

Measuring the Value of Official Statistics: testing and developing a measurement framework



UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

Measuring the Value of Official Statistics: testing and developing a measurement framework

Prepared by the Conference of European Statisticians Task Force to
test and develop the framework for measuring the value of official
statistics



Preface

'Paradigm shift' can be an overused phrase. It often seems to get deployed to dress up unremarkable proposals in overly dramatic language.

Yet when used in its proper sense, it describes a revolution in thinking, which makes us look at issues and problems from a completely new direction. And that's the case with this Conference of European Statisticians work on Measuring the Value of Official Statistics.

Because the 'paradigm shift' phrase is very well deserved to describe this superb piece of work. The report proposes moving away from an indicator-led approach. Instead, it argues that we should be working out what it is that people really value, and build from there – an outside-in perspective as opposed to the top-down, inside-out approach that can easily prevail. This perspective represents a twist of the lens to one that brings everything into sharper focus, and highlights new perspectives. For example, the report makes the profound point that quality is not the same as value; and similarly it argues that an organization's values are not the same as what its users value about its outputs. And above all, the report proposes understanding value by asking people what they care about and how they use statistics.

The report also demonstrates an impressive humility and honesty. It describes the process that the Task Force went through. The Task Force started with an indicator set before realizing this was precisely the wrong place to start. Instead, it shifted to a focus on clarifying the concept of value, leading to a broad, rather than narrow, conception of value. In this broad idea of value, a distinction is drawn between measuring value with indicators that monitor production, and measuring value from the point of view of consumers. It is the latter, of course, that statistical producers should be trying to optimize as providers of a public good. In setting out this process, the authors not only help us understand how they got to the shift in thinking; but they also demonstrate the openness and rigour of the national statistical organization at its best – never shying away from addressing a difficult question. As the report says at one point: "Indeed, we of all people, as statisticians, should not fall into the trap of saying that if something is hard to measure then we won't even try: instead we should attempt to develop a way to measure it!"

I am very impressed with this report, which sets a new agenda for thinking about value. At its heart the report a simple yet passionate plea to start from the perspective of the user.

Is this a paradigm is shift? Absolutely.

Ed Humpherson

Director General for the United Kingdom's Office for Statistics Regulation

Acknowledgements

The work presented here results from the combined efforts of a very large number of individuals from a diverse mix of roles from across many countries all over the globe. The statistical offices of 18 countries and organizations provided detailed information about their current practices and plans, by providing case studies and/or by completing the measures template developed by the Task Force: **Armenia, Australia, Belgium, Canada, Finland, Gulf Cooperation Council, Hungary, Ireland, Israel, Mauritius, Mexico, the Netherlands, New Zealand, Poland, Romania, Slovenia, United Kingdom and the United States of America (Bureau of Labor Statistics)**. In the United Kingdom's Office for National Statistics (ONS), many staff members contributed substantive and logistical support for the kick-off workshop and subsequent online meetings. More people were involved in these efforts that can be listed individually here, but our thanks are extended to all of them.

Individuals who either participated in the 2019 kick-off workshop, joined regular Task Force meetings, coordinated their country's contribution to the measures review template and/or contributed to the detailed review of the measurement framework and subsequent write-up are listed below. Several people gave inspiring presentations at the kickoff event of this group, and although they were not Task Force members thereafter, their contribution was invaluable in shaping the direction of the work. These presenters are listed in Annex 4.

Angela Potter, United Kingdom (Chair)
Fiona Willis-Núñez, UNECE (secretariat).
Anahit Safyan, Armenia
Averil Templar, Australia
Olivier Goddeeris, Belgium
Wendy Schelfaut, Belgium
Gabrielle Beaudoin, Canada
Leila Boussaïd, Canada
Janice Keenan, Canada
Petra Kuncová, Czechia
Jesper Ellemtose Jensen, Denmark
Mikael Skovbo, Denmark
Timo Koskimäki, Finland
Angéla Kátainé Marosi, Hungary
Andrea Ordaz-Németh, Hungary
Elaine O'Mahoney, Ireland
Mirit Cohen, Israel
David Landau, Israel
Laima Grizaite, Lithuania
Set Fong Cheung Tung Shin, Mauritius
Fernando Lazo, Mexico

Andrea Fernandez, Mexico
Juan Ignacio de Anda, Mexico
Ottile Mwazi, Namibia
Michael Ackermans, Netherlands
Meindert Kappe, Netherlands
Kevin Sweeney, New Zealand
Olga Świerkot-Strużewska, Poland
Bogdan, Patarlageanu, Romania
Zala Jakša, Slovenia
Richard Heys, United Kingdom
Charlotte Deeley, United Kingdom
Aaron Davies, United Kingdom
Martin Nicholls, United Kingdom
Andrew Engeli, United Kingdom
Michael Levi, United States of America
Emily Liddel, United States of America
Suha Al-Ruqaishi, Gulf Cooperation Council
Suad Al-Salam, Gulf Cooperation Council
Najat Rashid Ali, Gulf Cooperation Council
Giles Sullivan, Independent expert (New Zealand).

The write-up of the work was developed and approved by all regular current members of the Task Force in August 2021. The drafting of parts of the text was led by Angela Potter, Averil Templar, Angéla Kátainé Marosi, Andrea Ordaz-Németh, Giles Sullivan, Olga Świerkot-Strużewska and Andrea Fernandez, while many other members of the Task Force contributed to parts of the text. Giles Sullivan led the activities relating to the Results Map as described in Chapter 5, including running two online workshops. Andrea Ordaz-Németh developed the concept map of value used in Chapter 2.

Contents

Preface.....	ii
Acknowledgements	iii
Contents	iv
Acronyms and abbreviations	vi
Executive summary	1
Chapter 1. Introduction.....	7
1.1 The measurement framework	8
1.2 Evaluating, testing and refining the framework	9
1.3 Overview of this report.....	11
Chapter 2. What is value and why do we want to measure it?	13
2.1 Introduction	13
2.2 The changing face of the value of statistics in the context of the Covid-19 pandemic.....	14
2.3 Delimiting concepts	16
2.4 Conclusion.....	24
Chapter 3. Daring to be different: a new proposal for understanding value	26
3.1 Introduction	26
3.2 Value from a customer perspective.....	26
3.3 The DARE Model	27
3.4 Conclusion.....	29
Chapter 4. Reviewing the measurement framework.....	30
4.1 Introduction	30
4.2 Guiding principles for refining the measurement framework.....	30
4.3 Overview of findings	32
4.4 Countries' experiences from efforts to measure value	33
4.5 Conclusion: towards a refined measurement framework?	39
Chapter 5. A Paradigm shift in measuring value: measuring for results	40
5.1 Introduction	40
5.2 Turning value inside out.....	40
5.3 What is a Results Map?.....	41
5.4 Experimenting with Results Maps for measuring the value of official statistics	43
5.5 Conclusions and next steps.....	46
Chapter 6. Conclusions, recommendations and further work.....	47

6.1	Conclusions	47
6.2	Recommendations	48
6.3	Further work	50
	References	52
	Annexes	54

Acronyms and abbreviations

ABS	Australian Bureau of Statistics
Armstat	Statistical Committee of the Republic of Armenia
API	Application Programming Interface
BLS	Bureau of Labor Statistics (United States)
BSC	Balanced Scorecard
CEBR	Centre for Economics and Business Research (United Kingdom)
CES	Conference of European Statisticians
CMS	Content Management System
CSO	Central Statistics Office (Ireland)
CVA	Consumer Value Added
DARE	'Dependable-Applicable-Relationship-Ease of Use' model of value
DDI	Data Documentation Initiative
DOI	Digital Object Identifier
EFTA	European Free Trade Association
ERP	Estimated Resident Population
FAO	Food and Agriculture Organization of the United Nations
FTE	Full-time equivalent
GCC	Gulf Cooperation Council
GDP	Gross Domestic Product
GLOS	Generic Law on Official Statistics
HCSO	Hungarian Central Statistical Office
HTML	HyperText Markup Language
IAS	International Accounting Standards
IFRS	International Financial Reporting Standards
ILCS	Integrated Living Conditions Survey
IP	Internet Protocol
ISI	International Statistical Institute
ISO	International Organization for Standardization
IT	Information technology
KPI	Key Performance Indicator
LFS	Labour Force Survey
MPC	Monetary Policy Committee (United Kingdom)
NADA	National Data Archive (Armenia)
NatCen	National Centre for Social Research (United Kingdom)
NBER	National Bureau for Economic Research (United States)
NiGEM	National Institute Global Econometric Model
NOAA	National Oceanic and Atmospheric Administration (United States)
NSO	National Statistical Office
NSS	National Statistical System
NZ	New Zealand
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics (United Kingdom)
PDF	Portable Document Format
QAF	Quality Assurance Framework

SRS	Secure Research Service
Stats NZ	Statistics New Zealand
UK	United Kingdom
UKSA	United Kingdom Statistics Authority
UNECE	United Nations Economic Commission for Europe
UNFPA	United Nations Population Fund
UNSD	United Nations Statistics Division
URL	Uniform Resource Locator
USA	United States of America
USD	United States Dollar
WHO	World Health Organization

Executive summary

Any attempt to quantify how valuable official statistics are must begin by asking what is meant by value, and whose perceived value matters.

This report reviews potential ways of measuring value, supported by case studies demonstrating their use and suitability (or non-suitability) for assessing the value of official statistics. The overarching conclusions of the document are:

- Consumer-based approaches to value result in a different set of potential indicators than those arising from a more traditional, 'production-based' approach to value.
- While production-based indicators can be very useful for operational and management purposes (including areas like quality and budgeting etc), they do not necessarily reflect the value of our outputs in the sense understood here.

Moving towards a consumer-centred approach to measuring value—and, by extension, to creating, maintaining and improving that value, which is the ultimate goal—necessitates a fundamental shift in direction.

This document offers a suggested method to navigate this change, which entails using a 'Results Map' to define a clear path to achieving the central goals of official statistics; working outwards from core strategic goals, to measurable outcomes, to quantitative indicators of value.

Measuring the Value of Official Statistics

The Task Force reviewed a set of measures proposed to help show the value of official statistics. They concluded that:

- Many measures were not sufficiently developed to show value in the way that it was intended.
- Measuring value relies on knowing why want to measure it and value to whom?
- Value is determined by 'users', therefore, measures should be consumer-based rather than production-based.
- A paradigm shift is needed to develop measures from an inside-out approach (working towards a strategic goal) rather than an outside-in approach, where measures are often decided first.

Before we can measure value we need to know what it is

Value is a fuzzy concept and has a range of meanings, so we need to be explicit about what we are talking about when trying to measure it.

We measure value both to prove and to improve

Measuring value relies on knowing why we want to measure it. One reason, is to prove to others that what we do has worth and is a good use of public resources. Another reason is to enable us to monitor the effectiveness of our efforts to increase our value.

To determine value we must seek users' views

First determine what it is that 'users' value. We should not assume that we know what people value; or what we think is important is the same as what they think is important, or that our views are more important.

Value ≠ ValueS


The properties of official statistics are expressed through the core valueS we hold. Just communicating our values cannot convince anyone of our worth. We can say why we think users 'should' find properties valuable, but it is users who will define value by what meets their needs.

Value ≠ Quality

Quality is a well-defined concept in official statistics with clear and agreed dimensions which can be objectively measured. In contrast, value is inherently subjective. For some users, the gravitas that official statistics offer may make them more valuable than unofficial ones.

Developing measures through a Results Map

A paradigm shift is needed to develop measures that work outwards from a central goal. A Results Map defines a clear path to central goals of official statistics, that works outwards through organizational strategies helping to achieve goals; tactical outcomes and measurable areas for which indicators can be developed.


