Measuring Public Administration: A Feasibility study for better comparative indicators in the EU

Written by Prof. Dr. Wouter Van Dooren
March 2018

Consortium for contract
European Commission
DG EMPL - VC/2016/0492
EUPACK-Project

Hertie School of Governance

Social Europe

EUROPEAN COMMISSION

Directorate-General for Employment, Social Affairs and Inclusion
Directorate F — Investment Unit F1 — ESF and FEAD: policy and legislation

Contact: EMPL-F1-UNIT@ec.europa.eu

European Commission
B-1049 Brussels

Measuring Public Administration: A Feasibility study for better comparative indicators in the EU

Directorate-General for Employment, Social Affairs and Inclusion
"Support for developing better country knowledge on public administration and institutional capacity building" (VC/2016/0492)

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

More information on the European Union is available on the Internet (http://www.europa.eu).

Luxembourg: Publications Office of the European Union, 2018

ISBN: 978-92-79-91249-8
doi:10.2767/819180

© European Union, 2018

Reproduction is authorised provided the source is acknowledged.

CONTENTS

1. EUPACK – PROJECT INTRODUCTION

Good governance and quality of public administrations is recognisably in the interests of the EU citizens and Member States, to achieve maximum value from finite public funds and create a public-private interface that raises employment and growth. Worldwide, the evidence is irrefutable: high productivity, high income per head economies have the most effective and efficient public institutions. The internal market cannot be completed, the EU acquis cannot be effectively implemented, and the goals of smart, inclusive and sustainable growth cannot be realistically achieved without good governance.

Member State administrations currently face the triple challenge of: delivering better with less - meeting societal & business needs in times of tighter budgets; adapting service provision to demographic, technological and societal changes; and improving the business climate through fewer and smarter regulations and better services in support of growth and competitiveness.

Experience in Europe in the past two decades shows different administrative reform paths and results[1] mainly due to different degree of reform capacity, sustainability of reform approach, coverage and a 'fitting context'. The incentives that triggered the "New Public management" wave of reforms in older Member States, addressed domestically recognised needs to reduce the size of government and make administration more efficient. Change has been rationalised through the accumulated management experience and exchange with peers. In new Member States, the "first wave" of reforms began with the EU-accession requirements[2] for establishing professional and depoliticised civil service systems. The limited internal capacity was compensated with externally managed support. Limited strategic orientation and ownership of reforms[3] led to mixed results[4].

Recently, the fiscal crisis has reinforced the relevance of public administration downsizing, outcome and result-orientation, and reduction of bureaucracy across Europe. Administrative culture however tends to produce important differences in the operationalization of these principles in management-oriented public administrations and in more legalistic

---

[1] Christopher Pollitt and Sorin Dan. 2011. COCOPS Policy Brief 1: The Impact of New Public Management (NPM) Reforms in Europe. http://www.cocops.eu/wp-content/uploads/2013/07/COCOPS_PolicyBrief_1_newlayout.pdf

[2] http://www.sigmaweb.org

[3] For more information see thematic evaluations of the PHARE programme.

[4] Meyer-Sahling, J. (2009), "Sustainability of Civil Service Reforms in Central and Eastern Europe Five Years After EU Accession", SIGMA Papers, No. 44, OECD Publishing; Also WB, Administrative capacity in the new EU member states : the limits of innovation? http://documents.worldbank.org/curated/en/2007/06/8187914/administrative-capacity-new-eu-member-states-limits-innovation

ones[5]. The need for quick results is another reason why on many occasions the focus is only on budgetary consolidation, cutting staff and salaries, instead of rethinking the scope of government and investing in the capacity of civil servants, as a basis for designing and delivering better quality of policies and services.

The EU has no specific competences in the administrative sphere but still has a strong indirect impact on the administrative practice in Member States through the administrative standards set in the acquis, the transfer of best practices with EU financial instruments, the promotion of management practices of its own institutions, etc.

Smart administration, development of human capital and related ICT of administrative and public services were seen as a fundamental requirement for economic growth and jobs already with the renewed Lisbon agenda. In response to the needs, in the 2007-2013 programming period[6] institutional capacity building became a key policy priority for the European Social Fund. The support was intended to go beyond the technical assistance for the better management of EU funds and assist the ongoing administrative reforms. Altogether, about EUR 2 billion of European Social Fund (ESF) and European Regional Development Fund (ERDF) were allocated to measures supporting the quality of public administration in 19 Member States.

In 2014-2020, the European Structural and Investment (ESI) Funds[7] should be the catalyst for achieving the objectives of the Union Strategy for smart, sustainable and inclusive growth. Through the European Semester[8] the European Commission undertakes every year a detailed analysis of EU Member States' programmes of economic and structural reforms and provides them with proposals for Council recommendations (Country Specific Recommendations, CSRs)[9] for the next 12-18 months. The ESI Funds will serve as an effective means to support the implementation of the CSRs. In 2014 some 20 Member States have received country specific recommendations (CSRs) in the area of public administration. 17 of them have programmed support to address the challenges under the specific thematic objective "*enhancing institutional*

---

[5] Gerhard Hammerschmid, Steven Van de Walle, Anca Oprisor and Vid Štimac. September 2013. COCOPS Policy Brief 4: Trends and Impact of Public Administration Reforms in Europe: Views and Experiences from Senior Public Sector Executives. see http://www.cocops.eu/wp-content/uploads/2013/10/Policy-brief-wp3.pdf

[6] Community strategic guidelines on cohesion (2006/702/EC) For more information, see: http://eur-lex.europa.eu/legal- content/EN/TXT/HTML/?uri=CELEX:32006D0702&from=EN

[7] From the European Social Fund (ESF) and the European Fund for Regional Development (ERDF)

[8] For more information, see: http://ec.europa.eu/europe2020/making-it-happen/index_en.htm

[9] CSRs adopted for the coordination of the economic policies (Article 121(2) of the Treaty) and CSRs adopted for the coordination of the employment policies of the Member States (Article 148(4) of the Treaty. For more information see http://ec.europa.eu/europe2020/index_en.htm

capacity of public authorities and stakeholders and efficient public administration" (TO11)[10] for a total of about EUR 4.2 billion.

In this context, understanding of public administration characteristics and dynamics in Member States is critical for the Commission in order to be able to provide for effective implementation of the ESIF investments, and/or other support and maximise EU value added. Furthermore, any future EU initiatives in this area - be they related to funding, policy or dialogue with Member States - need to be based on a sound understanding of context, needs, opportunities and challenges, as well as drivers and obstacles to administrative reform, in order to be able to respond with a targeted and customised approach that fits the specific needs of the respective Member State.

The amount and detail of information on functioning of national public administrations available to the Commission services tends to vary across (sub-)sectors and countries concerned. This assignment therefore needs to support the Commission in: ensuring consistent and coherent knowledge on the characteristics of public administrations across all EU Member States; deepening its understanding of public administration functioning based on common approach and methodology, and capture of reform initiatives and dynamics; understanding the role of external (EU funded) support to administrative reform process. While there is obvious and particular attention on countries that receive EU funding for public administration reform from ESF/ERDF, and on those countries with specific recommendations from the European Semester, most of the work under this contract will cover all Member States, to have a wider and more varied basis for comparison of characteristics and factors driving change of public administrations in the EU.

The current paper is prepared in the framework of the European Commission project "Support for developing better country knowledge on public administration and institutional capacity building" (hereafter EUPACK – EUropean Public Administration Country Knowledge). The project aims to ensure consistent and coherent knowledge on the characteristics of public administrations across all EU Member States; to deepen the understanding of public administration functioning based on common approach and methodology, and capture of reform initiatives and dynamics; to understand the role of external (EU funded) support to administrative reform process.

[10] Full title of the thematic objective: 'enhancing institutional capacity of public authorities and stakeholders and efficient public administration'

2. EXECUTIVE SUMMARY

In many policy fields, international comparative indicator sets have been used to fuel policy debate. The **evidence base for Public Administration** – the machinery of government - however remains thin.

The **existing measurement initiatives** allow for discerning big differences and large trends in governance. They also have contributed to the agenda setting of public administration issues and have raised our awareness for the potential of comparative measurement.

While progress has been made, a lot of ground still needs to be covered. Comparative Public Administration needs **better data** to credibly underpin the conceptual models of Public Administration performance.

Five strategies are proposed.

First, we suggest a **focus on sub-themes** rather than on comprehensive measurements. Well-chosen themes, such as the *performance of tax administration* provide actionable insights that also give an indication for the system as a whole.

Secondly, we propose using **employee surveys** to collect comparative data on public administration. While perceptions of citizens, businesses and experts are frequently used in current indicator schemes, employees are left out of the equation.

Thirdly, **administrative systems** can be mined for better data. A good understanding of differences in definitions is needed and capacity for interpretation of results is required.

