scale, which is often used to evaluate the evidence in justice cases (Sherman et al. 1997, 1998), and its modification by the UK What Works Network (Madaleno and Waights 2015).

In the EU policy context, and in particular in the assessment of regulatory policies, counterfactual impact evaluation methods are still rather novel. Limitations, heterogeneity of study designs and differences across areas and countries in administrative data access are factors that have limited the speed at which these methods have been introduced into the policy cycle.

To imagine how methods and policy applications might develop, it is useful to extrapolate trends in data gathering, access, linking, storage and availability. To this end, the following simplifying assumptions are made:

1. Microdata will cover all economic and social fields, and will be instantaneously updated.
2. Datasets will be linked, anonymised and made available to the research community.
3. (Unique) identifiers will allow seamless data linkage across policy areas (e.g. health, education, taxation).

4. Personal data protection will be fully ensured through effective legislation and oversight.
5. Micro-datasets will be available internationally in a fully comparable manner (e.g. pan-EU).

These assumptions may not necessarily appear realistic at present; however, they provide a useful framework for exploring policy opportunities and risks.

3 Stylised Aspects of a World with Perfect Administrative Microdata Availability

This section explores the implications of more, better and faster data in shaping policies, under the above assumptions (1. to 5.) reflecting the continuation of existing trends.

3.1 Breaking Silos: Multidimensional and Integrated Policy Design

Social and economic reality is shaped by the complex interplay of many factors reflecting numerous causal relations due to the actions of individual agents and groups. To a very significant degree, many policy-makers think in terms of isolated sectors, i.e. 'in silos'. Here, an integrated approach, trying to understand the entire economic and social structure of an economy across sectors, can be contrasted with a pragmatic one, attempting to carry out intelligent policy analysis at a sectoral level.

The 'silo culture' is in no small part due to the data and information that policy-makers typically receive. Officials working on health policy, for example, generally read reports providing information and analysis on the workings of the health system for which they are responsible. Similarly, education departments tend to look at data from schools and universities, and tax collectors focus on company accounts and personal tax returns. The data used to evaluate these organisations, in turn, are based on distinct 'health', 'education' and 'revenue' metrics, respectively, which creates incentives to maintain the silo approach to policy.

Figure 2 shows this diagrammatically, using three traditional areas of public policy: education, health and tax collection. It exemplifies some arbitrarily chosen interlinkages between relevant variables in each silo. (Each link is represented by a double-headed arrow, even when there can be a one-sided causal link; this is to simplify exposition.) For example, it is not *prima facie* controversial to assume
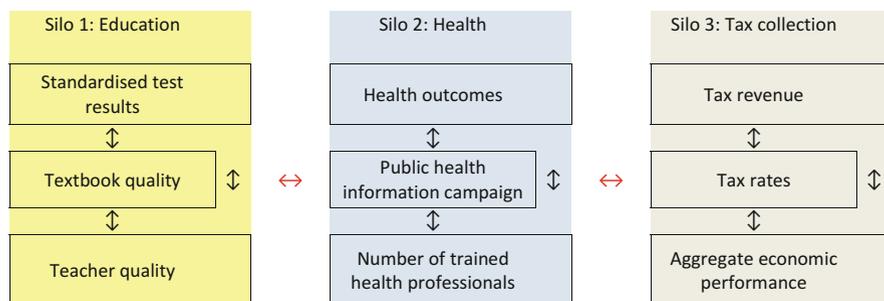
Silo 1: Education		Silo 2: Health		Silo 3: Tax collection	
Standardised test results		Health outcomes		Tax revenue	
↕		↕		↕	
Textbook quality	↕ ↔	Public health information campaign	↕ ↔	Tax rates	↕
↕		↕		↕	
Teacher quality		Number of trained health professionals		Aggregate economic performance	

Fig. 2 Vertical and horizontal policy links. A stylised example with three silos: education, health and tax collection. Links are represented by double-headed arrows (even when there can be a one-directional causal link) to simplify exposition. Vertical arrows are within-silo links and horizontal arrows represent links across silos

the existence of causal links between standardised test results, textbook quality and teacher quality. For instance, teacher quality and textbook quality may have an effect on standardised test results. Tax revenue depends significantly upon taxation rates and aggregate economic performance. These vertical linkages represent the within-silo links. Moreover, links can exist also across silos. For instance, overall tax revenue is important for funding expenditure on health and education; these links are indicated by the horizontal arrows.