UNECE

social.stats@un.org
valuing.official.statistics@ons.gov.uk

Official statisticians are driven by the conviction that their products, underpinned by the Fundamental Principles of Official Statistics, are uniquely valuable and essential for evidence-based decision-making, for political accountability and for democracy. But there is an increasing awareness within the community of official statistics producers of a disconnect between their own understanding of the benefits and value of their products, and the public perception of the value that official statistics can offer. It is not merely a matter of 'proving our value'. Official statistics are not the only source of statistical information. Users of statistics, or, more broadly, consumers of statistics (see Chapter 2 for an explanation of these terms) justifiably will choose, use and trust official statistics over other sources only when their own weighing of the pros and cons leads them to do so. Governments and other funders want to know if the resources invested in official statistics offer a good return on their investment. Official statisticians want their commitment to the Fundamental Principles, and what they view as the value resulting from this, to be recognized and understood.

Driven by all these motivations, in 2015 a Task Force on Valuing Official Statistics was established by the Conference of European Statisticians (CES), resulting in the 2018 [Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics](#). The Recommendations promulgated a measurement framework comprising a set of proposed indicators for measuring value, and some suggested methods for producing them. In concluding their work, the group suggested that 'pathfinder' NSOs should pilot test the proposed measurement framework, with the support of

UNECE, and should share their findings and experiences with the wider international statistical community.

The present work originated from plans to undertake such pilot testing, via a Task Force established in 2019 for this purpose. Yet two things became clear very early on in the exercise:

- First, the proposed framework inherited from the 2018 Recommendations was a brainstorming exercise, not yet sufficiently developed to be able to ‘pilot test’ without first refining and developing it.
- Second, the whole exercise of developing and testing a measurement framework rested on the untested assumptions that it was already clear what ‘the value of official statistics’ meant, and for what reasons the official statistics community would wish to produce such measures of value. In fact, as the present document shows, these two assumptions needed to be examined and deconstructed in depth before any ‘pilot testing’ could be meaningful and before a useable measurement framework can be developed.

This publication presents the findings of the CES Task Force to test and develop the framework for measuring the value of official statistics. **The main argument put forward by this group is that ‘value’ is determined by the customer.** Therefore, any attempt to quantify how valuable official statistics are must begin with an investigation into **what it is that people value**. A measurement framework comprising indicators of how well official statistics live up to the standards set by the statistical community themselves, no matter how laudable and universal those standards and no matter how detailed and carefully crafted the indicators, is not, in fact, a framework to measure the value of official statistics.

The report documents the process followed by the Task Force to reach this core argument, including a detailed review of the measurement framework, which found that a significant share of the original measures cannot be proposed as indicators of customer-perceived value. Many are still useful for operational reasons, including quality assessment and budgeting, but a much-reduced number are retained as having potential as indicators of value. Amongst those retained are a number of longstanding measures that show or have the potential to show value, and there is ample evidence of methodologies becoming more and more developed in the quest to promote and improve official statistics. Many countries have undertaken or are undertaking targeted work to evaluate and assess their impact, covering a range of objective, subjective and monetary angles.

A wealth of country case studies is presented, to illustrate the diverse experiences, and lessons learned among countries in their journeys towards assessing and quantifying the value of their work.

Key messages

“It’s important that we think not just about what we produce, but also the value of that in the hands of the users” Jonathan Athow, ONS, interviewed during Task Force kick-off event in London, September 2019 ([see full video](#)).

The key messages arising from the two years of deliberations and investigations by the Task Force are the following:

Before we can measure value we need to know what it is

Just as with any topic in statistics, an essential precursor to measurement is defining concepts and delineating what is to be measured. 'Value' is a fuzzy concept with a range of meanings, making it hard to give a precise definition—but this only makes it all the more necessary to be explicit about what we are talking about.

We measure value with a view both to proving and to improving

Developing a meaningful framework for measuring the value of official statistics relies on us knowing why we want to measure value at all. One reason is to prove to others that what we do has worth and is a good use of public resources. Another reason is to enable us to monitor the effectiveness of our efforts to increase our value. Both reasons are valid.

The only way to determine what people value is to seek their views

We should not assume that we know what people value; or that what we think is important is the same as what they think is important; or that if we explain well enough why we think things are important, people will come to share our views. To really know what elements comprise the value of official statistics, we need to ask. This doesn't have to mean direct questioning, which could be poorly understood, but it does have to entail some method of probing users to identify their criteria.

We must not lose sight of the fact that users are also customers and consumers

'User' is a widely-used, conventional term within statistical organizations, referring to those who make some sort of active use of the NSO's statistical products and services. However, it is important not to lose sight of the fact that that users are also customers and consumers in a wider data market. As such, they may source information elsewhere and make choices as to whether and when to do so. Furthermore, the term 'consumer' is broader than 'user' as it is not restricted to those who actively use statistics, but can encompass also those whose interaction, consumption and benefit from statistics may be somewhat more passive—which does not necessarily mean the statistics are any less valuable to them. In this report, while all three terms are used somewhat interchangeably, 'consumer' is therefore the preferred term (as discussed in Chapter 2).

Value is not synonymous with values

Value, like beauty, is in the eye of the beholder. What makes official statistics valuable is not for us to say (unless we ask—see previous point). We can state our values, the things that drive and motivate us and the reasons that we do what we do. We can state the properties of official statistics that result from the values we hold. And we can even state why we think users or society at large 'should' find these properties valuable. But there is no automatic link between values and value. In the end, users will value what fits their needs and it is they who define those needs, not us. Relatedly, communicating our values is no guarantee that this will convince anyone of our worth. Like the foreign tourist who speaks louder and louder in their own language in the hope of being understood, we risk alienating

our users by simply stating our values and hoping that people will 'get it'. More helpful would be communicating how we fulfil *their* criteria of value.

Value is not synonymous with quality

In official statistics, quality is already a well-defined concept with clear and agreed dimensions. It amounts essentially to 'how good our statistics are'. Although one of the dimensions in standard quality frameworks is relevance, which is defined as meeting users' needs, the majority of other quality dimensions are aspects of the statistics themselves which can be relatively objectively measured within an NSO. In contrast, value cannot be determined without reference to the perception of the one doing the valuing. It is inherently subjective. For some users, it may be synonymous with quality, if quality is the main thing they value, but for others it will encompass more intangible aspects including relationships and opinions. For example, for some users, the simple fact of coming from an NSO and the gravitas that offers may be enough to make official statistics more valuable to them than unofficial ones, even if all other aspects of the statistics could be held equal.

Measures of value are not (necessarily) the same as measures of quality nor of adherence to values

Since value is not a synonym either of values or of quality, a framework for measuring value will not consist of identical indicators to those found in quality frameworks, nor will it consist of measures of how well we uphold values and embody the Fundamental Principles. But there may be some overlap, since for some users these things are indeed central components of their perception of value.

Testing and refining the measurement framework has highlighted underlying issues with the exercise

The mandate of the Task Force was to test and refine the framework as formulated in the 2018 Recommendations. The content of this publication demonstrates that this was done. But a principal finding is that the exercise of formulating a measurement framework "from the outside in", starting with available indicators and testing them to see whether they provide a good picture of the thing being measured, follows a flawed logic. The resulting refined framework proposed here, therefore, both sets aside a significant number of measures from the framework because they were found not to provide a good picture of the thing being measured, and contains notable gaps, for which new measures should be developed as this work is taken forward.

Excluding measures from the framework does not mean they are unimportant

The work undertaken by this Task Force involved reviewing each of the proposed measures included in the 2018 framework, in light of the collated experience of 18 countries and organizations who shared their views. This review resulted in a significant number of those measures being excluded from the framework, for a range of reasons: because they are not considered to be actually indicative of value; because they are not quantitative indicators; because they are not well-grounded in a logical pathway from intended impact, to outcomes, to measures indicative of achieving those outcomes; because the behaviours implied by tracking that measure are not desirable; because it is not evident

what a 'good' level of the indicator would be or it does not suggest a target towards which to strive; and/or because there is no obvious monotonic relationship between the indicator and the 'value' it purports to measure.

Nevertheless, excluding a measure from the framework for any or all of these reasons does not mean there is no point in producing it. It just means that it is not proposed by this group *as a potential indicator of the value of official statistics*. It could be an indicator of something other than value. There may be many other reasons to produce the measure. The case studies illustrate a number of examples of this.

A framework for measuring value can include things we don't currently know how to measure

There are many possible reasons to exclude a suggested measure from the measurement framework, but the difficulty or current impossibility of measuring it should not be one of them. If we think something would be a good indicator of value but we don't currently have a means of quantifying it, that should not be a reason to throw it out. Indeed, we of all people, as statisticians, should not fall into the trap of saying that if something is hard to measure then we won't even try: instead we should attempt to develop a way to measure it! This being said, of course, the converse is also true: just because a quantitative measure of something *does* exist does not mean it necessarily is relevant and should be used.

A useful approach to developing a value-measurement framework is to use a Results Map

In addition to reviewing and assessing the 2018 measurement framework, the Task Force considered an alternative means of developing a measurement framework. Instead of beginning with possible indicators and testing them to see what works, this approach works in the opposite direction. It begins by defining the central goals of official statistics, then works outwards through organizational strategies that help achieve these goals; to tactical outcomes; to measurable areas for which indicators can be developed; and finally to the indicators themselves. While much more time-consuming and with the real possibility that not all proposed indicators can (yet) be easily measured, the approach has the major strength that every resulting indicator in the framework traces a clear path from the central goal.

The refinements proposed to the measurement framework are only a first step: this Task Force proposes a fundamental shift in approach to take the work forward

The reviewed and refined framework proposed in this publication is based on assessing the measures suggested in 2018. Following through the Results Mapping process in depth, either on a country-by-country basis or collectively on an international level to develop a generic measurement framework capable of being adapted to different countries' circumstances, is a task beyond the mandate of the present Task Force; but one which this group argues could be extremely useful to help inform ongoing efforts to prove and improve the value of official statistics. Rather than pilot testing individual indicators of value, therefore, this group proposes that future work centre on *pilot testing the Results Map procedure*, drawing heavily on user (and non-user) consultation, to guide official statistics in its endeavours to offer value to society.

Chapter 1. Introduction

“What makes official statistics valuable to you?”

1. More than a thousand people viewed this question when posted by a statistics professional in a highly engaged network on LinkedIn. Yet only two offered substantive answers.
2. Posted on official social media accounts, the question prompted support in the form of ‘likes’ and ‘retweets’, but not one reply.
3. What does this tell us? That those who interact with official statistics via social media platforms cannot think of a single thing that makes our work valuable? Should we just pack up and go home?
4. The work presented here argues that official statistics, as an industry, is most likely immensely valuable to users and to society as a whole, but that our introspective habits built up and institutionalized over many years have led us to take this value for granted, leaving us without a clear means either of proving or of improving it.
5. One reason why people do not, or perhaps even cannot, answer the question posed above is that it is too abstract. No industry, no provider of any product or service, would expect their customers to be able to articulate directly and consciously the range of criteria they apply, or the relative weights they assign to these criteria; for example when choosing a car, a pair of jeans or a brand of cookies. If it were this simple, market research in any field would be a much smaller operation. In fact, it is a whole industry based on subtle and complex ways of eliciting information from people about the criteria they apply, often subconsciously, and how they relate these to value, whether in terms of value for money, time, prestige, or other factors.
6. So it must be for official statistics.
 - First, *ascertaining what criteria our users (or potential users, passive users and non-users) apply*¹ when determining the value they assign to our products and services is a complex operation, because such determination is rarely conscious or explicit, and because it is almost certainly not fixed but varies across user, use, time and space.
 - Second, *the task of ascertaining these criteria is not one that can be done purely internally without consulting those very users*—even though such consultation is challenging, and as the social media example above illustrates, cannot be achieved by simple direct questioning alone. If we opt instead to come up with our own components of value based on what we assume to be valued by users, we risk identifying the wrong criteria and then being led by these towards erroneous goals.
 - Third, even once the criteria—the constituent elements that together result in a user’s perception of value— have been established, operationalizing them in a way that permits

¹ This variety of terms is given rather than the simple word ‘user’ since some people use official statistics explicitly and consciously; some use them without being aware of it; some benefit from their existence without actually ‘using’ them *per se*; and some do not use them or benefit from them at all. The value, or lack of value, of our work to each of these groups is something we may wish to understand and potentially to alter. For more on this as well as a discussion of the differentiation between the terms ‘user’, ‘customer’ and ‘consumer’, see Chapter 2.

us to monitor, compare, set targets and assess progress towards them, is an equally large and multifaceted task.