Fourthly, **open data strategies** enable analysts in the academic and non-academic research community to study issues of Public administration performance. The *Digiwhist* project based in EU tender data is a good example.

Finally, apart from government finance statistics, public administration data are not included in the **Eurostat** system. Eurostat does have a programme for experimental statistics where they develop new data sources. Several subthemes of PA may be eligible for experimental development into an established data stream.

3. INTRODUCTION

In the last decades, measuring the performance of public administration has been an incessant ambition of international, European as well as (some) national governments. The desire to develop comparative indicators has become stronger in recognition of the foundational role of public administration and governance in building prosperous societies[11]. Several indicator sets come out of the development sector: most notably the World Bank's governance indicators. In particular in fragile states, the consequences of an ailing public administration apparatus become apparent and the case for measuring public administration can be made. Beyond international organisations (such as the World Bank, the World Economic Forum, and the OECD), also foundations (e.g. Bertelsmann), academic institutions (e.g. the University of Gothenburg and the Blavatnik School in Oxford) and consultancies (e.g. Gallup and companies providing risk assessments) have resorted to measuring public administration.

Successful measurement projects in other policy fields fuelled the drive towards measurement in public administration. Health indicators, environmental indicators and the PISA project for educational performance are some examples of international comparative indicator sets that have been used in many policy documents and discussions. Public administration - the machinery of government - has however long been a statistical wasteland, with the exception of budget data and the national accounts. From the early 1990s onward, several large organisations have promised to fill the void. Yet, in spite of the lofty ambitions, much remains to be done. Indicators of public administration raise important concerns of validity and face serious measurement issues. Definitions are fraught with imprecision and differences between scores are often hard to interpret.

The capabilities of the measurement community should not be put into question. Increasingly sophisticated techniques for data cleaning and analysis are employed. Increasingly more source data and detail about methodologies is provided. Increasingly attractive data visualisation allows for terse, interactive web analysis by users. Yet, too often sophisticated data portals and reports are principally repackaging the same source data. More statistical legwork is needed to build a solid evidence base for comparative public administration.

This paper proposes an agenda for moving ahead with the measurement of public administration. There is no magic bullet in measurement. Many valuable efforts have been undertaken, but much more work is needed. As in other sectors, the development of international indicators will require a continued and coproduced effort of government, civil society, commercial and academic actors.

---

[11] See for instance the thematic factsheet on the quality of public administration added to the European Semester.

The paper consists of two major parts. First, it discusses some of the previous work and assesses where we now stand with the measurement of Public Administration. Secondly, an agenda for future measurement efforts is presented.

4. REVIEW OF PREVIOUS WORK

4.1. Indicator sets with global coverage

The **World Bank Governance Indicators** are probably the best-known and most widely used indicators for measuring public administration. The WBGI covers 200 countries since 1996. It is an aggregation project. Source data from commercial, non-profit and for profit providers is analytically combined into an index. Each source is considered to be a signal of governance – albeit an imperfect one. By bringing all those signals together, a more robust assessment of governance quality is produced[12].

Governance is a broad concept in the World Bank project. It refers to the conditions that need to be met for being a modern, democratic state: voice and accountability, political stability and the absence of violence, government effectiveness, regulatory quality, rule of law, and the control of corruption. Some of these dimensions deal with Public Administration in the strict sense (defined as the apparatus of implementation). Some dimensions reflect on the broader institutional context of a country. Some parts of the *Government Effectiveness*, the *Regulatory Quality* and the *Control of Corruption* indices deal with Public Administration.

In the early 2000s, the validity and reliability of the WBGI's has been questioned. It was argued that the WBGI measures different things in different countries, which hampers comparison. Moreover, the aggregation of data sources into one index would obfuscate rather than enlighten our understanding of trends in governance[13]. In response, the WBGI improved the transparency of the measurement process, with a more detailed account of the methodology and the release of most of the source data. Thanks to this debate, we now have a better understanding of how the indices are compiled – including its limitations.

The WGI now calculates a margin of error around the estimates[14]. This is useful because it allows assessing how certain we can be of difference

[12] Kaufmann, Daniel and Kraay, Aart and Mastruzzi, Massimo, The Worldwide Governance Indicators: Methodology and Analytical Issues (September 2010). World Bank Policy Research Working Paper No. 5430. Available at SSRN: https://ssrn.com/abstract=1682130

[13] Oman, C. P., & Arndt, C. (2010). Measuring governance. *OECD Development Centre Policy Briefs*, (39)1. See also: Arndt, C. (2008). The politics of governance ratings. International Public Management Journal, 11(3), 275-297.

[14] There is no sample in the traditional sense from which the confidence intervals can be calculated. The WGI therefore uses another statistical technique called unobserved components model. Overall, the technique estimates the likelihood of an unobserved mean given the data sources available. More data sources that are more in unison will lead to lower standard deviations and smaller confidence intervals.

between countries and over time. Figure 1 compares the scores of 27 EU member states[15] on government effectiveness for the year 2016. The confidence intervals are also included. A 90% confidence interval means that, if we measure government effectiveness 100 times, the true value would be within the confidence interval in 90 % of the cases. The World Bank reports 90% confidence intervals. In research, a (stricter) 95% confidence interval is more common. Figure 1 reports both intervals.

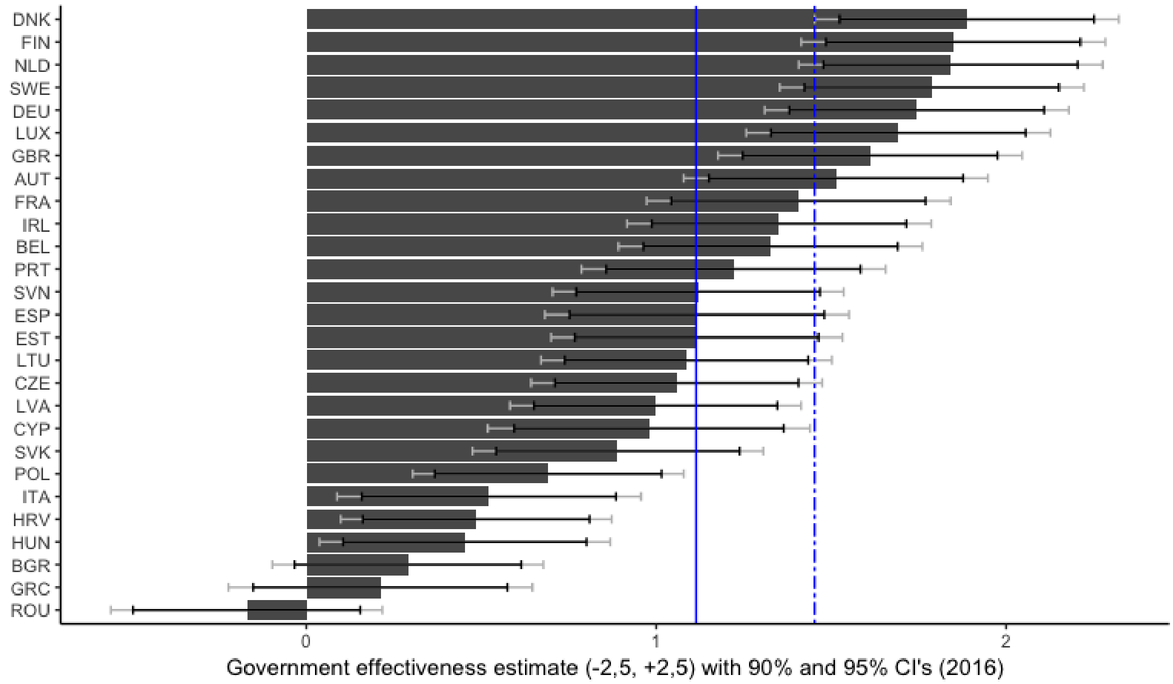

Figure 1: confidence intervals around the government effectiveness estimate

From figure 1, we can infer that all EU28 countries but Romania score above the world's mean (which is at 0), although for Greece and Bulgaria, the score above the mean can be due to chance. We also can assess some large differences in governance. We can be confident that the Scandinavian countries as well as Germany, Luxembourg and the United Kingdom are above the European mean (blue line) and that South-eastern Europe together with Poland and Italy are below the mean. Beyond that, comparisons within Europe should be made cautiously. The five or so places above and below an observation have intervals that strongly overlap. Comparisons of countries close to each other are not useful. Even the confidence intervals of Denmark and the Czech Republic (Rank 17 - 95%CI) and Portugal (Rank 12 - 90% CI) are slightly overlapping. The big lesson to draw from this is that the WBGI are more useful for big differences that can be found on a global scale, and probably less so for smaller distinctions within Europe.

[15] Data for Malta were missing in the WBGI portal.