The vertical, within-silo stratification of policy silo thinking is legitimate and has certain advantages. For example, while there are certainly interactions between education policy and health outcomes, the process of trying to quantify the linkages can be difficult. In the absence of abundant information about how different fields of government policy interrelate, there is a certain wisdom in focusing one's analysis in an area in which one has a certain level of expertise and understanding. However, under the assumption of perfect microdata availability, the silo mentality might be expected to attenuate over time. Subject to their availability, microdata that are easily linked across policy areas can potentially facilitate research into such cross-silo linkages.

One of the most important impacts of more readily available micro-datasets might then be a cultural shift in terms of how policy is designed and evaluated. If looking at data from the citizens' perspective—rather than the policy perspective—were to become predominant, then one might expect a more integrated approach to policy-making to develop over time.

While the speed at which changes in policy-making will occur is uncertain, the direction of the impact of greater availability of microdata is relatively clear. More integrated policy design should be the result, and it will hopefully drive better targeted and more effective public policy interventions. Indeed, such a 'cross-silo' approach is advocated by Gluckman (2017).

3.2 Ever More Precise Proxies for Citizen Well-Being

While the ultimate goal of public policy should surely be that of enhancing human well-being, data availability has been a major constraint on measuring the impact of government policies throughout history.[7] For example, in spite of a growing body of evidence that happiness and well-being are far from perfectly associated with wealth beyond some modest thresholds (Layard 2005), possibly due to the lack of better statistical proxies, government policies have focused mostly on measurable monetary and aggregate growth objectives.

Complete and linked microdata should allow better setting of policy objectives and targets, better policy design and better measurement of policy outcomes in terms of the ultimate objective of citizen well-being. Better proxies for well-being can first be determined on a range of factors including health, social and economic engagement, access to green areas, etc.

3.3 Reducing Evaluation Lags in the Policy Cycle and Adjusting Policies 'On the Go'

At present, both spending and regulatory policy evaluations are often available only many years after implementation. In many cases, quantitative evidence is missing, or based on ex post surveys, which are subject to well-known shortcomings. Where rigorous causal evaluation is undertaken, counterfactual impact evaluation methods are used. In practice, in the felicitous cases where good data are available, the most time-intensive part of this work is usually associated with obtaining the relevant datasets, and cleaning and linking the data across data sources, leaving very little time for analysis.

Subsequent econometric analysis, report writing and review procedures are also time consuming. With perfectly available microdata, the time lag from implementation to evaluation ought to be massively reduced. An evaluation strategy could be designed upfront, such that results would be available shortly after implementation. Depending on the type of policy, this could sometimes even be in near-real time.

This would further allow policy-makers to design policies conditional on intended outcomes (both on desired targets and on side effects), such that policy adjustment and fine-tuning could—within limits—become semi-automatic after a given probation or review period. *De facto*, in many cases, monitoring

---

[7] 'Whatever the form or Constitution of Government may be, it ought to have no other object than the general happiness' (Thomas Paine, *The Rights of Man*). The US Declaration of Independence of 1776 specifies 'the pursuit of happiness' as one of the principal inalienable rights of all citizens. The 1942 Beveridge Report in the United Kingdom, which laid the groundwork for Britain's welfare state, referred to 'the happiness of the common man' as the basic objective. The above citations are taken from Duncan (2010).

and evaluation frameworks could be merged. All this would have profound consequences for the duration of the policy cycle, and the effectiveness and efficiency of policies.

3.4 Reducing Other Lags in the Policy Cycle

3.4.1 Need for Policy

At present, the process of problem (or needs) identification in public policy typically arises from a government or institutional agenda, and/or from popular concern as expressed in the media and/or by civil society groups. Where an administration is tasked with designing in detail a policy measure to respond to this demand for action, the first phase is typically problem identification. This might take in the order of 1–2 years, to pull together the necessary data and analyse them and their evolution over time. With perfectly available microdata, this 'outside lag' ought to be cut dramatically, and perhaps to as little as a few months.