7. This report represents an attempt to expand upon each of these three ideas, within the mandate of the CES Task Force to test and further develop the framework for measuring the value of official statistics (hereafter the Task Force on measuring value, or just the Task Force).
8. Official statisticians are driven by the conviction that their products, underpinned by the [Fundamental Principles of official statistics](#), are uniquely valuable and essential for evidence-based decision-making, for political accountability and for democracy. But there is an increasing awareness within the community of official statistics producers of a disconnect between their own understanding of the benefits and value of their products, and the public perception of the value that official statistics can offer. It is not merely a matter of 'proving our value'. Official statistics are not the only source of statistical information. Consumers justifiably will choose, use and trust official statistics over other sources only when their own weighing of the pros and cons leads them to do so. Governments and other funders want to know if the resources invested in official statistics offer a good return on their investment. Official statisticians want their commitment to the Fundamental Principles, and what they view as the value resulting from this, to be recognized and understood.
9. Driven by all these motivations, in 2015 a Task Force on Valuing Official Statistics was established by the Conference of European Statisticians (CES), resulting in the 2018 [Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics](#). The Recommendations promulgated a measurement framework comprising a set of proposed indicators for measuring value, and some suggested methods for producing them. In concluding their work, the group suggested that 'pathfinder' NSOs should pilot test the proposed measurement framework, with the support of UNECE, and should share their findings and experiences with the wider international statistical community. When endorsing the Recommendations, CES called for pilot testing of the framework for measuring the value of official statistics, and for follow-up of the outcomes of such pilot testing in order to further develop the framework.
10. The present work originated in the intention to undertake such pilot testing, coordinated by the Task Force on measuring value which was established in 2019 for this purpose.

1.1 The measurement framework

11. The 2018 Recommendations put forward a framework made up of three components, or kinds of measures of value:
 - observable or **objective indicators** (reflecting actual use of statistical products; measures of quality assumed to be adding value; and/or demonstrating adherence to the Fundamental Principles of Official Statistics)
 - **subjective indicators** (covering perception, trust, support, satisfaction etc.), and
 - **monetary valuations** (quantifying the impact of statistics in monetary terms and/or weighing the value of outputs against the cost of inputs).

12. In general terms, the objective component consists of things which could be measured with relative ease from existing sources, such as download counts, numbers of citations in various different types of media, social media interactions, etc. The inclusion of these as proposed indicators of value rests on the assumption that when people view, download or use our statistics or mention our work, this is an indication that they find our work valuable.
13. The component comprises measures which would primarily be obtained from user satisfaction surveys—either dedicated full annual surveys, or continuous mini surveys on web pages. This component offers crucial insights into user confidence and trust in official statistics, and into how useful, relevant and accessible they are found to be by users.
14. The monetary component merits a heading of its own due to its communicative power, even though it is not in the same conceptual or logical category as subjective and objective measures (it is not a mutually-exclusive three-way classification). Much of the discussion around the value of official statistics relates to the desire to investigate whether official statistics are good value for money, as a public good resourced by public money. Monetary measures of value are the most complex to produce but particularly easy to communicate. Given the unique position of NSOs being in the field of quantification, there is an expectation that they should be able to defend in quantitative terms the value that they add.
15. Within these three components, the 2018 framework suggested wide range of indicators or measures, grouped into sub-components (e.g. the objective component is divided into indicators of quality, transparency, use and relevance; the subjective component includes, among others, the sub-components ‘awareness of brand and message’ and ‘user support’). The full original proposed framework can be seen in Annex 1.

1.2 Evaluating, testing and refining the framework

16. The CES work on measuring value was established with the intention of ‘pilot testing’ the measurement framework. But early on in the activities of the group two things became clear. First, it was found that the framework was not yet something that could simply be pilot tested. The loose constellation of measurement ideas proposed in the 2018 Recommendations was an early brainstorming exercise, not yet sufficiently sophisticated to be operationalized into actual measurable indicators. ‘Number of tailored services by user group’, for example, or ‘share of users whose information needs were met’ are far from being sufficiently clear to be able to produce figures without first establishing a variety of necessary definitions and a standard methodology.
17. Second, the group found that some of the items in the proposed framework were not, in fact, measures of value, while others that perhaps ought to be included in a comprehensive consideration were absent from the framework. Indeed, the whole exercise of developing and testing a measurement framework rested on the untested assumptions that it was already clear what ‘the value of official statistics’ meant, and for what reasons the official statistics community would wish to produce such measures of value. In fact, as this review shows, these two assumptions needed to be examined and deconstructed in depth before any ‘pilot testing’ could be meaningful and before a useable measurement framework could be developed.

18. The group therefore expanded its aims, to include both testing *and developing* the measurement framework. To develop it, they first needed to achieve clarity and consensus on what value actually means, and why NSOs would or should want to measure it. On this basis, they designed a more coherent and complete approach to measuring the value of official statistics. In combination with detailed input from each participating country about their own experiences in attempting to produce measures of value in the various aspects of this framework, and complemented by a rich set of case studies, the Task Force was then able to put forth generalized recommendations for NSOs on whether, why and how to apply the framework.

1.2.1 Work undertaken

19. The activities undertaken by the group in direct fulfilment of the tasks foreseen in its terms of reference were:
 - a) An initial scoping exercise in which countries shared information on which areas of the 2018 framework they were testing, or planning to test, or areas on which they already regularly produce indicators. This helped shape an understanding of where the biggest gaps are—not surprisingly, monetary valuations stood out as being the least produced.
 - b) A kick-off event in London in September 2019 bringing together Task Force members and many UK-based experts to share initial thoughts and findings of research into measuring value
 - c) Brainstorming a long list of possible measures to allow the possibility to consider new ones not included in the original framework
 - d) Developing and having each participating country complete a template in which they reported on what measures they produce and for what purposes, as well as to rate them according to how useful and important they are within the NSO as indicators of value. Completed templates from 18 countries and organizations provided a rich source of information upon which to base the group’s ensuing recommendations.
 - e) On the basis of the collated answers to these completed templates, Task Force members reviewed and evaluated each measure (both those from the initial framework and the small number of newly-proposed ones from the group brainstorming), using an online polling tool to score each measure, decide whether it should be kept, removed or refined, and share proposals for refinement, interpretation, etc.
20. In parallel with these activities, the Task Force developed an agreed template and collected country case studies to describe work related to measuring value using any of the identified measures. The case studies are designed to be a succinct way for countries to share the key points relating to their rationale, methodology and lessons learned. A number of countries have indicated willingness to provide further case studies, which could be added to the online version of this work on an ongoing basis.
21. The group extended their work in order to properly fulfil the mandate of the Task Force. The group’s objective, as set out in its terms of reference, was to assess the feasibility of applying the framework and its utility for producing helpful measures of value, and the expected outcome was a refined measurement framework. Furthermore, paragraph 8 of the terms of reference stated that “a simultaneous and equally important goal of the exercise is to provide

a collaborative forum for exchange of ideas, methodology and lessons learned between participating countries.” The following extended activities enabled the Task Force to fulfil these objectives:

- f) Regular online meetings in which group members also invited colleagues from their offices to present research findings and projects related to measuring value.
- g) Deconstructing the meaning of value and the purposes for which official statistics would want to measure it, including a consideration of the ways in which the Covid-19 pandemic has influenced both of these. The outcomes of this are presented in Chapter 2 and provide the basis for many of the conclusions and recommendations offered in Chapter 6.
- h) Two online workshops to develop and test the idea of using a Results Map technique (see Chapter 5 for details) to better formulate an approach to measuring value. The method is described and the results of the workshops are also included in Chapter 5.

1.3 Overview of this report

- 22. This report comprises five further chapters after the current one, followed by a series of annexes containing more detailed material:
- 23. Chapter 2 examines the concept of value and what we mean by the value of official statistics. Building on the related discussions in the 2018 Recommendations, it considers why we as official statistics producers might want to produce measures of value, and what purpose such measures might serve. Informed by two years of discussions, it is underpinned by a critical examination of some of the assumptions often made within the official statistics community, and questions the common assumption that our value is high and just needs to be better communicated. The chapter also discusses how the conception of value has been influenced by the Covid-19 pandemic.
- 24. Chapter 3 proposes a simple four-component model, termed the ‘DARE’ framework, to guide us in conceptualizing customer-centric value. The model invites us to broaden the scope from a consideration of characteristics of our products and services themselves, to encompass also aspects such as relationships, relevance, and ease of use.
- 25. Chapter 4 presents a high-level summary of the findings of the review of the original measurement framework, explaining that the review found a significant number of the measures to be either unsuitable as measures of customer-perceived value, or else in need of development to render them suitable. It describes in summary form some of the experiences of countries which have produced measures of value, distilling the key findings and lessons learned for the overall endeavour of measuring value (the experiences of countries are described more fully in the corresponding annex).
- 26. Chapter 5 offers a proposal for a radical shift in the approach to take for conceptualizing and measuring value. Recognizing the limitations of the measurement framework reviewed in Chapter 4, it argues for a performance measurement approach, starting from formulating a goal and defining value in terms of outcomes that contribute to fulfilling that goal. The approach offers a method for developing indicators and demonstrates an experiment in applying this approach, undertaken by the Task Force. It ends by suggesting that this approach, referred to

as the 'Results Map' approach, could be piloted and eventually adopted by NSOs as a means of integrating a consideration of value into their organizational strategies.

27. Finally, Chapter 6 outlines some important conclusions and recommendations. It also offers proposals for continuing and expanding this work in light of the findings presented here.
28. Five annexes can be found at the end of the document:
 - Annex 1 shows the originally proposed measurement framework which the Task force was asked to review and develop.
 - Annex 2 describes in detail the ways in which the measures were reviewed and the key findings from countries' responses. On this basis, it then states which of the measures are proposed as potential indicators of value; which are proposed for exclusion; and offers suggestions related to some of the measures, including refining definitions, guidance on interpretation and some groupings of measures to increase their utility.
 - Annex 3 consists of country case studies showcasing work on measuring value. The case studies include work related to measures that are maintained in the framework, as well as measures that have been proposed for exclusion from the framework. It is explained that this is because even where a measure is not included in the refined framework, it may still have other uses beyond value measurement.
 - Annex 4 acknowledges individuals who, while not members of this Task Force, inspired their work at a kickoff event of the group held in London in September 2019.
 - Annex 5 (a temporary annex until the document is prepared in clickable online format) offers brief explanations of terms and references for the concept map of value used in Chapter 2.

Chapter 2. What is value and why do we want to measure it?