The **World Competitiveness Yearbook (WCY)** of the World Economic Forum is another long-standing measurement of governance. Since 2004, a yearly report of more than 100 countries (138 in 2016) is provided. The purpose is to explain competitiveness of the economy. One of the 12 pillars of competiveness measures the quality of the institutions. The *institutions* pillar mainly deals with the political and judicial institutions. Items on public administration measure transparency and corruption. Policy fields where the government plays an important role are not covered by the institutions field (as opposed to the government effectiveness measures of the World Bank), but included in sectorial pillars (education, health, infrastructure). The macro-economic pillar includes items on public finance. Indicators on regulatory burden on business are scattered in other pillars.

The WCY relies heavily on a survey of business executives. They are quite open about it. The WCY reflects the voice of the business community. The big question of the WCY is how to remove barriers for free trade. The role of government as a regulator is particularly seen in this light. Regulation is generally seen as a burden. A similar perspective is found in the regulatory quality dimension of the World Bank Governance indicators or in the analysis of the Swiss IMD business school, which administers a similar study amongst business executives[16].

In recent years, the French **Institutional Profiles Database (IPD**) has gained currency in international measurement. Several indicators of the IPD are included in the World Bank Governance indicators. The data of the IPD is obtained from a large survey that is sent out to representatives in the French economic and development missions around the world. The dataset has a stronger focus on Public Administration, with indicators on the reliability of economic data, efficiency of tax administrations, and the capacity for reform. The dataset also has several items on strategic thinking and coordination in the policy process. The limitation of the data is however the limited empirical foundation. There are only one or two experts per country[17]. Several private providers also rely on a limited number of experts. Examples are the Economist Intelligence Unit (which works with correspondents) and the Political Risk Services group.

Arguably the most complete repository of governance indicators can be found at the **Quality of Government** institute of the University of Gothenburg (QoG). The QoG collects data on the general quality of government, on the political system, on civil society, on conflict and violence, on religion, and on most policy sectors in society (education, energy, infrastructure, environment, health, migration, and welfare). As

---

[16] https://www.imd.org/wcc/world-competitiveness-center-rankings/World-competitiveness-yearbook-ranking/

[17] Regional checks have been made to counter personal biases. Yet, while the most excessive outliers may be corrected, the limitations of the method remain. The IPD is transparent with regards to the method.

a result, the dataset has a large number of variables: approximately 2000. The quality of government proper is measured by means of 113 indicators worldwide (with 68 variables available for OECD countries)[18]. Most of the 42 indicators that relate to the bureaucracy are measuring corruption.

The QoG also administers its own expert survey on the quality of government. The survey is mainly filled out by academics. For EU member states, the total number of respondents is 377, with a range between 2 for Luxemburg and 35 for Germany, a mean of 14 and a median of 11. Compared to other expert surveys, this seems to be a decent response. Yet, with on average 14 responses per country, the impact of individual assessments on the total remains high. Conceptually, the survey taps into several core issues of Public Administration. Ten items deal with recruitment and seven items concern working conditions of the public service. Fifteen items probe for the quality of policy making and implementation. Nineteen items deal with transparency, corruption and embezzlement in the public service.

4.2. Indicator sets of developed democracies

Global indicators by definition have to cover a lot of ground. Fragile states have to be measured with the same yardstick as developed democracies. As a result, differences between developed countries are not always meaningful. Moreover, data has to be available globally. Better data from developed countries therefore cannot be used. Better yardsticks that would be able to pick up more subtle distinctions are left unused. Several organisations have attempted to fill the gap by providing assessments of developed countries.

OECD's **Government at a Glance** report is a biennial publication that is taking stock of what the OECD calls public service performance[19]. The underlying database (2017) contains 200 indicators related to the organisation of government. The OECD collects most of the source data for indicators. Some data comes from public finance statistics. Most of the data however is collected from surveys to countries. This is the major weakness of the OECD data. Most of the data is self-reported by country representatives. Besides the potential for strategic behaviour in filling out the surveys, the points of contact for the OECD in the national administration may also not have a good overview of the whole public sector. In particular in federal countries, practices may differ between entities. The OECD tries to mitigate the potential problems of self-reporting by doing peer reviews of the data.

---

[18] Teorell, Jan, Staffan Kumlin, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Anna Khomenko & Richard Svensson. 2017. The Quality of Government OECD Dataset, version Jan17. University of Gothenburg: The Quality of Government Institute, http://www.qog.pol.gu.se doi:10.18157/QoGOECDJan17

[19] OECD (2017), *Government at a Glance 2017*, OECD Publishing, Paris. http://dx.doi.org/10.1787/gov_glance-2017-en

The OECD's Government at a Glance has a strong focus on the Public Administration. The indicators are organized around eleven topics. Considerable attention is devoted to the administrative tasks of government, such as staffing, HRM, budgeting, procurement, digitisation and risk management. Open government and public sector integrity are also included as well as a wide range of indicators on public finance. Some of these topics return in each edition. Other topics are only picked up every four years or even once. Table 1 provides an overview of the topics in the last four editions. Note that some of the indicators underneath the topics shift from one topic label to another from year to year. Women in government for instance was a separate topic in 2013, but is reported under public employment in 2017. In the 2017 edition, the OECD structures the wide range of sector data in health, education and justice based on three general principles of public service delivery: access, responsiveness and quality.

	2011	2013	2015	2017
Employment	X	X	X	X
Procurement	X	X	X	X
Public Finance and Economics	X	X	X	X
Budgeting		X		X
Compensation	X			X
Core results			X	X
Digital government and innovation			X	X
HRM	X			X
Open government		X		X
Serving Citizens (sector data)			X	X
Strategy	X	X		
Delivery modes (outsourcing)	X			
Institutions (Centre of Government)			X	
Institutions (regulators)				X
Integrity			X	
Regulatory government				X
Risk management				X
Women in government		X		

Table 1: topics in Government at a Glance[20]

A second comparison of governance in developed countries is provided by the German Bertelsmann foundation. Their project, the **Sustainable Governance Indicators (SGI)** recently produced a third edition. Indicators are compiled from objective, statistical sources as well as from a survey filled out by an expert network. The expert network consists of two or three academics that evaluate their own country. The country experts, together with a coordinator from Bertelsmann, draft a country report that provides background to the scoring on the criteria. The spread around the scores is not high, which raises questions about

[20] Based on the labels in the data portal: http://stats.oecd.org/Index.aspx?QueryId=78408

the validity of the differences. Figure 2 includes a boxplot of the scores of the EU28 countries on the topics proposed by the SGI.

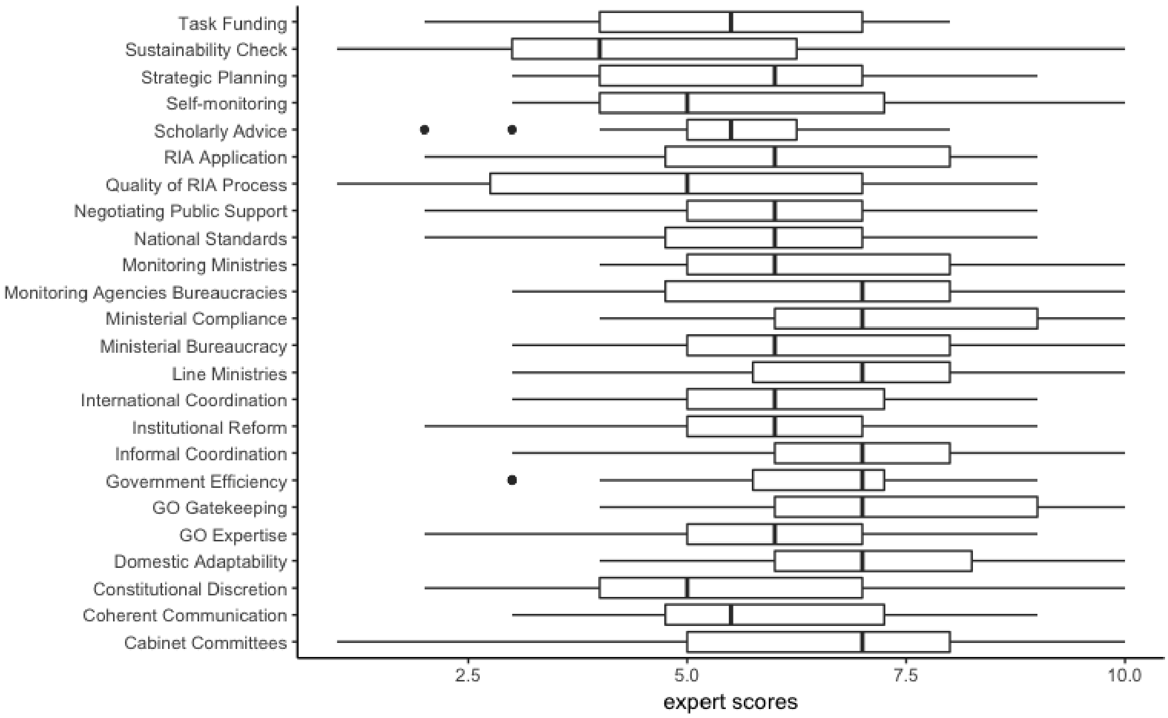

Figure 2: range of the expert scores of EU28 countries in the SGI

The conceptual framework has policy sectors (economic, social and environmental), democracy indicators and indicators of executive capacity and executive accountability. Indicators that are relevant for public administration can be found under the executive capacity header. Executive capacity mainly deals with capacity to coordinate, even across tiers of government. Indicators cover strategic capacity, inter-ministerial coordination, evidence based policy, consultation, implementation, adaptability and reform. All indicators in the executive capacity are assessments of the country experts.