There would, in theory at least, be far less scope for controversy—and therefore far less need for analytical refinement—if microdata could instantaneously deliver a 'clear and transparent picture' of the *status quo ante* in a particular field of policy.

Moreover, open (but privacy-safeguarding) access to linked data for policy researchers could even trigger crowdsourcing of such analysis. For example, experimental opening of a linked labour market dataset at the IAB has led to a large number of policy research papers of the highest quality. This has placed the IAB in the top 6% of economic institutions as of September 2017 (for more details, see Chap. 7).[8]

However, if the use of such techniques is to become more widespread, they will have to move in step with measures that respond to citizens' fully legitimate concerns about personal data privacy and security, which are discussed in Sect. 4.

3.4.2 Policy Design and Consultation

At present, policy design and consultation for policy areas within the competence of the EU are the subject of a structured process implementing the Better Regulation Agenda.[9] Feedback is sought from interested stakeholders and impact assessments are prepared. As a rule of thumb, this process might typically take 1–2 years. Policy design and consultation processes are fundamentally human, democratic interactions and should certainly not be made subject to full automation simply because of the instantaneous availability of microdata.

---

[8]See https://ideas.repec.org/top/top.inst.all.html.

[9]See https://ec.europa.eu/info/law/law-making-process/better-regulation-why-and-how_en.

However, if such data were to become more widely available, there would be great potential to foster a better informed level of policy debate; see Spiegelhalter (2017). Robust evidence on policy impact would become more available, and at a more rapid rate. Enhanced knowledge about what works and for whom gained from fully linked data should facilitate enhanced policy research. The consultation and engagement process could then focus more on impacts that have been overlooked in past analyses. This 'inside lag' in policy design and consultation could therefore be shortened, to some degree, by better informed debates.

For the sake of completeness, as regards the legislative process, the main benefit would be through better informed discussions based on more robust evidence, rather than any significant time saving. Democratic due process would not necessarily be sped up by the availability of microdata.

3.5   The Potential of Administrative Data for Better Policy

Under the given assumptions, one would therefore expect to see a very considerable reduction in time lags in the policy cycle, from the identification of a policy problem to evaluating the policy's impact. Moreover, the availability of real-time administrative microdata would probably encourage processes of near-real-time policy adaptation. For example, one might imagine that policy redesign might be directly incorporated into some sort of dynamic evaluation process even during the course of the implementation phase. In democratic systems of governance, the challenge here is to ensure that more rapid policy analysis and adaptation, through improved administrative data availability, foster better informed policy design and consultation procedures. In this way, policy legitimacy can be ensured.

Summarising, both spending and regulatory policies could benefit from administrative data in a number of ways: (1) breaking silos to better incorporate interactions across policy areas; (2) allowing policy decisions and adaptation based on more robust evidence on what works for whom; (3) better measuring of impacts at individual or disaggregated group levels, reflecting distributions across income, age, ethnicity, location, etc.; and (4) much more efficiently and effectively targeting policy's ultimate objective of increasing citizen well-being.

4   Avoiding the Pitfalls

The potential benefits of perfect microdata for policy-making are great. Progress towards perfect microdata for policy could take policy-making to a different level. However, several threats are evident from both public and private sources. A number of obvious risks are considered in turn below, indicating how they could best be mitigated.

4.1   Data Privacy and Security

Perhaps the most obvious threat comes precisely from the risk of its de-anonymisation of microdata. Perfectly linked microdata allow the researcher to create an entire life narrative according to an individual's full set of recorded interactions with the organs of the state, which is, from a research point of view, a goldmine. From a privacy point of view, it is a potential danger, both at a personal and at governmental level; the concern is that data could be de-anonymised and released into the public domain. On this, all democratic countries have developed some level of right to privacy in their legal structure. This is certainly the case in the EU, where personal data protection has been significantly enhanced in recent years.[10]

Data security is an important issue around the globe. The fact is that, in contrast to the stylised assumptions in this chapter, data protection standards vary across different parts of the globe, as do levels of awareness among individual citizens about how to protect themselves from data theft. Moreover, hacks and data 'leaks' are a daily reality,[11] and some governments are themselves in the business of seeking to obtain personal data from other countries by illicit means.