2.1 Introduction

29. In today's data environment, it is no longer enough to take the value of official statistics for granted: to assume such value exists and that stakeholders share our perceptions of it. Big Data and the accompanying boom in statistical products has accelerated an evolution in customers and in the products that they demand, as policymakers, businesses and the general public increasingly look for timely and reliable data to guide their decisions. This growth is combined with an increased ease of access, greater use (and misuse) of statistics, and circulation via social media platforms that makes the data market surrounding NSOs more competitive than ever before.
30. The Covid-19 pandemic has deepened and accelerated these trends, acting as a catalyst for changes that were already underway, not only in the production of statistics but also in the way that NSOs assess and reflect upon the value of their contributions to society. Section 2.2 below shines a spotlight on the ways in which the pandemic has affected society's and NSOs' understanding of value.
31. Given these changes, NSOs have an opportunity to measure the value of official statistics and to communicate it in a compelling way to stakeholders including to the wider public. Currently, NSOs are broadly recognized as independent bodies, whose products and services can be trusted by society. But they should not be lulled into complacency by relying on institutional trust or prestige to navigate the new statistical landscape. Nor should they underplay the value of official statistics, particularly as they look to defend their budgets or advocate for more resources.
32. The ability of NSOs to review the value of their outputs from the perspective of users, and to set improved value targets that respond to user needs, will define the continuation into the future of the traditional high esteem and relevance they currently enjoy.
33. In addition to highlighting the positive impact of the work of NSOs and the larger National Statistical Systems (NSSs) on society, the process of measuring the value of official statistics will intensify dialogues with users, potential users, and 'passive users' (anyone who benefits in some way from the use of the official statistics, from their use in decision making; they may not consider themselves users of statistics but they benefit from their existence). This in turn will enable NSOs to monitor the effectiveness of their efforts and support product development to better meet the needs and interests of users.
34. Private statistical providers track the value of their products by contrasting input efforts against revenue, as their survival is tied directly to the responsiveness of their customers. Prices reflect willingness to pay and patterns of consumption, which are aligned with trends and deeper changes. NSOs differ from this, both in their core function – to provide their countries with data that informs decision-making in every layer of society – and circumstances – being publicly funded and forming part of the national institutional fabric. Therefore, as with any public good, the process of weighing the value of their contribution must expand beyond a purely monetary

approach: ‘value’, for official statistics, does not only mean ‘value for money’. Good value implies that the benefits outweigh the costs, and in the case of a public good neither the costs nor the benefits are necessarily only monetary in nature. Ultimately, the value derived by society depends on whether society’s needs are met in order to make well-informed decisions with reasonable confidence.

2.2 The changing face of the value of statistics in the context of the Covid-19 pandemic

35. Within days of the World Health Organization (WHO) declaring a global health emergency, a change began to emerge in the way that statistics and data were being used and perceived across the world. Daily press briefings from the WHO and other international organizations, national governments and news bulletins began to centre around numbers and graphs; people who had not needed to look at or interpret a graph in years found themselves studying trendlines and axes; whole editorial news articles were devoted to picking apart the nuances of different indicators of mortality, hospitalization and positive cases.
36. As the main custodians of official data on these questions, this threw NSOs into uncharted waters. They were expected to provide and stand behind figures, more rapidly than ever and against a level of scrutiny never before witnessed. Official statistics producers have always been proud of the fact that they are transparent, scientifically robust and politically independent. Some users, in the past, may perhaps not have placed especially high value on these characteristics, but as national responses to the pandemic often became politicized, these qualities suddenly took on a central importance for many. At the same time, the relative importance of various quality dimensions—timeliness, punctuality, accuracy, accessibility and others—in the eyes of users underwent shifts, as the nature of the users and uses transformed.
37. This is not to say that the *meaning of the term value* has changed. The idea of value being shaped by “fitness for use” remains unaltered. But the uses to which statistics are now being put and the demands being made of the offices producing them are in many ways very different from what they were before the onset of the pandemic. Policymakers, scientists and the public have needed information urgently to inform their rapid-response decisions related to the progress or impacts of the pandemic; and as such, they may find themselves placing greater value on the timeliness of statistics to fulfil these urgent needs, potentially placing less emphasis on precision and level of detail, often touted as the ‘gold standard’ for official statistics.
38. Many NSOs were already starting to assess the value of their outputs prior to the Covid-19 pandemic, responding to the proliferation of statistical products that have become available in the marketplace. Today, however, we are witnessing how the perception of value changes in line with circumstances and the needs of the user or customer, making it clearer than ever that these value assessments need to be regular and ongoing, not one-time-only evaluations.
39. For example, with the onset of the pandemic, governments were confronted with very different data needs to support their pandemic response planning. GDP metrics are of course still required to measure the economic impact, but a highly precise, quarterly product with a six month time lag is inadequate in circumstances such as lockdowns where changes are large, rapid and subject to major short term fluctuations. Instead, ‘flash GDP’-type products have evolved. These products feature much higher frequency and timeliness than traditional

measures. Whilst these products diverge somewhat from the traditional value model of official statistics, they respond directly to customer needs (and their perception of value) for dynamic decision-making.

40. A further example of the way that the constituent aspects of the value of statistics have changed, is the way in which demand shifted towards compact, 'pre-digested' statistics. Products that were easy to interpret, based on dashboards, visualizations, storytelling and simple graphics with clear tag-lines gained prominence. This corresponded both to increasing diversity of users (with more non-expert and less statistically literate users among them), and a need for more rapid processing even among more expert users such as journalists and civil servants who are used to working with numbers but lack the time to process them themselves. Many NSOs introduced their own national Covid dashboards or portals², and Eurostat developed the [European Recovery Dashboard](#), which was launched in December 2020.
41. In almost direct contrast to this heightened value of aiding user interpretation through increased pre-processing, a parallel shift took place. For some users, the availability of detailed, transparent, verifiable metadata was what gave official statistics the most value in the early days of the pandemic. For many, the special value of official statistics as compared to other sources derived from the ability to check sources, understand concepts and definitions, investigate the reasons for differences in mortality as measured in different ways and the reasons for certain peaks or troughs in infection rates, unemployment rates, testing frequency etc.
42. A final major shift in the value of official statistics — understood as an industry, system or NSO rather than in terms of individual statistical products or services — brought about by the pandemic is the greatly increased importance of official statistics' role in safeguarding correct and fair use of statistical information. The role of chief statisticians, chief economists and statistical regulatory bodies in calling out deliberate or accidental misuse has always been a unique selling point of official statistics. The rise of Covid-related misinformation, combined with widespread mistrust of governments and politicization of Covid response policies, led NSOs right into the spotlight as 'custodians of statistical facts'. Extra efforts have had to be employed to ensure public understanding of the political independence of NSOs. From the point of view of a member of the public looking on as politicians debate their points using statistics, the knowledge that the figures being cited are produced independently from those citing them, and the knowledge that egregious misuse of statistics will be publicly called out, is likely to be more valuable than ever.
43. Such shifts in customers' assessments of the value of various aspects of statistics are triggered by specific situations which are unique in time and place. These shifts can, and must, lead to concomitant shifts in the focus of NSOs to ensure their work continues to be valuable. The landscape is complicated by the fact that different customer segments have different criteria, as well as different degrees of 'negotiating power' or influence over what kinds of statistics get produced as well as how, when and in what format they are published. The consequences of not fulfilling a customer group's needs also varies from group to group. Hence, during the pandemic, the imperative to adjust outputs to the immediate needs of government health policy might have been particularly strong. The specific changes identified might well be short term. Some of the new users may indeed gradually move back to being non-users or passive users, while the regular and expert users may eventually return to a situation where they are

² For example, [Poland](#), [Ireland](#) and [the Netherlands](#), [OECD](#), among many others.

more interested in the degree of detail and accuracy than in how rapidly they can obtain preliminary figures. But what will remain as a permanent impact on official statistics is the realization that *what makes statistics valuable is neither immutable nor universal*. Statisticians are eager to measure the ‘quality’ of their products but (as discussed in section 2.3.3), quality as understood by statisticians is not a synonym for value.

2.3 Delimiting concepts

44. Statisticians know well that conceptual clarity is key to producing meaningful, comparable and interpretable measurements. This is especially true for terms originating in everyday language and now used in a specialized context, as in the case of the word ‘value’. Hence rather than assume a common understanding of the concepts that will serve as building blocks, the following sections aim to establish the terminology of this report.
45. The discussion differentiates what value is from what it is not; reflects on the characteristics that differentiate official statistics from other statistics; and explores the actors that make today’s data market so sharply competitive. The chapter aims to clear a path towards the factors that determine the elements of value, which NSOs can measure and set as targets for themselves.

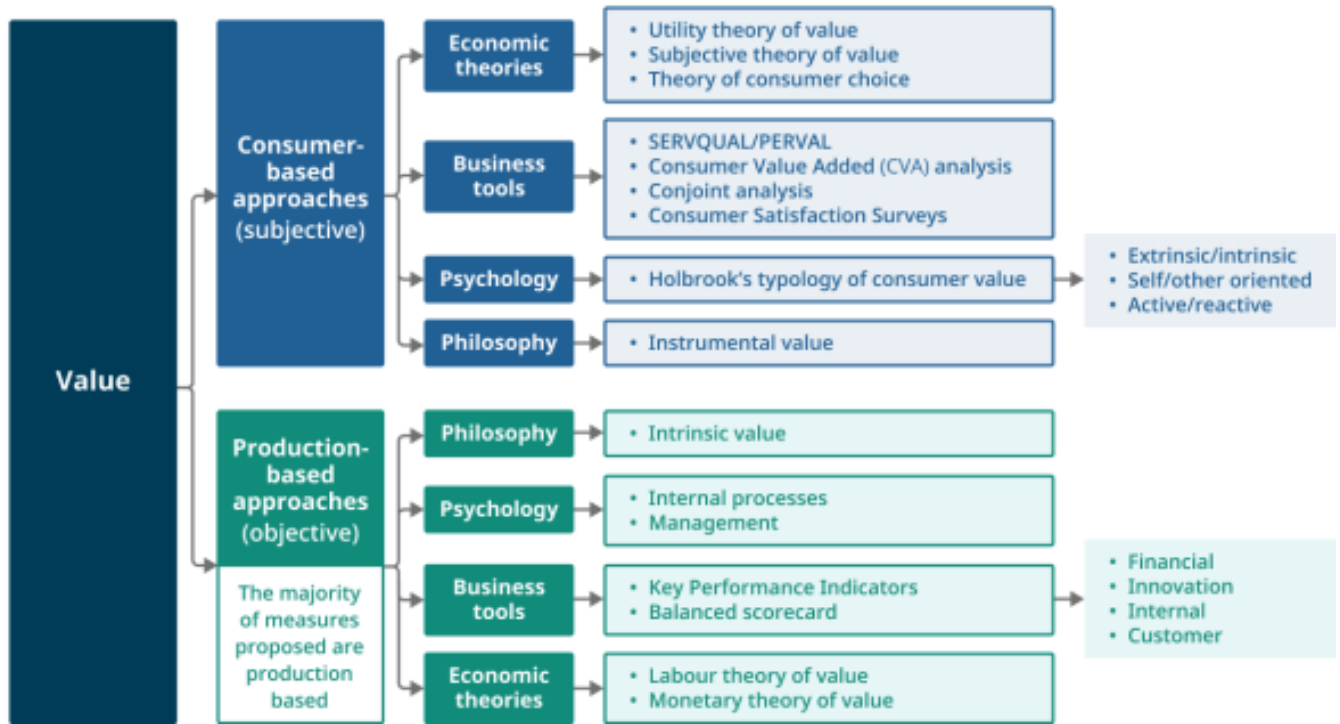
2.3.1 Value

46. The word value can be understood in many ways, and has been unpacked and defined in a variety of schools of thought, as illustrated in *Figure 1*, a concept diagram developed by the Task Force to map out the principal strands of thinking. While no attempt is made here to detail all of these approaches in an exhaustive manner, it is important to note that (as the diagram illustrates) there is a major bifurcation between **production-based** and **consumer-based** understandings of value. A production-based understanding of value is associated with the scarcity or abundance of a product (a good or service). This gives rise to the exchange or market value, i.e. the price, of that product. When used in this way, the word ‘value’ refers to something objective, determined by the market. A consumer-based understanding of value, in contrast, is associated with cultural traditions, human emotions and memories, and is therefore subjective and variable. Value understood in this way is not an intrinsic quality of any one thing: it is subjective, always perceived by someone within a context. Since official statistics are not a simple market good, their value cannot be understood by an objective consideration of price, but rather must be examined through the second, consumer-based lens. As the preceding Task Force on the Value of Official Statistics stated, it is rare for a product or service to be timelessly valuable. Therefore, value is a dynamic quality that has to be assessed and re-built continuously (UNECE 2018).
47. Terminologically speaking, value is what is known as a fuzzy concept. As such, it is first necessary to specify a determined context for what is meant by value, and then identify its component elements before measuring them. The present context is an assessment of the value of official statistics, from the perspective of users, for the demonstration and enhancement of that value by NSOs. The elements of value are determined by the qualities that users seek in statistics to inform their decision-making.