The **International Civil Service Effectiveness** (InCiSE) Index is the latest addition to the governance indicators universe. InCiSe is produced by the Blavatnik School of Government at the University of Oxford. It draws together existing data to study effectiveness of the civil service. Missing data points are imputed. Normalisation of scores is done based on a Min-Max method. The country with the lowest value on an indicator is scored 0 and the highest is scored 1[21]. The 2017 edition is seen as a pilot study, with a limited set of countries (31 developed economies). The study also acknowledges the limitations in data availability. The report warns that given '*the limited availability of some data and complexity of the subject area, InCiSE is not claiming at this*

[21] One concern with this method is that the differences between scores are also standardised. Values that are close to each other in the original dataset are treated similarly than values that lie apart from each other. For example, values of 4,5 and 6 on a scale of 10 translate in a score of 0, 0.5 and 1. Values of 3, 6 and 9 also translate in a score of 0, 0.5 and 1. The former difference may be irrelevant, while the latter is probably significant.

stage to be a robust, comparative measure of civil service performance (p.8).' [22] The technical report reports on the challenges for future editions and suggests some analytical improvements[23].

Conceptually, InCiSE makes a distinction between three core functions and attributes of government action that apply throughout. The first core function is a well-organised central executive that is able to make and coordinate policies, manages fiscal resources, assesses regulation and manages risks. The second core function is mission support: procurement, HRM, ICT, and financial management. The third core function is direct service delivery (the outcomes). InCiSE does not go into policy sectors, but assesses the tax administration, digitisation of services and the administration cost of social security. Six attributes are identified: integrity, openness, capabilities, inclusiveness, staff engagement; and innovation.

4.3. Summary findings and discussion

The main value of the existing indicator sets is that they allow for broad-brushed assessments of governance quality in a country. However, they generally do not sustain more fine-grained analyses within groups of similar countries. While not without its problems, the OECD's government at a glance provides more opportunities for learning. They do not aggregate into overall indices. Moreover, most OECD indicators are backed by more substantive reports and datasets.

The meaning of labels such as *governance*, *institutions*, *quality of government,* or *public service effectiveness* is quite diverse. Sometimes measures of policy sectors are included. In other datasets, they are measured separately or left out. When dealing with public administration in a stricter sense, three perspectives prevail: a political, business and a public administration perspective. The Quality of Government dataset, home in a department of political science, heavily relies on the political perspective. Quality of government is measured as the quality of the political institutions, the lack of undue interference of politics in the bureaucracy and the absence of corruption. The World Competitiveness yearbook of the WEF is taking a business perspective, with a lot of emphasis on removing administrative barriers to free enterprise. The IPD and the WB provide a mixture of both. Public Administration components (budgeting, HRM, contracting, reform) receive much less attention. The OECD's Government at a Glance reports more closely on traditional public administration topics.

[22] Blavatnik School of Government (2017). The International Civil Service Effectiveness (InCiSE) Index. University of Oxford. p.82
[23] Blavatnik School of Government (2017). The International Civil Service Effectiveness (InCiSE) Index. Technical Report. University of Oxford.

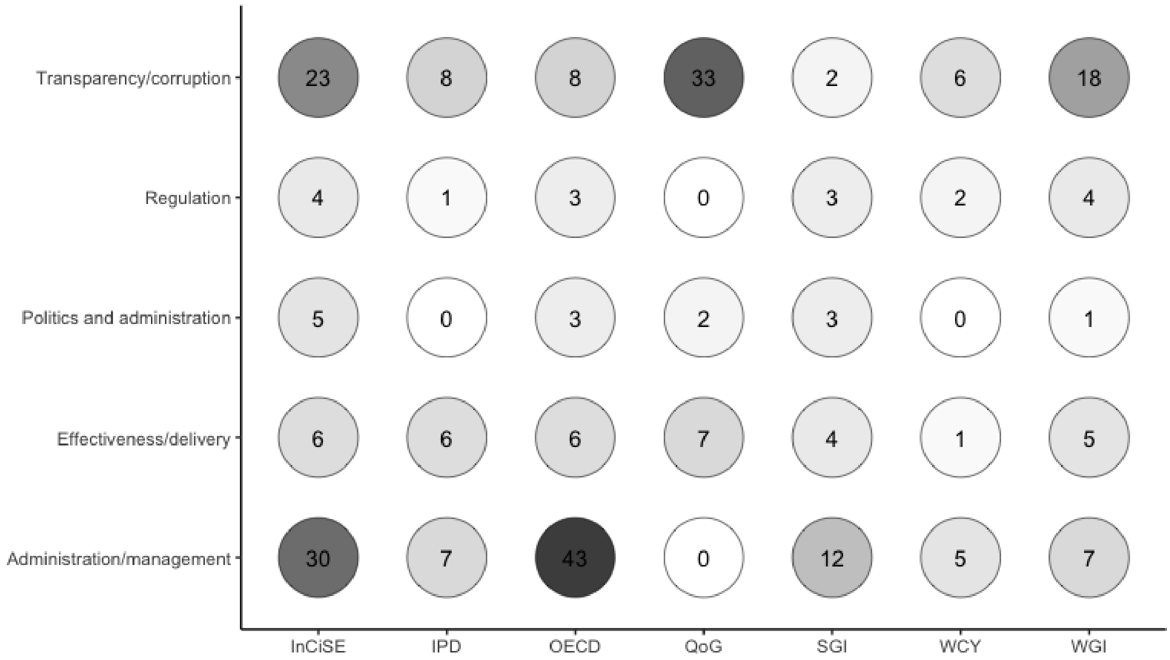

Figure 3: number of indicators for each indicator scheme, clustered in categories.

The seven indicator sets have in total 268 indicators that refer to role of public administration[24]. Figure 3 provides an overview of the number of indicators, clustered in five broad clusters. Administration and management refers to indicators that deal with processes within the bureaucracy. The OECD and InCiSe have many indicators in this field. Transparency and corruption in the public service are the most populated category. These indicators are used in all datasets, and in particular in the QoG database. The regulatory role of government in society, the relation of politics and administration and the effectiveness and delivery of public services are less common.

The indicators sets exchange a good number of indicators (Figure 5). Data is re-used. IPD, SGI, and OECD (except for a few indicators) only report what they produce. WGI, QoG and InCiSE only import data. The OECD exports only to InCiSE. SGI data is used by InCise and QoG. QoG is used by InCiSE. WCY is used by QoG and InCise. WGI uses data from WCY and vice versa.

Under the hood, the indicators mainly rely on expert opinions of academics, diplomats, correspondents, non-profit staff or business executives, which makes the empirical evidence vulnerable for outliers. Two to three experts per country for SGI, one or two experts for IPD, about 15 experts that fill out the QoG survey and the country contacts of the OECD surveys. The World Bank also includes some survey data in

[24] The datasets contain many more indicators, but we only considered the indicators that refer to the public administration system. Indictors of the justice system, policy sectors, parliaments and parties, and democracy in general are not taken into account.

its PA items, but this is an exception. The validity and reliability of low-n expert surveys is debatable. In particular, when no opportunity is provided for argumentation and interpretation of scores, expert judgement may be problematic.

We do have interesting conceptual schemes of governance (see for instance the InCiSE and the SGI models). We also have increasingly sophisticated means of analysis (see for instance the WGI unobserved components model that allows to report standard errors). Despite these advances in our understanding, much of the empirical legwork of collecting data remains to be done.