Against this backdrop, citizens of any jurisdiction may not be entirely comfortable with perfectly linked and perfectly anonymised datasets being widely available to researchers around the world. Safeguards can, however, be implemented in the short term, along the following lines: (a) access to linked datasets can be limited to authorised researchers and officials; (b) work with those datasets can be restricted to controlled environments such as secure remote access, rather than the open Internet; and (c) access to data can further be limited to researchers and institutions that follow relevant procedures for democratic oversight and potential scrutiny of their activities.

These measures comply with the 'five safes' now common in this area: safe people (researchers can be trusted to use data appropriately), safe projects (research focuses on different populations of interest, and must be in the public interest), safe settings (data can be accessed in a secure environment), safe data (data is de-identified), safe output (research output must be confidentialised before it can be released); see Gendall et al. (2018).

Restricted access to microdata may in turn have implications for the development of open science and for academic peer-review procedures, as currently only a small number of researchers can replicate the work of selected peers. Moreover, international cooperation may also be curtailed to some extent by these data security concerns.[12] This appears to be one reason, for example, why some member states

---

[10]See http://ec.europa.eu/justice/data-protection/ and https://www.eugdpr.org/.

[11]See http://breachlevelindex.com/.

[12]The General Data Protection Regulation (GDPR) came into force in May 2018 and now sets the framework for the treatment of personal data throughout the EU; see https://www.eugdpr.org/.

of the EU are only at the earliest stages of discussions about granting access to non-national researchers.

4.2   Dislocation from Consultative and Legislative Due Process

As briefly discussed in Sect. 3, compressing many elements of the 'inside lag' in the currently conventional policy cycle will bring with it the challenge of ensuring that improved administrative data availability fosters better informed policy design and consultation procedures. As a result, some guidelines may become necessary to ensure enough time for society at large to engage with the policy implications brought about by faster policy research due to microdata availability.

4.3   Data Accuracy

An additional risk is simply a restatement of the well-known GIGO (garbage in, garbage out) phenomenon. Clearly, if linked micro-datasets became near-perfectly available and used in close-to-real-time policy adaptions, then any data imprecisions would be transmitted through the policy-making cycle at much higher speed. The implication is that even more attention will need to be paid to ensuring that data is as accurate as possible at the time of its being inputted into recording systems.

5   Concluding Remarks

The potential of using linked administrative microdata for better targeted policies in support of well-being appears to be very great, and possible associated pitfalls can be avoided. The likelihood of the assumptions on the availability to policy researchers of microdata being borne out in the future in Europe will depend on actions taken by public administrations, member states and the EU as a whole.

There are therefore strong arguments in favour of the increased use of administrative data to improve the quality and precision of impact evaluation and related public policy research. Use of public funds, such as EU strategic investment funds, could be envisaged for investment in this context. For this to happen, a shared set of objectives needs to be developed across the research, policy-making and data-holding communities. This will take time and will certainly need to take full account of concerns about data privacy and security.

Further refinements of the vision depicted in this chapter are desirable. Implementation of some of these ideas into legal frameworks and institutional processes will certainly require additional contributions from many stakeholders; this chapter attempts to get this discussion started.