Figure 1: a conceptual map of value within different schools of thought

A conceptual map of value within different schools of thought

An objective of this report is to encourage a shift from how statistical organisations think about the value of official statistics, from production based to consumer-based approaches as value is determined by the customer.



[n.b. the final version of this diagram will have clickable links in which explanatory caption texts will appear as popups when the reader clicks any items in the diagram. For the purposes of CES consultation prior to making the clickable diagram, these texts can be found at the end of the document as Annex 5]

2.3.2 Value and values

48. A major trend in current discussion among chief statisticians and other key players in official statistics at present is to refocus attention on, and reflect on the meaning of, the 'core values of official statistics'. To a large extent these core values centre on the UN [Fundamental Principles of Official Statistics](#), as well as the ways in which these are enacted by NSOs in their work.
49. There is a very important distinction to be made between the value perceived by users and these values that NSOs appraise in themselves as institutions and in their own products and services. They are intermingled concepts that have important points of interaction, but they are not fully interchangeable.
50. First, while the value of official statistics changes as decision-making users react to the world and as social conditions evolve, the core values are set in place, designed to remain stable above temporary circumstances (as are the Fundamental Principles: hence the use of the word 'fundamental' to emphasize this continuity).

51. Second, value, like beauty, is in the eye of the beholder. ‘Values’, on the other hand, are things that we hold dear to ourselves, that shape our attitudes, motivate us, determine our priorities and guide our behaviours. For official statistics, these core values shape the way we see ourselves and the way we would like others to see us.
52. A person’s own deeply-held values may or may not align with what others find valuable in that same person, however. For example, someone might consider honesty and integrity to be the things that drive them, but ask others ‘what do you value in this person?’ and they might say their efficiency or good humour. There is no contradiction, it is simply the case that the two words refer to different things. That said, the values might lead directly to, or play a role in creating, the value: the values of honesty and integrity may lead the person to work hard, resulting in the efficiency that others value. The same is true in official statistics: our core values underlie and motivate what we do, and, we hope, result in features of our work that society values. But we cannot assume that they are, or should be, the main or only things that society wants from us.
53. NSOs would dearly love for their core values to be known and understood by society at large. Efforts to better communicate the core values and the Fundamental Principles are ongoing both in the international statistical community and in individual countries. Underlying these efforts is an assumption that if they are known, then official statistics will be better appreciated, or valued.
54. But this assumption makes a logical leap that may not be entirely founded. Just because someone understands your values and what drives you, this does not mean that they will necessarily share your view. It *might* influence their perception of you, but it might not. Indeed, it might even have a negative influence. When we learn a person’s motivation for their behaviour in daily life, this might lead us to hold this person in higher regard, or quite the opposite, depending on our own point of view. By the same token, simply ensuring that the core values and Fundamental Principles driving official statisticians are widely known, laudable as this goal is, is not enough. The present Task Force argues that communicating the values *and how they add value* is essential, as we cannot assume that this latter part will simply be deduced by society. And to do this, we must find out—from members of society—*what constitutes the value of official statistics* and hence how our work adds to it. This latter step is not something we can do alone. Introspection will get us so far, allowing us to examine our own beliefs and the central tenets upon which we have built our work, but without also examining the views of those assigning value to our work these efforts may be in vain (for more on this, see section 3.2).

2.3.3 Value and quality

55. For the purposes of this work, value must also be distinguished from quality. The official statistics community prides itself on quality. Indeed, quality is often touted as the hallmark of official statistics, the thing that sets our industry apart from statistics and data obtained from other sources. What do we mean when we say that official statistics have, or strive to have, high quality?
56. The International Organization for Standardization’s (ISO) definition of quality is the “degree to which a set of inherent characteristics of an object fulfils requirements” (ISO 9000:2015). The UN Statistics Division (UNSD) (2019 p.7) puts this definition into more everyday language:

“Quality: ...A simple definition is “fit for use” or “fit for purpose”. It is the users’ needs that define the quality. Different users may have different needs that must be balanced against each other.”

57. The [United Nations Fundamental Principles of Official Statistics](#), developed in the UNECE region in 1992 and subsequently adopted as a global standard by the General Assembly, describe the core precepts that underpin all the work of national statistical systems: not only statistical products and services themselves but also the processes, perceptions, relationships with government and with stakeholders, and ethical standards.
58. Quality Assurance Frameworks (QAFs) and related guides on their application, such as the [United Nations National Quality Assurance Frameworks Manual for Official Statistics](#) and the [Quality Assurance Framework of the European Statistical System](#) and various national and international adaptations of these³, are very closely linked to the UN Fundamental Principles. They operationalize quality as defined above in terms of a set of criteria or dimensions, grouped into those which refer to the *institutional environment*, those related to the *statistical production process*, and those which are concerned with the *statistical outputs* themselves. While the specific dimensions included in these frameworks can differ slightly across different countries, organizations, data types and product types, in general they comprise the following (UNSD 2019 pp.7-8):
- **Relevance:** the extent to which the statistics satisfy the needs of the users.
 - **Accuracy:** the closeness of estimates to the exact or true values that the statistics were intended to measure.
 - **Reliability:** the closeness of the initially estimated value(s) to the subsequent estimated value(s) if preliminary figures are disseminated.
 - **Timeliness:** the length of time between the end of a reference period (or date) and the dissemination of the statistics.
 - **Punctuality:** the time lag between the release date and the target date by which the data or statistics should have been delivered.
 - **Accessibility:** the ease and conditions with which statistical information can be obtained.
 - **Clarity:** the availability of appropriate documentation relating to the statistics and the additional assistance that producers make available to users.
 - **Coherence:** the ability to reliably combine statistics and data sets in different ways and for various uses. *Consistency* is often used as a synonym for coherence.
 - **Comparability:** the extent to which differences in statistics from different geographical areas, non-geographical domains, or over time, can be attributed to differences between the true values of the statistics.
59. It is possible, then, to use these standardized QAFs to assess the quality of anything an NSO does, ranging from an individual statistical product up to a statistical system as a whole.

³ e.g. Statistics Canada, Statistics Sweden, IMF, OECD.

60. So why not just rely on these frameworks? Why do we also need a framework to measure value? The simple answer is that quality and value are not synonyms.

The difference between quality and value

61. The two terms, and the concepts they imply, are of course very closely interlinked and significantly overlapping. Parts of the two concepts, and their component elements, are at times even identical. Adding to the complexity of untangling the two concepts is that *the extent to which they overlap differs according to the specific user and use*.
62. The principal difference between quality and value is the perspective from which they are construed. Statistical quality is, at its core, a concept designed *for the statistical office* in order to monitor and assure the standards of the work they conduct and the products they disseminate. As such, even though the definition of quality given above is ‘fitness for purpose’, a large majority of its dimensions can be understood and even measured without reference to the user and without recourse to investigations of users’ perspectives. Of the nine quality dimensions in the UNSD framework, only one, relevance, absolutely necessitates explicit reference to the user and their intended use (a second, accessibility, can be interpreted in terms that depend on characteristics of the user such as their degree of statistical literacy or on physical and mental capacities). This means that with the exception of the relevance dimension, quality assessments can be prepared on the basis of largely internal exercises.
63. In contrast, value, as understood by this Task Force, is intended to be *entirely customer-defined*: very similar, in fact, to the relevance dimension of quality, but also comprising other, broader aspects (see Chapter 3 for a discussion of these aspects). A framework for measuring that value may comprise elements that are quantifiable internally (everything from download numbers to customer service response times to timeliness of releases), but only a user can say whether these elements are actually components of value for them. **Quality, in essence, is the ‘degree of excellence’, while value is the subjective assessment of that quality that makes something desirable.**
64. An example can help to illustrate this difference. A child has a favourite soft toy. It is a rather poorly-made, cross-eyed, second-hand teddy bear. Whatever realistic criteria one might use to define the ‘quality’ of a soft toy, this toy does not meet many of them. But it is certainly of high value *to the child*. She loves it. The softest, handsomest, most well-stitched teddy bear with a silk bow around its neck would never meet her criteria for a valuable teddy bear. So it is with quality and value. For some children, the criteria will align entirely: some will love the ‘perfect’ silk-bowed bear. For others, the criteria will intersect, with certain being features of both quality and value (softness, for example, might make bears both higher quality and more valuable for many children). But for others—such as this child—the criteria she uses to decide that her teddy is valuable have next to nothing in common with those we might use to say that it is high quality. And we have no right or power to suggest that her criteria are ‘wrong’, that she should change her mind, nor should we have any interest in doing so. We can communicate quality, but we cannot *make* users value it.

So which matters?

65. Quality frameworks are designed to ensure that the statistics produced by NSOs are good, or ‘high quality’. They have a clear and obvious purpose and the work of this Task Force does not in any way intend to detract from the concept of quality, or from well-established procedures to define, assess and improve it. But quality is not a substitute for value. Both need to be

measured, for different reasons. **We measure quality to ensure that we are producing the best statistics we can, and we measure value to ensure we are doing what people want.** A restaurant producing excellent *haute cuisine* will not do well in a town where most people want fast food. The successful restaurateur conducts market research to determine the wishes of their potential customers, and so must NSOs.