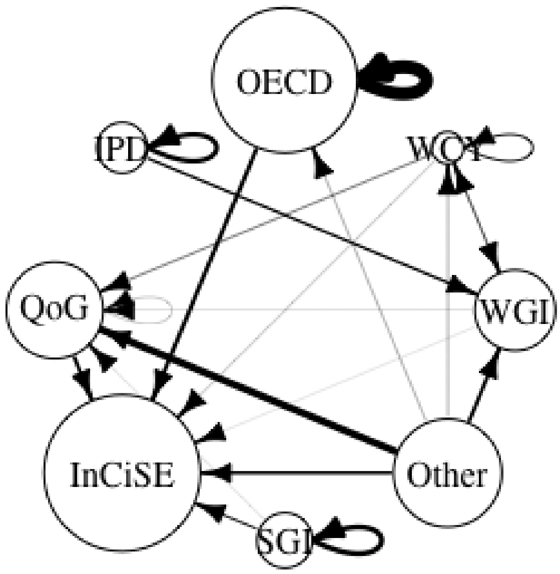

Figure 4: producers and consumers of governance and PA indicators (circles are scaled by the number of core PA indicators)

Finally, an important assumption of international measurement is that governance *can* be compared. Comparative measurement assumes that different societies share in common what they expect from their governments. This assumes that there are some common ways of organising governance, and that there is a shared normative understanding to signpost good and bad performance. As obvious as this may sound, we should not treat this issue lightly. Indicators are performative. They not only measure social reality, they also create the reality they are supposed to measure[25]. When the focus is on deregulation for businesses, then government is reduced to a regulator. Fewer rules, often called administrative burdens, become the standard. These assumptions should not be implicit. A strong awareness of the

[25] Merry, S. E., Davis, K. E., & Kingsbury, B. (Eds.). (2015). The Quiet Power of Indicators: Measuring governance, corruption, and rule of law. Cambridge University Press.

conceptual assumptions behind each measurement effort is vital to understand its meaning and relevance for a particular context.

5. NEW DATA STRATEGIES

There is a good variety of conceptual models for measuring public administration and an increasing sophistication in analysis techniques. Yet, the empirical foundation remains rather thin. This section discusses some strategies for strengthening the empirical base. First, we argue that measurement on subthemes may be more promising than all-encompassing indicator schemes. Next, we discuss two sources of data that could be further explored: employee surveys and administrative data. Finally, we discuss two strategies for making data available in a systematic way to a broad community of analysts. Therefore, we need first, more open data and secondly, more PA data in Eurostat's data portals.

5.1. Selective measurement on sub-themes

A first strategic choice is to lower the ambitions of measurement. Instead of measuring the whole system of governance, we could also focus on subthemes. Most existing datasets work from the top down. They first identify an overall framework and next attempt to fill the concepts with indicators. The whole conceptual model needs to be covered for aggregated reporting. However, the boxes/fields/dimensions are usually still quite broad. Think of concepts such as institutions, government effectiveness or open government. Those concepts cover a wide array of potential subthemes. Open government can refer to information policies, to the use of freedom of information claims, to open data strategies, or to participation in policy-making. Government effectiveness can apply to all policy sectors, but can also apply to the administrative apparatus itself[26]. The concept of 'institutions' can refer to the bedrock of governance – the courts, parliaments – or can be the conceptual *garbage can* for indicators that have no home yet. The high-level concepts do not give much direction towards concrete meanings.

There is a gap between the high-level concepts and the indicators. In research terms, there is an issue of validity of measurement. Do the indicators measure what they purport to measure? Compare this to judging the quality of a bike by the quality of the saddle. Probably, the wheel, the frame and the chain may give more useful indications. Unlike a bike, public administration consists of a vast number of parts that

[26] Van Dooren, W., De Caluwe, C., & Lonti, Z. (2012). How to measure public administration performance: A conceptual model with applications for budgeting, human resources management, and open government. Public Performance & Management Review, 35(3), 489-508.

could be selected for measurement. Hence, the selection of indicators becomes even more critical. Therefore, it may be more realistic to measure (significant) parts rather than the whole.

Rather than measuring against an inclusive framework of public administration, a selection of more concrete subthemes can be made. The OECD's Government at a Glance largely follows this approach. For each edition, they select concrete themes where data collection is feasible. Good examples of subthemes are levels of outsourcing, performance budgeting, centres of government, or compensation of employees.

Indicators need not only to be valid. They also need to be reliable. Reliable indicators allow for consistent measurement over time and place. Suppose we want to measure levels of compensation of employees, then we need a good understanding of how compensation is organised in different countries. If we measure outsourcing, are we measuring the same thing in different countries?

Reliability and validity issues are in particular problematic for objective indicators. The differences in national administrative systems have a large impact on the data. A deep dive into the national context is usually needed to understand the precise meaning of the indicators and to detect measurement problems. In a comparative study of sick-leave statistics, Hoffmann and Van Dooren (2017) compared national data of Austria, Finland, Denmark and the Netherlands. The data were adjusted in order to standardize the methodology.[27] Sick leave in the Netherlands and Austria was substantially higher compared to the self-reported data from the OECD. This is not necessarily evidence for gaming. More likely, it is the result of differences in definitions. Such an in depth analyses are not feasible for encompassing conceptual models.

For perceptual indicators, validity and reliability issues are (in theory) easier to solve. Sample sizes of respondents can be increased and response scales can be standardized. Perceptual indicators also more straightforwardly measure what they claim to measure: i.e. perception of a population of respondents. Explanations for those differences in perceptions of course can go beyond the objective performance of government. Cultural factors for instance can also explain differences in perceived performance. Yet, the fact that perceptions can be explained by many factors does not invalidate the perceptual measure in itself. It is not because there are many causes for CO2 emissions that we should stop measuring. As long as perceptual indicators are treated as being subjective, they do have value. Expert judgements can be problematic in this regard. They are also perceptual, but because the respondents are called experts, the measures get an appearance of objectivity. This

[27] Hoffmann, C., & Dooren, W. V. (2017). Towards Good Enough Measurement–Sick Leave Statistics as a Case of the Measurement Challenges in Comparative Public Performance. Journal of Comparative Policy Analysis: Research and Practice, 19(4), 362-376.

is particularly problematic when the expert only provides a score without a qualitative assessment.

Analyses on subthemes may well tell us something about the system as a whole. The subthemes can be regarded as case studies of the larger system. If selected wisely, it is possible to generalize from cases[28]. One approach would be to take a typical public service that is found in all public administration systems. An example could be the tax administration. Levying taxes is a core function of the state. Both the OECD and the EU have studied the structure and performance of the tax administration[29]. The EC for instance reports on indicators of VAT compliance (the extent to which the theoretical income of VAT is also collected in reality), which differs substantially between member states. According to the EU, *'the variations reflect the existing differences in Member States in terms of tax compliance, fraud, avoidance, bankruptcies, insolvencies and tax administration. It offers an indication about the performance of national tax administrations, but should not be looked at in an isolated way.'*[30] Arguably, the chances are high that tax evasion is found in other branches of the tax system, given that tax evasion is high in the VAT. Similarly, given that the fiscal administration is managed properly, the chances that other core bureaucracies are managed well will be higher. We cannot be certain, of course, but the probability is higher.

The idea that some functions may represent broader notions of public administration performance is also used in a historical study on state capacity and democracy[31]. The authors argue that it is generally better to first build state capacity before engaging in democratization efforts. They measure state capacity by means of one sub-theme: the quality of the cadastre. A cadastre is a systematic inventory of land and land ownership. Cadastres give states detailed information that enables the state to levy tax, enforce property rights, and administer land. The authors argue that the capacity of developing a capacity is at a core of what it means to be a capable government.

The selection of the subtheme does not necessarily need to be a *typical* case of public administration. It may also make sense to select a subtheme that is *a-typical*. Take for instance the performance of prisons. Prisons are a-typical because compared to most other services the state is playing a strong, repressive role. Moreover, prisons are a-typical because compared to public services such as taxes accountability mechanisms are not so strong. In most countries, prison policies affect a small and underprivileged part of the population who, also by nature

---

[28] Bent Flyvbjerg, "Five Misunderstandings About Case-Study Research," Qualitative Inquiry, vol. 12, no. 2, April 2006, pp. 219-245. DOI: 10.1177/1077800405284363

[29] European Commission (2015) Tax Reforms in EU Member States: 2015 Report. Taxation papers N. 58 – 2015. Directorate General for Taxation and Customs Union & Directorate General for Economic and Financial Affairs OECD. Tax administration 2017. Paris: OECD

[30] https://ec.europa.eu/taxation_customs/business/tax-cooperation-control/vat-gap_en

[31] D'Arcy, M., & Nistotskaya, M. (2017). State first, then democracy: Using cadastral records to explain governmental performance in public goods provision. Governance, 30(2), 193-209.

of their imprisonment, have little political voice. Compare this situation with a tax administration that has dealings with the whole population, including strong upper and middle class groups with strong political voice. The external pressure on tax services is arguably stronger compared to prisons.

The argument for selecting subthemes could be that if a state provides for the weakest of its people, the chances are higher that it will provide for the general population as well. The selection of the subtheme should concern those who are often considered undeserving of public services: think of inmates or refugees. Number of detainees and occupancy statistics of prisons for instance are available and may be a good indicator of the criminal justice system. The first aim should be to keep people out of prison, but once convicted, people should have decent living conditions. Lecturers of case study research call this selection strategy the Frank Sinatra inference. If you can make it there, you can make it anywhere.