References

Angrist JD, Pischke J-S (2017) Undergraduate econometrics instruction: through our classes, darkly. J Econ Perspect 31:125–144

Angrist J, Azoulay P, Ellison G et al (2017) Economic research evolves: fields and styles. Am Econ Rev 107:293–297

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? Q J Econ 119:249–275

Cisco (2017) The zettabyte era: trends and analysis. Cisco public white paper. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf. Accessed June 2017

Duncan G (2010) Should happiness-maximization be the goal of government? J Happiness Stud 11:163–178

Feeney L, Bauman J, Chabrier J et al (2015) Using administrative data for randomized evaluations. J-PAL North America, Boston. https://www.povertyactionlab.org/sites/default/files/resources/2017.02.07-Admin-Data-Guide.pdf

Gendall K, McDowell A, Blackburn A (2018) Linking data for better policy: New Zealand's integrated data infrastructure. Stats NZ, Wellington, New Zealand

Gluckman P (2017) Using evidence to inform social policy: the role of citizen-based analytics. Office of the NZ prime minister's chief science advisor. www.pmcsa.org.nz/wp-content/uploads/17-06-19-Citizen-based-analytics.pdf

Hamermesh S (2013) Six decades of top economics publishing: who and how? J Econ Lit 51:162–172

Head BW (2015) Policy analysis: evidence based policy-making. In: International encyclopedia of the social and behavioral sciences, 2nd edn. Elsevier, New York, pp 281–287

Layard R (2005) Happiness: lessons from a new science. Penguin Books, London

Madaleno M, Waights S (2015) Guide to scoring methods using the Maryland scientific methods scale. What Works Centre for Local Economic Growth. http://www.whatworksgrowth.org/public/files/Scoring-Guide.pdf

Mastenbroek E, van Voorst S, Meuwese A (2015) Closing the regulatory cycle? A meta evaluation of ex-post legislative evaluations by the European Commission. J Eur Publ Policy 23:1329–1348

Panhans MT, Singleton JD (2017) The empirical economist's toolkit: from models to methods. Hist Polit Econ 49(Suppl):127–157. https://doi.org/10.1215/00182702-4166299

Sherman LW, Gottfredson DC, MacKenzie DL et al (1997) Preventing crime: what works, what doesn't, what's promising. National Institute of Justice. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.130.6206&rep=rep1&type=pdf

Sherman LW, Gottfredson DC, MacKenzie DL et al (1998) Summary. National Institute of Justice Research in Brief. https://www.ncjrs.gov/pdffiles/171676.PDF

Spiegelhalter D (2017) Trust in numbers. J R Stat Soc A 180:948–965

Stern E (2009) Evaluation policy in the EU and its institutions. N Dir Eval 123:67–85

Sven Langedijk is the Head of the Modelling, Indicators and Impact Evaluation Unit of the European Commission at the Joint Research Centre in Ispra (IT). The unit's research and analytical activities cover the full range of EU's policy areas, including education, competitiveness, employment, innovation, the internal market, judiciary, health, poverty, migration, cohesion, and other

socio-economic dimensions. Previously, he headed the European Commission team responsible for developing, negotiating and monitoring the EU-IMF financial support and economic adjustment programme for Ireland. Until 2010, he contributed to the development of the EU fiscal governance framework and the Stability and Growth Pact and was editor of the Commission's annual flagship report on Public Finances in EMU.

Ian Vollbracht is the Deputy Head of the JRC unit Modelling, Indicators and Impact Evaluation. An economist and a researcher at the JRC, over the past 20 years he has worked in a variety of posts in the three main EU institutions: Commission, Council and Parliament. His work has mainly focused on economic, trade and budgetary policies. As a result of this varied career he has deep experience in the reality, as well as the theory, of policy development, implementation and evaluation within the EU institutions. Ian also has a deep interest in behavioural and psychological aspects of economics and policy-making. He is the co-author of the non-academic book *The Karmic Curve*, which seeks to bring some of these insights to a wider audience.

Paolo Paruolo is the Coordinator of the Competence Centre on Microeconomic Evaluation (CC-ME) at the European Commission—Joint Research Center, Ispra, IT. He has a Master in Economics and Statistics from the University of Bologna (1987) and a PhD in Mathematical Statistics (Theoretical Econometrics) from the University of Copenhagen (1995). He has taught Econometrics at the University of Bologna and at the University of Insubria (Varese, IT). His research interests are in econometrics (theory and practice) and in counterfactual methods. Because of his publication record, he was ranked among the best 150 econometricians worldwide in Baltagi (2007) "Worldwide econometrics rankings: 1989–2005", Econometric Theory 23, pp. 952–1012.