2.3.4 Official and non-official statistics

66. With today's overabundance of data, NSOs may be gradually shifting towards a role that includes curation of data in addition to their existing role and creators of data. They are widely trusted by decision-makers and the public as a source of carefully selected information, processed to comply with the highest standards. Numerous systems are in place to safeguard the curation process and therefore the trust placed in the outputs of NSOs, particularly where it concerns official statistics. These structures ensure the credibility and impartiality of official statistics, whereas other products may be of varying quality and subject to external influences and interests.
67. In the context of this report, 'official statistics' refers to the complete catalogue of statistical products (supported by services in some cases) offered by NSOs and the organizations of their national statistical systems. The criteria for official statistical authorities and their products are stipulated in national laws and international guidelines, such as the United Nations Fundamental Principles of Official Statistics, or in the case of European Union Member States, the European Code of Practice. The main common criteria are:
- Official statistics are founded on legal mandates for NSOs and the national statistical system, who are helped in their tasks by established quality frameworks and guidelines for best practices.
 - Official statistics are publicly available.
 - The methodology and interpretation of data are non-partisan, objective and transparent.
68. Official statistics stand apart from any other statistics in the market by virtue of being a public good founded upon these criteria. The benefits and conditions that NSOs retain for the production of official statistics do not apply to any other private statistical providers. Among these feature the scope of official statistics, mandated by law at national level. The state-funding of NSOs also means that they have the resources to conduct useful research. On the other hand, the obligations of NSOs to carry out statistical work on all relevant sectors do not allow them to specialize in the same way that private statistical providers can. Additionally, regulations surrounding data confidentiality have different applications in the public and private spheres, meaning companies may publish some data more readily.
69. There may be some overlaps in the content and methodology of official and privately-produced statistics. However, users do not turn to NSOs only for specific products for which competitors offer no alternatives: they assess the value of official and non-official statistics for their own purposes and make a choice suitable for themselves.
70. The concrete reasons behind a user's preference of a product of official statistics over non-official statistics – or vice versa – is known only to the user, which is why it is essential for NSOs to intensify their dialogues with users of all levels. Here it is important to keep in mind that the

ability of diverse user groups to articulate their criteria and their data needs may be varied, and thus challenging to pin down. Nevertheless, NSOs should not presume to know pre-emptively how and where users perceive value in statistics. Only users can answer this.

71. NSOs should also understand that official statistics are not indispensable to all users at all times. Competitors will surge where there are gaps of information, but based on assessments of value, it may not be necessary for NSOs to attempt to fill in such gaps. More importantly, they cannot cater to every single arising statistical need.
72. As NSOs step more into a data stewardship role and expand their ability to influence their data ecosystems⁴, they will be better prepared if they have a concrete understanding of where official statistics deliver value. Re-thinking this role in an increasingly commercialized sector is a challenge, but remaining static in such a volatile environment would risk erosion of the public perception of value of official statistics compared to the wider data market. Any perceived disparity may also increase the level of scrutiny on the value delivered by NSOs through official statistics.

2.3.5 The lifecycle of statistics

73. There is an established tool for the analysis of the lifecycle of products⁵. It is focused on the strategic choices and actions of businesses as they attempt to best position their products in the market while gaining an understanding of the changes in the behaviour of other industry players during the lifecycle phases of their products. Such an analysis is also applicable to official statistics in the data market
74. The general stages of a product lifecycle are: introduction, growth, maturity and decline. The introduction phase is characterized by fewer industry players and inconsistent quality in an innovative product. With increased growth, new players emerge in the market while the quality and design of the product improve; when maturity is reached only a few larger players remain and the standardization of products take place. Finally, the stage of decline is characterized by a falling demand, with less emphasis in the differences of the products, but more on brand loyalty.
75. NSOs should remain aware of their position vis-à-vis other providers in the data market and consider how these market shifts may impact their role and the value they offer in the future. The stages above can also be observed in the supply and demand of statistics. However, a difference between any other products and statistics is that, in the case of the latter, the usability of the data upon which statistics are built is time sensitive. Data may peak in importance at a given moment and later will either lose or gain relevance by facilitating tracking points over time. Moreover, official statistics are often revised as a solution to improve accuracy without compromising the timeliness desired by users. This way, the stages of the lifecycle of statistics reinforce the earlier characterization of value as dynamic.

⁴ <https://unece.org/statistics/documents/2021/05/working-documents/how-national-statistical-systems-adhere-core-values> [accessed on 25 June 2021]

⁵ <https://kfknowledgebank.kaplan.co.uk/business-strategy/strategic-analysis/product-lifecycle-analysis> [accessed on 5 August 2021]

2.3.6 Actors in the statistics market

76. Before embarking on modes for the measurement of official statistics, it is important to understand the role of the most relevant stakeholders involved in the supply, production, consumption and multiplication of statistics. The following list describes their relationship to official statistics through the lens of value.
- **Users** . as consumers of statistics, are at the core of the activity of national statistical systems. In their capacity, they ultimately decide which products better suit their purposes, and are therefore the most valuable to them. As a very diverse group, users carry different weight and significance to national statistical systems. Users can also be classified based on sectoral, demographic, decision-making or (statistical) literacy levels, among other features. They are the determining factor on whether the production of a statistical products is worth it, as they decide whether it is worth it for them to make the effort of using statistical products for meaningful decisions in their spheres of influence.
 - **Non-users or passive users**. They can also be regarded as potential users, which makes them a group of high interest to NSOs for further research. Some non-users or passive users are not active stakeholders in official statistics, but they are affected by decisions that other users make based on statistics. Other non-users might be actively engaged as stakeholders (e.g. as survey participants, members of census focus groups, etc.), but still do not actually use the statistics themselves.
 - **National statistical systems**. The structure of NSSs are varied, and their position in the centralized to decentralized continuum is dependent on the national institutional context. NSSs can entail a large number of authorities or institutions, each authorized by law to collect data for official purposes. NSSs approach value from a production (as opposed to consumption) perspective, and therefore place great emphasis on standards and guidelines.
 - **Data providers**. The persons or entities providing data in response to direct data collection. They are the traditional source of official statistics – although waning in magnitude as other methods of data collection emerge and become popular. Data providers value the data they transmit to the statistical system as raw material for the statistical process. Their task can require significant efforts in time and capacity, and this burden cannot be neglected by NSSs.
 - **Private data holders**. They are relatively new entrants in the data market and own datasets in an extensive range of topics. In many countries, they are considered in the group of competitors of NSOs. For private data holders, such data sources may also offer a different lens on the value of statistics and the monetization of data as a source of revenue and competitive advantage.
 - **Information multipliers**. A group comprised by media, educational institutions, libraries and other entities that support the dissemination activities of statistical systems. Also worth noting that they can also be intermediaries or stats propagators who share statistics publicly in their own right
 - **Governments and budget offices**. Governmental organizations, including public administration, can be regarded as an exceptionally important user group. However, they also make decisions on the resources allocated to NSSs. From the perspective of budget offices

and their equivalents, the value of statistics lies in monetary efficiency and financial transparency.

- **International statistical organizations.** They are statistical divisions or directorates of intergovernmental organizations to which national statistical systems transmit their official data, and through which standards and best practices are developed. Like national statistical offices, their approach to value is through the perspective of production.
77. As shown, the production of relevant, high quality statistics involves the direct or indirect participation of many different actors who approach value from a perspective of either production or consumption. These are not clear-cut categories: for example, data providers may also be users. Governments may decide on the budgets of NSOs, but they are also a very important user group. There are also other overlaps between academia and businesses, or private persons and civil organizations. Regardless of the precise categorization, users' perception of value depends on their needs and whether they are being met.
78. It is well understood that official statistics are not meant to stay within the walls of NSOs: the statistical process is only completed when users refer to them to learn and make evidence-based decisions. This is the reason why this report argues, as will be seen in the next chapter, for a shift in the mindset of NSOs to understand that a consumer-based approach to value is fundamental. Among the actors in the statistics market, users are at the centre and all other actors face them.

2.4 Conclusion

79. In belonging to the public sector, NSOs have grown accustomed to thinking of users in terms of entities or citizens classified into segments to reflect their use of official statistics. In doing so, they should not lose sight of users also being customers and consumers in a wider data market, whose behaviour is shaped by their perception of the value of *all* the statistical goods and services they encounter.
80. In the private sector, value is at the centre of transactions. Monetary gauges, in the form of prices, revenues and willingness to pay, are readily available and routinely used to calculate targets and movements in the market. For private companies, therefore, it is easy to see how price can usually be thought of as an excellent proxy for value. This is the origin of 'production-based' understandings of value.
81. Any attempt among NSOs to quantify the value of their products and services in terms of solely money can, however, never be simple or self-evident, nor can such a measurement hope to encompass the full impact of official statistics on society. Nevertheless, it would be greatly beneficial for both NSOs and the public to have the ability to refer directly to that impact as part of a dialogue on the role of official statistics and the wider national statistical systems in responding to present-day circumstances, such as the Covid-19 pandemic, and other situations that can be expected in the future. As the data landscape continues to expand and NSOs evolve into a role as data stewards in an increasingly commercialized sector, they will be favoured by their ability to demonstrate the value delivered to users through their products and services.

82. Recognizing the wide variety of concepts encompassed by the word 'value', as was illustrated above in **Error! Reference source not found.**, is key to realizing that NSOs do not have a simple task ahead of them if they wish to measure the 'value of official statistics'.

Chapter 3. Daring to be different: a new proposal for understanding value

3.1 Introduction

83. The preceding chapter made the case that NSOs must strive to understand and quantify value, not only quality; and that they should do so in a way that recognizes that value is determined by the user, unique to each user and use, and not necessarily or perfectly aligned with what we as statistical producers think are or should be the most valuable features of our work.
84. But if value consists of different criteria and components for each ‘valuer’, how can we hope to measure it? Before moving on to discussion of a measurement framework *per se* in Chapter 4, this chapter proposes a model to help understand the space which these varying criteria and components occupy.

3.2 Value from a customer perspective

85. In the official statistical community, the most common term for those who access our products and services is ‘user’. But we must remember that users, and all members of society, are also customers whether that be as a direct user or as a non-passive user or beneficiary. Indeed, the ISO defines the word ‘customer’ to encompass other terms such as consumer, client, end-user, beneficiary, etc. (ISO 9000:2015). Customers get to choose what they want to consume; and if one provider does not fulfil that wish, they are free to look elsewhere.
86. Customers come in all shapes and sizes: agencies, organizations, individual citizens and everything in between. On the surface, these customer segments are very different. However, what they have in common is that they all determine value based upon their internal perceptions: perceptions of their needs, perceptions of the product, perceptions of the provider, and perceptions of how these aspects are interlinked. It is these customer-held perceptions of value that shape and influence their choices regarding which products and services they will use (if any) and when, including the selection of data sources.
87. For a long time, much of the focus of NSOs’ efforts to assess their own work has been centred on assessing characteristics of the products and services themselves—whether they are released according to schedule, whether they are accompanied by comprehensive metadata, whether they were produced in accordance with the latest international standards, the size of their confidence intervals, their completeness, the time lag between collection and release, etc. That these are the things that are valued by customers has been taken as given. Until recently, little has been said about whether things other than the characteristics of the statistics themselves play a part in customer-defined value.
88. This has begun to be rectified with discussions at the international level about the trustworthiness of official statistics – as an entity or an endeavour, rather than in reference to individual statistics – as one of its defining features. Even so, however, while much has been said about the importance of demonstrating trustworthiness and maintaining trust, how to measure it, and how to achieve and safeguard it, much less has been said about how far

customers truly do put trustworthiness at the core of how they determine value, and how we know this.