Selective measurement on subthemes allows investing resources and research capacity into more fine grained, in depth analyses of well-chosen topics. This strategy offers the best chances for picking up upon the often-subtle differences between developed bureaucracies. Moreover, a focus on subthemes usually provides more actionable insights than government-wide measurement. If the purpose of the indicators is to provide evidence for policy learning, they need to be linked to specific public services or structures.

5.2. Data from standardized measurement scales in (staff) surveys

A second strategy for obtaining good international comparative indicators is to include standardized measurement scales in existing measurement efforts of member states. More specifically, staff surveys are an underdeveloped source of PA data. Existing governance indicators use perceptual data from expats, from business leaders, from academics and from citizens. The opinion of the people who work within the administration is not taken into account. This is a missed opportunity, in particular because many member states are already using staff data to assess performance[32]. International comparison can be improved by coordinating the survey items in national measurement efforts.

---

[32] The argument for using staff as a source of PA performance data has been made elsewhere: see for instance the report of the Dutch Ministry of the Interior for the EUPAN network. Lamboo, Van Dooren and Heywood (2016) Prime Witnesses? Case studies of staff assessments for monitoring integrity in the European Union. Ministry of the Interior and Kingdom relations: The Hague. See also a report by the OECD: OECD (2012b). *Performance of Integrity Measures: Proposal for an Integrity Module for Staff Surveys,* OV/PGC/ETH(2012)1/REV1, Paris: OECD. Another report is by: Hoffmann, C., & Van Dooren, W. (2013). *Regional benchmarking of public administration performance : towards a construction of an international comparative dataset*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen.

There are several reasons why asking staff about dimensions of public administration performance is a good idea[33]. Staff assessments are one of many sources for monitoring integrity, but potentially a very strong one. In the first place, staff knows best what is happening in the back office. They are prime witnesses of improvements or decline in public services. Their judgement is not reputational or based on hearsay, but instead based on what they see in their daily job. Secondly, staff assessments can serve as an early warning device. When performance or integrity deteriorates, staff will perceive the decline before outsiders do. When bad performance or corruption is out in the open, the machinery of government "*under the hood*" probably will be affected severely. Finally, staff assessments are actionable. They can be linked to concrete working contexts or concrete policy instruments.

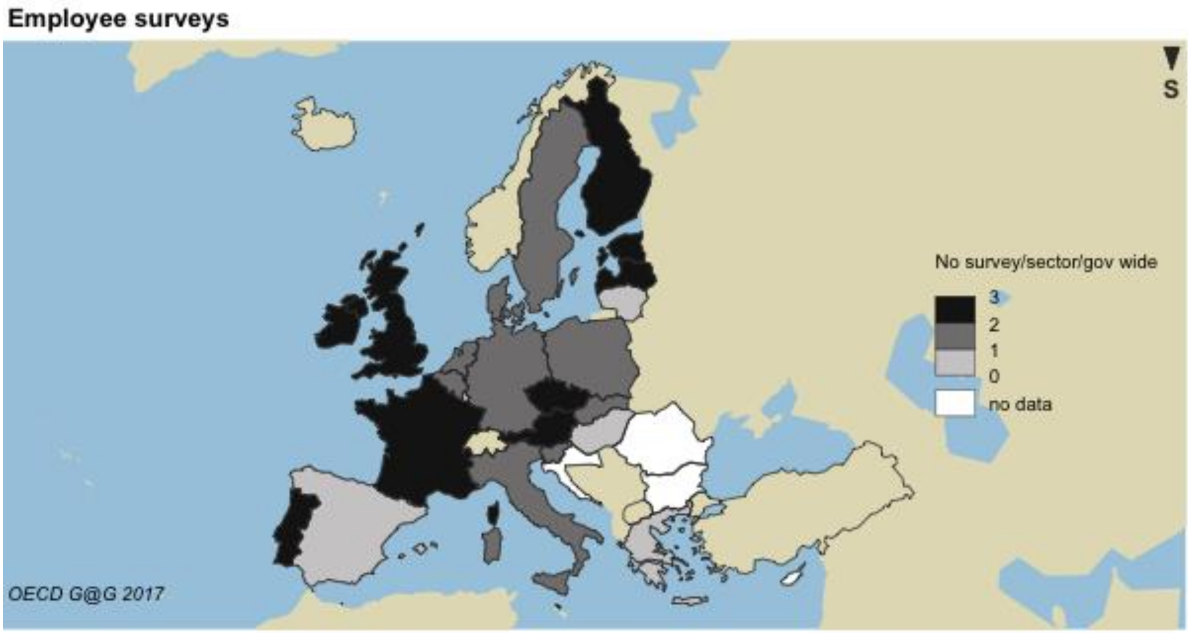

Figure 5: availability of employee surveys (government wide: black, sectoral or ministerial: dark grey, no survey: light grey, no data: white)

The infrastructure to organise the measurement of staff perception is already in place. Almost all countries regularly conduct employee surveys. The OECD's Government at a Glance report gives an overview of the use of employee surveys. Twenty-four European countries filled out the OECD survey on employee surveys. Only Greece, Hungary, Lithuania and Luxemburg do not use employee surveys at all. Eight countries use surveys for the whole central public administration – allowing for comparison across sectors and organisations. Most countries however do not have a centrally administered employee survey. They leave the responsibility to the ministries or sectors, which would allow for international comparison across those sectors. The map (figure 6) shows that there are no regional clusters of countries

[33] See Lamboo, Van Dooren and Heywood (2016) for an extended discussion.

adopting employee surveys, which seems to suggest that administrative cultures do not have a strong bearing here.

	General survey	Interval	Ministry or sector survey		General survey	Interval	Ministry or sector survey
AT	Yes	> 2y	Yes	**IT**	No		Yes
BE	No		Yes	**LT**	No		No
CZ	Yes	Irregular	Yes	**LV**	Yes	2 year	Yes
DK	No		Yes	**LU**	No		No
EE	Yes		Yes	**NL**	No		Yes
EL	No		No	**PL**	No		Yes
ES	No		No	**PT**	Yes	2 year	No
FI	Yes	Yearly	Yes	**SE**	No		Yes
FR	Yes	> 2y	Yes	**SI**	No		Yes
DE	No		Yes	**SK**	No		Yes
HU	No		No	**UK**	Yes		No
IE	Yes	2year	No				

Table 2: the use of employee surveys in government (based on OECD Government at a Glance 2017: p.149). Data from Romania, Croatia, Bulgaria, and Cyprus are missing).

A study by the University of Antwerp of employee surveys in Austria, the Netherlands, Scotland, Flanders and Finland also found a lot of similarities in the surveys[34]. Questions asked are rather uniform (see Table 3). There is no great variation with regard to wording or content. However, some countries use more detailed questions, whereas other countries ask for rather general impressions. Furthermore, most countries group questions according to subthemes such as work environment, teamwork, leadership and career development.

Concept	Austria	Finland	Flanders	Nether-lands	Scotland
Work-life balance	x	X	X	X	X
Work content	X	X	X	X	X
Autonomy in the job	X	X		X	X
Availability of information for doing job	X	X	X	X	X
Availability of tools for job	X	X	X		X
Information on matters	X	X	X	X	X

[34] Hoffmann, C., & Van Dooren, W. (2013). Regional benchmarking of public administration performance: towards a construction of an international comparative dataset. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen: pp.48-65

related to job					
Feeling appreciated	X	X	X	X	X
Workload	X	X		X	X
Skills and task correspond	X	X	X	X	X
Access to learning and development	X	X		X	X

Table 3: overview of employee satisfaction concepts in 5 surveys (based on Hoffmann and Van Dooren 2013)

A key insight is that employee surveys can be used for more then HRM issues. The OECD also reports on the aspects assessed in the survey. HR themes are still most important, with subjects such as employee engagement, work/life balance, workplace diversity and skills match between job and person. Two themes that are not core-HR are also included in most surveys: Effectiveness of management (frontline supervisors and general management) and integrity at the workplace. Issues such as performance, coordination, or administrative burdens are not included, while employees may well have an informed opinion on this.

To conclude, the use of existing staff surveys can help to obtain the view from the employees on how public administration is working – a view that is currently not included. In order to make staff surveys useful for comparison, items need to be validated based on scientific knowledge. Beyond HR, other topics could be considered for inclusion into the surveys.

5.3. Data from administrative data mining

A second strategy to obtain better data is to use the administrative data systems of the member states. All states have built statistical infrastructures. Most departments and agencies of a certain size have management information systems to manage their operations. Can we use these data infrastructure for comparative purposes, maybe after some recalculations?