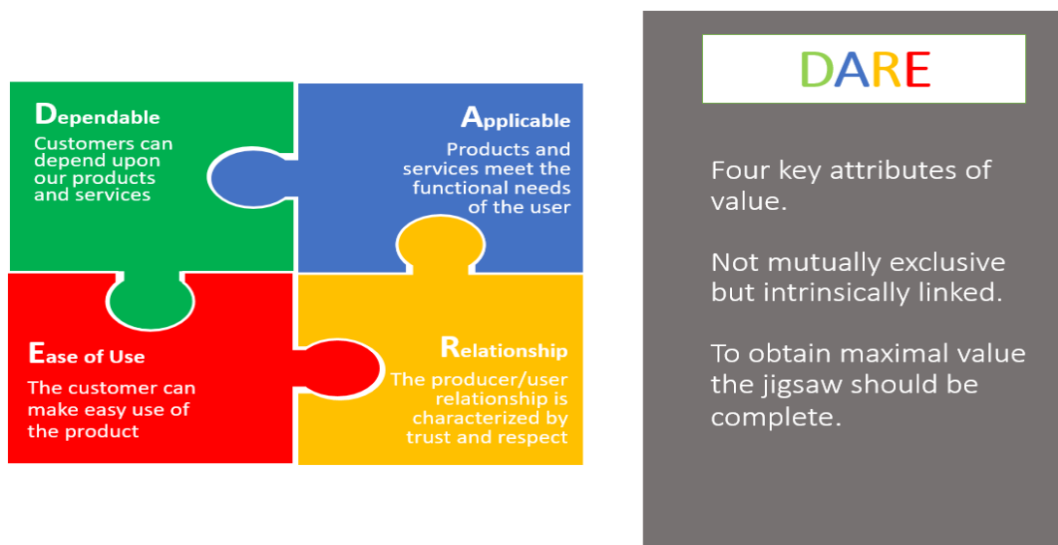
89. The solution may then seem obvious. Let us simply ask customers what they value in official statistics. Let us ask them if they care about trust, ask them which of the quality dimensions are most important to them, and ask them what they think about our organizations.
90. In fact it is not so simple, since we cannot expect all customers to be conscious of or able to put into words what it is that makes official statistics valuable (or not) to them, just as market researchers cannot expect customers to be able to articulate directly the range of criteria they apply, or the relative weights they assign to these criteria, when choosing a car, a pair of jeans or a brand of cookies.
91. The decision on what is valuable still needs to be made by the customer. The DARE model proposed below suggests that thinking of value along four interlinked dimensions is a first step towards being able to elicit the value criteria that our customers apply. Customer-perceived value will differ across customers and product segments as the criteria they apply differ, but all will fall somewhere within the space identified. Direct research with users is needed using the DARE model as a lens to inform decisions in order to be customer centric.

3.3 The DARE Model

92. The DARE model builds on the framework that was developed by Stats NZ (2018-19) that explored the nature of value determination of four attribute types that customers frequently consider when making a value-based usage decision. Collectively, these four attributes offered a set of lenses through which to contextualize the customer perspective of value for a given product or service.
93. Using this model as a starting point, this Task Force defined a model in which the four attributes are clearly and deliberately distinguished from quality dimensions and from terms which have an accepted specific usage in statistics (terms such as relevance, reliability and accessibility, which feature in the original model). The model is illustrated in *Figure 2*. The single-word descriptors shown in the figure are complemented by statements to further explain what each dimension means.
94. The dimensions are wide, and overlapping. Their broadness reflects the fact there are many possible value criteria within any given dimension. Within 'ease of use', a school student might value easy website navigation, while a civil society activist might value clearly explained visualizations. Within 'relationship', some may find rapid customer service responses to be valuable while others are impressed by outreach sessions on planned census dissemination products. Dimensions may overlap, for instance where the extent to which a statistical product meets a user's needs (applicability) is seen by that user as resulting from effective NSO-stakeholder consultation (relationship). The relationship of these dimensions to the dimensions of a typical quality framework is not clear-cut. 'Ease of use' and 'Applicable' certainly encompass some of the standard quality dimensions (relevance, accuracy, clarity, accessibility, for example). 'Dependable' also requires some degree of accuracy and punctuality. But it is not possible to neatly slot quality dimensions into the DARE quadrants, since, as explained in section 00, quality and value are not synonyms.

95. The DARE model is not, in itself, a measurement framework. It does not offer indicators, or scales. Instead, its function is in allowing an organization to have a structured and considered conversation about the value that a given product, or indeed the office as a whole, is offering.
96. To do this, consideration should be given to each of the four attributes for a product relative to the customers' preferences. Different customer types may have a different value profile. That is, not only do the specific criteria within each dimension differ, but the relative weight assigned to each of the four dimensions can vary. For example, a corporate customer might attach more value to product dependability rather than the relationship, whereas an ordinary citizen might favour ease of use.
97. Nevertheless, in principle, maximum value is delivered where all four attributes are present; the absence of anyone, or more, dimensions might be an indication that something is lacking in the value proposition. For this reason the DARE model can be illustrated as jigsaw puzzle (see *Figure 2*), where all of the pieces are required to complete the overall picture. Agencies must therefore consider how best to address these areas in a manner that recognizes that they are not mutually exclusive, but rather intrinsically interlinked and requiring holistic consideration. The reality is that true value is achieved where the product profile matches the customer profile.

Figure 2: DARE – Jigsaw of the four attributes of value



3.1.1 Using the DARE Model

98. The DARE model is offered as a tool through which to consider customer value when designing or evaluating a product or a service with respect to a given customer segment. Quantitative indicators of value (by whatever means they are developed—see Chapter 4 and Chapter 5) can be situated within the four dimensions. This will shine a light on the role of each of the four attributes in characterizing a given product or service: a 'value profile'. By reconciling this profile against the known customer preferences, organizations can understand where their products offer the most and least value relative to each customer type.

99. In this manner, the DARE model can be an effective tool to identify areas where the value proposition can be improved. This means that organizations can make informed investment decisions based upon what is truly adding value to their customers.

3.4 Conclusion

100. Recognizing, explicitly that users (and non-users and passive users) are customers can help to focus effort on fulfilling their needs. The customer-centric perspective afforded by the DARE model can help us to look beyond features of our products and services alone, to encompass also the relationships that are important in determining value. The four overlapping dimensions can help us to map out the diverse criteria for value that customers may apply, and that may be fulfilled to different degrees by different products and services as they relate to specific users or user groups. Viewed through this lens, it becomes clear that 'a measurement framework for measuring the value of official statistics' must be something flexible, wide-ranging, outward-looking and inclusive of subjective perspectives .

Chapter 4. Reviewing the measurement framework

4.1 Introduction

101. The principal activity assigned to the Task Force according to its terms of reference was to review the proposed indicators of value included in the 2018 framework, assess their suitability as measures of value, and provide refined guidance on producing, interpreting and using them. An explanation of how this work was undertaken and the detailed results are included as a technical annex (Annex 2).
102. The review began with collecting information from participating countries using a template which was collectively developed by the Task Force. This exercise involved a wide consultation of many staff across a range of departments, asking countries to report on which of the 'objective', 'subjective' and 'monetary' measures proposed in the framework (including some additional ones brainstormed by the group) they currently produce, why, and how. Completed templates from 18 countries and organizations provided a rich source of information. The findings were analyzed to determine whether each measure offers potential as an indicator of value.

4.2 Guiding principles for refining the measurement framework

103. In reaching their decisions on measures to be retained in the framework, the Task Force was guided by the following principles:
 - **Measures should be clearly indicative of the value of official statistics.** As discussed in section 2.3.3, this is not simply a synonym of quality. Therefore, even though a measure may very obviously be a good measure of statistical quality or of adherence to core precepts of official statistics, this does not necessarily mean that it merits inclusion in a framework for measuring value
 - **Measures should be (at least theoretically) quantitative and have a monotonic relationship with the aspect of value being measured.** That is, a greater measurement indicates more value and a lower measurement indicates less value, even if the relationship is not necessarily linear.
 - **Measures should lend themselves to the formulation of actionable targets,** i.e. it should be evident what a 'good' level of the indicator would look like, and it should be possible to envisage how an NSO could take action to harness the information gleaned by producing the measure, to inform some action or behaviour that could alter the level of the measure (that is, that could improve the aspect of value being measured). Measuring something which is entirely externally caused may be of interest for various reasons (and may have important uses for the business operations of the NSO) but it would not permit any action to improve the value of official statistics based on the measurement.
 - **The potential unintended consequences of producing a measure should be considered.** The act of measuring anything is never neutral. By measuring something we imply that we find it

important, and whether deliberately or not we may turn it into a goal or target, which can affect the behaviour of those whose actions contribute to the thing being measured⁶.

104. Measures identified by many countries as long-standing key performance indicators are evidently considered important and are therefore typically maintained in the framework.
105. As well as categorizing the measures in the framework into those that are retained and proposed measures of value versus those that are not recommended, the Task Force also attempted to distinguish between measures that are grounded in a production-based conceptualization of value and those grounded in a consumer-based conceptualization of value. While this distinction is by no means clear-cut (some fall into both categories, and for some the categorization is unclear), this undertaking does go some way towards framing the selection of measures of value in a way that could help NSOs to reflect on the basis for their selection of measures. *Figure 3* illustrates an initial attempt to show such categorization (with those highlighted in green as production-based, those highlighted in blue as consumer-based or a combination of both).
106. It is important to note the choice of terminology used for the selection of indicators below. It was decided not to use the term ‘recommend’ for potential indicators, as it is difficult to recommend specific potential indicators as a means of measuring value across all national contexts. Rather, those retained as potential indicators of value serve as suggestions for areas that NSOs may wish to explore as they develop their country-specific strategic Results Map as described in the following chapter. Conversely, where measures are not retained in the framework when assessed according to the guiding principles detailed above, this should not be interpreted as “we do not recommend that NSOs produce this measure” but rather “we do not recommend, on balance, that NSOs include this measure *in a framework for assessing the value of official statistics*”.
107. **The resulting list of retained indicators is not an exhaustive list of recommended indicators for measuring value.** Far from it. The starting point for this exercise was simply the list of indicators contained in the earlier framework, which were reviewed and evaluated. The subsequent work of the group revealed that if a more targeted approach to measure development were used (see Chapter 5), entirely new and different measures may be proposed.

⁶ It is essential to recognize that, even if an NSO has not explicitly framed a measure as a target, it may become one. Even if we say that it is ‘just a metric’, the act of measuring influences behaviour, ultimately impacting on the thing being measured. If a team is judged on a key performance indicator, they will strive to improve that indicator, perhaps even at the expense of other tasks. If an NSO decides to measure ‘value’ using a particular suite of indicators (and assuming that ‘increased value’ is the overall goal), then the collective efforts of the NSO will, rightly or wrongly, aim at enhancing those chosen indicators, perhaps letting other aspects of value that are not included in that suite fall by the wayside. For example, if ‘number of retweets of NSO’s Twitter posts’ is taken as a measure of user engagement, might this inadvertently induce social media managers to post more potentially popular material, maybe at the expense of more ‘serious’ statistical material that would better inform users but be less susceptible to being retweeted? And is this a good thing or a bad thing, or does it not matter? Another example: if error detection time is taken as a measure of accuracy, might this lead to statistical teams taking longer to check and release their figures, and in so doing reduce timeliness? This is not to say that this would be a bad thing—but simply that this possibility, or more generally the possibility that adopting a measure might influence behaviours, should be taken into account.

4.3 Overview of findings

108. A full explanation of how this work was undertaken and the detailed results are included as a technical annex (Annex 2).
109. The major findings of this work were:
 - a. a large proportion of the indicators were not sufficiently developed to be able to ‘pilot test’ (at least not without first refining and developing them)
 - b. many indicators, while being potentially useful for a range of purposes, could not be proposed as measures *of the value of official statistics*
 - c. many indicators, even when they can be viewed as measures of value, are formulated from the point of view of the NSO and its production processes (‘production-based’ measures: see section 2.3.1). These may be of use to the organization for management and operational purposes, but they neglect or downplay the perspective of the customer.
110. *Figure 3* shows the list of measures proposed as potential indicators of value, categorized both along the objective/subjective/monetary classification and according to whether they are principally production-based or consumer-based. Note this is not a list of measures that have been developed, tested and are now being recommended by the Taskforce to show value, this is a just a tentative list of measures drawn from the framework that the Taskforce had to review.