This idea has been piloted before. Under the Belgian presidency in 2010, the EUPAN network produced one of the first attempts to acquire PA indicators from national administrative systems. The national representatives in the network scored the inventory of indicators on their utility and feasibility, which lead to a judgement on each indicator according to the scheme in table 4. Fifty-seven indicators, organised in seven themes, were discussed. The participants scored 30 of the 57 indicators as having a high feasibility and utility and 17 indicators as having a high utility and but low feasibility. Hence, 47 indicators should either be accepted or be developed. Only one indicator was rejected. The representatives of EUPAN may have underestimated the practical impediments to develop indicators and make them internationally comparable. Subsequent EUPAN presidencies also did not further pursue the path. Yet, the report remains a source of inspiration for potential fields of indicator development.

	Low Feasibility	High Feasibility
Low utility	Reject (1)	Reconsider (9)
High utility	Develop (17)	Accept (30)

Table 4: assessment scheme of the EUPAN project. Number of indicators assessed by the national representatives in EUPAN between brackets.

EUPACK also explored the possibilities of international comparison based on national sources[35]. National experts participating in the project were asked to assess the availability of indicators on the dimensions of the EUPACK project: transparency and accountability; civil service systems and HRM; service delivery and digitisation; organisation and management of government; policy-making, coordination and regulation; and overall performance. The report demonstrates that many indicators could be conceived based on the national data systems

Yet, while there is a potential in national administrative data, two issues remain.

First, the consistency between national definitions is not always strong. The EUPACK report however shows how different the indicators are in different countries. Many indicators are suggested, but not many indicators are available across a wide number of countries. Moreover, even well understood concepts such as absenteeism or turnover are defined and measured differently in different countries[36]. Definitional issues can have a strong impact on results. In fact, a good deal of the work of agencies such as Eurostat is to standardize definitions across countries. While it may not always be possible enforce standardization for indicators of Public Administration, we should at least be attentive to the effects of differences in definitions. We should account for the inconsistencies in definitions when discussing results.

Secondly, suppose we have consistent definitions, the interpretation of causes for differences can be challenging. Take for instance the statistics on the number of FOI requests per 100.000 inhabitants (figure 6). The indicator definition is straightforward; i.e. the total of FOI requests. Yet, the interpretation is difficult. The normative value of this indicator is difficult to interpret from the start. More FOI requests may point to issues of transparency. It may also be the case that the country

[35]Hammerschmidt and Thijs (2017) Interim Report on the Characteristics of Public Administration: Collecting Evidence on Existing National Sources/Indicators to Assess Government Capacity and Performance. Task 1 (Phase A). EUPACK project

[36] Hoffmann, C., & Dooren, W. V. (2017). Towards Good Enough Measurement–Sick Leave Statistics as a Case of the Measurement Challenges in Comparative Public Performance. Journal of Comparative Policy Analysis: Research and Practice, 19(4), 362-376.

has a well organised procedure for filing complaints, or that the legislation grants more rights to its citizens in terms of access to information. As a result, a higher number may be a good rather than a bad thing. Besides normative difficulties, contextual issues may also play a role. A tougher policy on fiscal fraud may lead to an increase in Freedom of Information (FOI) requests of citizens and companies who want to know what information the fiscal administration possesses. Finally, it may be that some countries take care of requests within the administration; before legal FOI requests are filed. The magnitude of the differences reported by the EUPACK network (see figure 7) suggests that not only performance of the FOI procedures is at play in the data. Different systems and different contexts need to be uncovered to make sense of the fact that Ireland has 500 times as many FOI requests then Belgium).

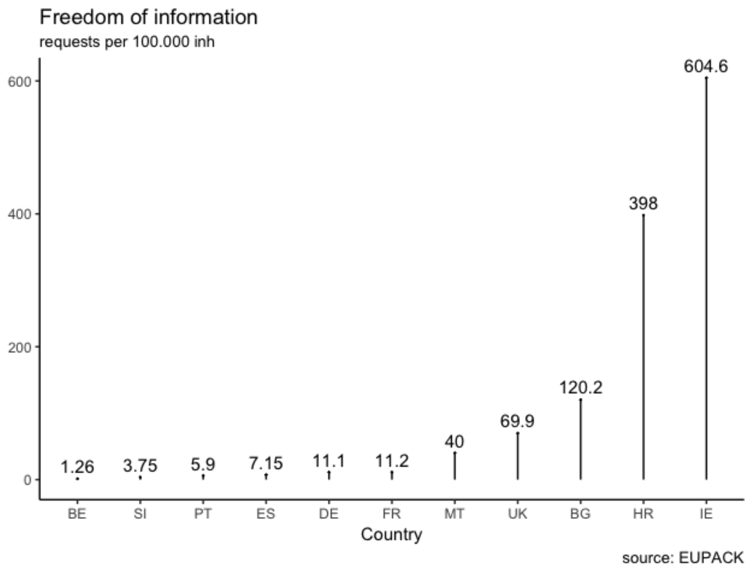

Figure 6: Freedom of information requests (data from the EUPACK network)

Difficulties in interpretation of the causes and the context that drive performance do not invalidate measurement. FOI laws, for instance, are a key mechanism of accountability and it does make sense to keep track and to compare. The surprisingly large variation in the indicator results however does point to the need for additional evidence that may help our understanding of what is going on. Measures of budget accuracy (budget estimates compared to actual accounts) are another example[37]. Good budgeting leads to accurate forecasts. This indicator can relatively easily be compared over time and across countries. Clearly, no one can expect budget departments to predict unusual situations such as the banking crisis. Similarly, economic growth is hard to predict. Yet, patterns of economic growth or decline often play out similarly in different countries. Still, faced with similar conditions, some budget departments' make more accurate budgets than others. The

---

[37] The University of Antwerp study on the benchmarking of public administration performance not only looked into staff surveys (see above), but also into the potential of national administrative data.

most interesting part of this indicator is the interpretation why differences are observed. Yet another example of the need for and difficulty of interpretation concerns measures of budget transparency[38]: Assessment of the levels of disclosure of budget information, of timely reporting to parliament and of comprehensiveness of information can also be obtained from administrative data. It is possible to use national administrative data for comparison, provided that there is sufficient room to discuss both the meaning of the results and how those results were obtained.

The need for interpretative capacity and for a good understanding of the definitions reinforce the argument made before that solid analyses on subthemes seem to be a more promising avenue for comparative measurement than the necessarily superficial measurements of the whole public administration system at once. Interpretation, based on additional evidence and policy debate is needed. More fine-grained expertise on subthemes is needed for such a performance dialogue. The EUPACK inventory of national indicators offers a good starting point for collecting indicators for studies on subthemes.

5.4. Open data strategies

Almost all governments, from local to European, are developing open data strategies. The centrepiece of the open data strategy is generally the open data portal. The inclusion of data on public administration in these data portals may be an important support for thematic, internationally comparative work on public administration performance. The most promising datasets provide large, granular data in a timely way.

Open data strategies tap into analytical resources in society. Open data can enable the research community and civil society to do their own analysis of the performance of government. As a result, the evaluation agenda of government performance is out of the hands of the government. The loss of control over the evaluation agenda may explain why public officials are sometimes reluctant in providing open data. Yet, overall, society will benefit from an open evaluation culture.

The European Union has a promising example of such an open data strategy that could lead to indicators of public sector performance: the database Tenders Electronic Daily[39]. All tenders are available for bulk download, which allows for analysis by researchers. The *Digiwhist* project at the Hertie School of Governance, funded under Horizon 2020

[38]Several frameworks for budget transparency exist; the OECD best practices for budget transparency, the IMF manual on fiscal transparency and the code on fiscal transparency, which form the basis for the fiscal Reporcs on Standards and Codes (ROSC), and the Financial Management Performance Measurement Framework (PFM).
[39] http://data.europa.eu/euodp/nl/data/dataset/ted-1

programme, is diving into the data[40]. Indicators of procurement performance are defined, based on the data. Examples are the share of single bidder contracts, the share of non-open procedures, the weight of non-price criteria, the advertisement period length and decision period length. More single bidders and closed procedures increase the risk for corruption. This is also the case when non-price criteria prevail and the advertisement and decision procedures are short. Again, the data does not tell whether corruption is occurring. It only provides an indication of the risk for corruption. More (and qualitative) evidence is needed to make statements on levels of corruption.

Another good practice is the online availability of employee surveys. The Office for Personnel Management of the US Federal government has a tradition of making the employee survey available at the level of the individual responses (without violating anonymity of the respondents): a dataset with 485 000 responses. The survey has not only been useful to the government. Also more than 40 research papers have been published based on the employee survey[41]. Fernandez et al (2015) argue for closer collaboration between government and academia in strengthening the staff survey design and implementation. In the EU, the Dutch government is also providing the raw data of the survey to researchers after having signed an agreement on its use.

For public administration, progress can be made by making some of the basic data available through open data. Budget data at the most detailed level for instance would hold the potential for more thorough analyses of budget allocation. The budget data which is nowadays available through the European system of accounts, is aggregated based on economic or functional classifications. These data are good for fiscal monitoring. Analysts however may want a more detailed insight in the budget. Another set of basic data concerns employment. The provision open employment data, including socio-demographic characteristics, level of employment, and maybe also compensation data may help analysts inside and outside government to assess the performance of government. Both budget and employment data should be retrievable from the administrative budget and personnel systems.

[40] http://digiwhist.eu/ A paper by Mara Mendez and Mihaly Fazekas lays out the requirements for good open data portals: e.g. low complexity, open and reusable data formats (xml, csv or json), more data depth and quality. See: Mendez and Fazekas (2017) DIGIWHIST recommendations for the implementation of open public procurement data: an implementers guide. Digiwhist: Berlin.

[41] Fernandez, S., Resh, W. G., Moldogaziev, T., & Oberfield, Z. W. (2015). Assessing the past and promise of the Federal Employee Viewpoint Survey for public management research: A research synthesis. *Public Administration Review*, *75*(3), 382-394.

5.5. Eurostat and the statistical system

Inclusion of indicators of public administration in the regular statistical system would leverage the development of solid databases. Today, coverage of *public administration* in the Eurostat database is mainly limited to the government finance statistics that are based on the national accounts. Other indicators are not included.[42] Indicators on public employment for instance could be a candidate for inclusion in the statistics system.

Today, not many PA indicators meet the levels of robustness and comparability required for inclusion. However, Eurostat also has a track of experimental statistics[43]. Experimental statistics are compiled from new data sources and methods. For example, Eurostat is estimating price changes in the food supply chain. Also, they are experimenting with the measurement of the joint distribution of income, consumption and wealth. The experiment is about finding data on income, consumption and wealth for the same household. They also make use of Wikipedia as a source to produce statistics on the online visits to UNESCO World Heritage Sites. The Wikipedia project is conceived as a big data project, with timely data on visits for 1000 UNESCO sites. Several data projects mentioned before, such as the tendering data or staff survey data, are promising avenues for experimental statistical development.

6. CONCLUSION

This paper has assessed some of the most cited efforts to measure the performance of public administration, identified some gaps, and suggested some strategies for improvement.

We analysed the indicators of the World Bank Governance indicators (WGI), the World Competiveness Yearbook (WYC) of the WEF, the Institutional Profiles Database (IPD) of French diplomacy, the Sustainable Governance Indicators (SGI) of the German Bertelsmann Foundation, the Quality of Government (QoG) database at the University of Gothenburg, the OECD's Government at a Glance (G@G) and the recent International Civil Service Effectiveness index (InCise) of the Blavatnik School in Oxford. The existing measurement initiatives allow for discerning big differences and large trends in governance. They also have contributed to the agenda setting of public administration issues. They have raised our awareness for the potential of comparative measurement.

[42] http://ec.europa.eu/eurostat/data/database
[43] http://ec.europa.eu/eurostat/web/experimental-statistics/introduction

However, some problems remain. First, the ambition of existing indicator schemes is too high. They want to measure public administration in its entirety. They typically develop a comprehensive intellectual framework that reflects all the functions of government. Next, the dimensions of the framework are covered with indicators. However, the validity and reliability of the indicators underpinning the conceptual framework is often limited. Most of the indicators are based on observations of a limited number of experts (academics, expats, correspondents) or a small sample of business owners. Indicators are re-used by different indicator schemes. Differences in scores between developed countries are small and probably not significant.

Comparative Public Administration needs better data to credibly underpin the conceptual models of PA performance. Some strategies are proposed.

First, we suggest a focus on sub-themes rather than on comprehensive measurements. Well-chosen themes, such as the *performance of tax administration* provide actionable insights that also give an indication for the system as a whole.

Secondly, we propose using employee surveys to collect comparative data on public administration. Employees are prime witnesses of public administration performance. Yet, while perceptions of citizens, businesses and experts are frequently used in current indicator schemes, employees are left out of the equation.

Thirdly, administrative systems can be mined for better data. Several studies have proposed indicators that could be obtained from administrative data. Yet, the difficulties of international comparison remain a hurdle.

Fourthly, open data strategies enable analysts in the academic and non-academic research community to study issues of Public administration performance. Good examples of open data in the field of PA already exist.

Finally, apart from government finance statistics, public administration data are not included in the Eurostat system. Eurostat does have a programme for experimental statistics where they develop new data sources. Several subthemes of PA may be eligible for experimental development into an established data stream.

7. REFERENCES

Arndt, C. (2008). The politics of governance ratings. International Public Management Journal, 11(3), 275-297.

Blavatnik School of Government (2017). The International Civil Service Effectiveness (InCiSE) Index. University of Oxford. p.82

Blavatnik School of Government (2017). The International Civil Service Effectiveness (InCiSE) Index. Technical Report. University of Oxford.

D'Arcy, M., & Nistotskaya, M. (2017). State first, then democracy: Using cadastral records to explain governmental performance in public goods provision. Governance, 30(2), 193-209.

European Commission (2015) Tax Reforms in EU Member States: 2015 Report. Taxation papers N. 58 2015. Directorate General for Taxation and Customs Union & Directorate General for Economic and Financial Affairs

Fernandez, S., Resh, W. G., Moldogaziev, T., & Oberfield, Z. W. (2015). Assessing the past and promise of the

Federal Employee Viewpoint Survey for public management research: A research synthesis. Public Administration Review, 75(3), 382-394.

Flyvbjerg, B. "Five Misunderstandings About Case-Study Research," Qualitative Inquiry, vol. 12, no. 2, April 2006, pp. 219-245. DOI: 10.1177/1077800405284363

Hammerschmidt and Thijs (2017) Interim Report on the Characteristics of Public Administration: Collecting Evidence on Existing National Sources/Indicators to Assess Government Capacity and Performance. Task 1 (Phase A). EUPACK project

Hoffmann, C., & Dooren, W. V. (2017). Towards Good Enough Measurement–Sick Leave Statistics as a Case of the Measurement Challenges in Comparative Public Performance. Journal of Comparative Policy Analysis: Research and Practice, 19(4), 362-376.

Hoffmann, C., & Van Dooren, W. (2013). Regional benchmarking of public administration performance: towards a construction of an international comparative dataset. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen and

Hoffmann, C., & Van Dooren, W. (2013). Regional benchmarking of public administration performance: a comparative analysis of eight European countries and regions. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen

Kaufmann, D., Kraay, A. and Mastruzzi, M, (2010) The Worldwide Governance Indicators: Methodology and Analytical Issues. World Bank Policy Research Working Paper No. 5430. Available at SSRN: https://ssrn.com/abstract=1682130

Lamboo, Van Dooren and Heywood (2016) Prime Witnesses? Case studies of staff assessments for monitoring integrity in the European Union. Ministry of the Interior and Kingdom relations: The Hague.

OECD (2012). Performance of Integrity Measures: Proposal for an Integrity Module for Staff Surveys, OV/PGC/ETH(2012)1/REV1, Paris: OECD.

Mendez and Fazekas (2017) DIGIWHIST recommendations for the implementation of open public procurement data: an implementers guide. Digiwhist: Berlin.

Merry, S. E., Davis, K. E., & Kingsbury, B. (Eds.). (2015). The Quiet Power of Indicators: Measuring governance, corruption, and rule of law. Cambridge University Press.

OECD (2017) Tax administration 2017. Paris: OECD

OECD (2017), Government at a Glance 2017, OECD Publishing, Paris.

http://dx.doi.org/10.1787/gov_glance-2017-en

Oman, C. P., & Arndt, C. (2010). Measuring governance. OECD Development Centre Policy Briefs, (39)1.

Teorell, J. Kumlin, S., Dahlberg, Holmberg, S., Rothstein, B., Khomenko, A. & Svensson, R. (2017). The Quality of Government OECD Dataset, version Jan17. University of Gothenburg: The Quality of Government Institute, http://www.qog.pol.gu.se doi:10.18157/QoGOECDJan17

Van Dooren, W., De Caluwe, C., & Lonti, Z. (2012). How to measure public administration performance: A conceptual model with applications for budgeting, human resources management, and open government. Public Performance & Management Review, 35(3), 489-508.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct Information Centres. You can find the address of the centre nearest you at: http://europa.eu/contact

On the phone or by e-mail

Europe Direct is a service that answers your questions about the European Union. You can contact this service

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by electronic mail via: http://europa.eu/contact

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: http://europa.eu

EU Publications

You can download or order free and priced EU publications from EU Bookshop at: http://bookshop.europa.eu. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see http://europa.eu/contact)

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en/data) provides access to datasets from the EU. Data can be downloaded and reused for free, both for commercial and non-commercial purposes.

KE-02-18-989-EN-N

Publications Office