Figure 3: Tentative list of measures drawn from the proposed framework that offer some potential to show value, indicated by whether they are production-based value (shaded green), consumer-based value (shaded blue) or both. (shaded blue and green)

Production Based		
Consumer Based		
Objective Measures	Subjective Measures	Monetary Measures
Punctuality of statistical releases (share of punctual/late/cancelled releases)	Average reported 'ease' of users finding what they were looking for (accessibility)	Cost based approaches - cost benefit analysis and ratios*
Share of error-free statistical releases	Average share of users who found what they were looking for (or those who couldn't find what they were looking for)	Stated preference methods - willingness to pay (includes a combination of the approaches)
Quick correction of errors (Average delay in correction of errors in releases)	Average share of users whose information needs were met	Newly suggested by Task Force Conjoint analysis that includes revealed preference methods and willingness to pay)
Accuracy of statistics - Average number of revisions required	Average rating of overall quality of website	Impact assessment and frameworks
Timeliness of statistical releases (weeks from the reference period)	Average rating of ease of navigation on website	
Number of website visits	Average reported ease of finding what they were looking for on website	
Downloads of statistical data by domain	Average ease of locating data in database/data warehouse	
Visits to digital library/publications webpage	Average ease of extracting/downloading/manipulating/visualizing data from database/data warehouse	
Number of social media followers	Average satisfaction with visualizations/interactive tools/maps/graphs/etc	
Number of agreements to use microdata for research	Average ease of locating relevant metadata	
Number of media citations	Average satisfaction with available metadata	
Number of citations in research/policy work	Degree of trust in official statistics	
Most cited statistics	Degree of belief in the political independence of NSO	
Most viewed/downloaded statistics	Degree of understanding of the remit of the NSO	
Number of retweets (comments, likes etc.)	Newly suggested by Task Force Future proofing of official statistics - E.g. data linkage across governments, analytical and coding capabilities of staff	
Newly suggested by Task Force Exploration of Digital Object Identifiers (DOIs) to monitor use of official data in online publications	Increasing statistical literacy through school engagement and resources	
Interoperability & link ability - how easy is it to work with, aggregate and join up data sets	Share of users who recognize the corporate branding of the NSO Share of users who have heard of a specific product	

4.4 Countries' experiences from efforts to measure value

111. A large number of case studies has been gathered from countries on one or more of the indicators included in the original proposed framework. It should be noted that initially case studies were sought which demonstrated the application of measures included in the original framework. However, as the preceding chapters show, many of these measures were found not to be clear metrics of value in the way that was intended. Nevertheless the case studies have still been included in this report since they offer an insight into NSOs' experiences in their attempts to show value, whether successful or less so.
112. The full set of detailed case studies is included at the end of this document as Annex 3. This section reviews the key findings of some of these case studies as they relate to the aim of refining and developing the measurement framework.
113. Some case studies refer to indicators which do have potential as measures of value (either production-based or consumer-based), while others detail experience with measures which have not ultimately been retained (in figure 3) as potential measures of value.

114. Of the 39 measures which have been tentatively retained from the original framework as having potential for measuring value (either production-based or consumer based), case studies have been gathered on all but eight measures. Further case studies are welcome and could be included in a final online version of this publication which the Taskforce is proposing, as an ongoing resource, that should be maintained. To note, of the retained measures (in figure 3) no case studies were submitted on the eight below.:

- *Objective*
 - Share of error-free statistical releases and quick correction of errors (Average delay in correction of errors in releases)
- *Subjective*
 - Share of users whose information needs were met
 - Ease of locating data in database/data warehouse
 - Ease of extracting/downloading/manipulating/visualizing data from database/data warehouse
 - Ease of locating relevant metadata
 - Share of users who have heard of a specific product
- *Monetary*
 - Impact assessment (and frameworks).

4.4.1 Case studies of countries' experiences with proposed 'objective' measures of value

115. Some of the case studies illustrate the ways in which NSOs have attempted to assess their adaptability and flexibility with respect to meeting changing user needs, including where those needs have been affected by the Covid-19 pandemic and the accompanying sharp surge in demand for timely and granular data. Case studies in these areas illustrate the ways that NSOs strive to monitor the value that their services and products offer, both in cases of well-established user needs and during unexpected crises.

- [“Covid impacts web analytics” from Canada](#) which provided insights into the increased interactions with their website during the pandemic and how COVID-19 products impacted the web traffic. Using this as a measure allowed for Web analytics to better understand user engagement and evaluate the reach, retention and effectiveness of website content.
- [“Measuring the impact of COVID-19 ONS publications” from the UK](#) which undertook a review using their existing measures and other newly sourced metrics to assess the impact of covid-19 publications (which included 119 different outputs released between March 2020 and June 2020). Part of that review was also to assess how well impact was measured. It found that quantifiable metrics were of little value and used for indication purposes rather than a measure of success or failure and not suitable for effective evaluation of output.
- [“Measuring value through website monitoring and evaluation in INEGI” from Mexico](#) focused on four main indicators for their website: Accessibility: Ease of the user in terms of access and location of information of interest; Level of detail: Satisfaction with the level of disaggregation of the information obtained; Navigation: Satisfaction with navigation on the portal and Site Rating: Rating given to the INEGI Site on the Internet in general. Three sources of information: a permanent survey held on the website, a quarterly questionnaire

(to one in 5 users) and Google Analytics. Results were used to contribute to institutional strategic planning and to support design actions to improve the usability of the website.

- “Communications and Covid-19: Assessing the impact of the ONS’s messages” from the UK provided insight into a ‘dashboard’ on communication, premised on media analysis, social media monitoring and website analytics which formed the basis of weekly and monthly evaluation of communication at ONS in response to the coronavirus pandemic. The dashboard used media monitoring tools, suited to tracking specific messaging and also used freely available analytical tools as well as paid-for services to gather data and monitor a rapidly changing situation. The key purpose of the dashboard was to enable action with timely insight, and to understand and demonstrate the impact of communications from a wide public audience (and therefore underline the value of ONS as a trusted and authoritative source of relevant information during the pandemic)..
116. “Internal dashboard tracking user activity” from Hungary uses an internal dashboard to provide concise and visual information with the purpose of supporting management decisions regarding the dissemination of statistical products with transparency on a quarterly basis. The Hungarian Central Statistical Office has taken this approach to measuring value, because in order to best fulfil users’ needs, it feels it necessary to understand who requests which data, how they use data, and how satisfied they are with the products and services provided by the office. One of the newly suggested measures put forward by the Task Force builds on the objective metrics above by tracking the ‘use’ of official data in online publications using Digital Object Identifiers (DOIs). These are already seen as an important factor in the academic and scientific world and are already used by some NSOs. The US and Canada shared their experience in case studies, and the UK has been exploring the potential of DOIs.
- “Digital Object Identifiers: Assessing the impact of our long-form analysis”. DOIs are alphanumeric strings that can uniquely identify an article, document, or dataset and are commonplace in academia as a way of citing research and data. The U.S. Bureau of Labor Statistics (BLS) primary objective in using DOIs was to maintain publication’s ability to be indexed. and to help inform on the value of the articles published in the Monthly Labor Review(the BLS flagship publication, publishing long-form data analysis for over a century). Though BLS does collect basic web metric information on each article, DOIs provide an opportunity to delve further into the use of articles as well as glimpse who is using the research. Citations can help provide a much more in-depth look at measuring the value of long-form analysis than Google Analytics which can provide views and downloads are but these measures are incomplete. Citations help show how many customers are citing the research and put a name to those using the information.. This type of information could help construct more qualitative measures around the citation.
 - “Canada: Exploration of Digital Object Identifiers (DOIs) to monitor use and impact of official data in online publications” - This feature was introduced to Statistics Canada’s web products in 2019 and appears as an alphanumeric string of characters that acts as an active link to the original digital object. It provides a way to cite a digital reference. It also allows the organization that owns the DOI to retrieve metrics on the number of times the DOI was used/cited. Still at the experimental stage, Statistics Canada does not currently use this indicator in any of its official reports.
 - Although the DOI process is easy to use, it is still in its infancy, as such it does not yet have widespread acceptance as an instrument to cite. There are limitations in its scope as a

potential for use as an indicator. Though, it shows potential for insight into the breadth and depth of the use of products in academic and scientific journals and allow tracking of the continued use of statistical products over time.

- “Recognizing and considering the value of DOIs for measuring impact”. The idea of using DOIs on outputs is also being explored at the ONS. Beginning with a pilot looking for the uptake of DOIs for digital publishing and the Secure Research Service looking at implementing to track outputs of technical users.

117. Other case studies illustrating experience with measures included in *Figure 3* include those dealing with accuracy, timeliness, reliability, transparency and punctuality.

- “Accuracy and reliability metrics, impact evaluation: Informing users of the accuracy and relevance of statistical data during a pandemic”. The U.S. Bureau of Labor Statistics (BLS) looked at the link between value and the needs of users. It reported on how the COVID-19 pandemic led to widespread changes in their standard practices in their surveys. This included recognising methodological challenges with changes to data collection modes and the introduction of new data sources in preparation for an anticipated decline in response rates. In recognition of potential issues on impact on accuracy, comparability and for relevancy and transparency for users’, in April 2020, the BLS standardized a set of accuracy and reliability metrics to compare all major data releases from May’ 2020 to pre pandemic.
- “Punctuality of statistical releases in INEGI”. Mexico developed a series of indicators as part of the Institute’s quality program that make it possible to monitor compliance with the quality principles established by the Quality Assurance Committee
- “Official Statistics Mentions in UK Parliament”. The UK has provided insight into some of the impact that official statistics has on debates in the UK Parliament in the House of Commons and the devolved nations. By measuring the number of mentions, by implication, the measure shows how often official statistics are used in order to bolster arguments for and against policy.
- “Measuring value through the number of agreements to use microdata for research” Armenia is currently using the use of microdata by researchers as a key performance indicator (<https://www.armstat.am/file/doc/99523493.pdf>). The measure facilitates the user-producer communication and users feedback, and helps to more effectively address users (researchers) demand and promote data use and analysis
- “Number of research projects with access to microdata (virtually and secure laboratory)” Mexico use this as an indicator to follow up the demand of the microdata access services, which type of service users prefer and the number of projects registered. Information on the measure is generated from a report on successful microdata applications. As a result more microdata labs were created where needed, making the service more accessible.
- “Producing datasets in open format” (interoperability & linkability, a suggested area to measure from the Task Force, looking at how easy it is to work with, aggregate and join up datasets). Canada has defined this indicator as the percentage of datasets that are publicly available in an open (non-proprietary) format. Non-proprietary formats allow datasets to be interoperable, allowing users to easily ingest data into their own system. Open data is an important element that is part of the Open Government strategy and national action plan on Open Government.

- Statistics Canada has been producing datasets in this open format for over 25 years which allows them to demonstrate their leadership in this space. Their organization's target is to have ALL datasets available in an open format.

4.4.2 Case studies of countries' experiences with proposed 'subjective' measures of value

118. Subjective measures such as perception, satisfaction, trust, public confidence and brand awareness etc offer a way in to 'consumer-based' understanding of value. It is essential to offer a crucial insight into the impact of the outputs produced by NSOs. 'Trust' and 'Branding' are areas highlighted by users as important.
119. Canada is in the process of developing its "Trust Centre" in order to promote openness and transparency. It is looking towards a whole agency effort to improve public trust through external engagement and partnerships areas of focus include identifying and segmenting audiences for focused communications and engagement. They are looking to earn social acceptance and not just to rely on legislative authority but to collect and disseminate data in new ways; Looking to conduct social-acceptance research and behavioral analyses to identify actionable insights, reach persuadable non-respondents along with developing narratives to improve public understanding; Develop crisis communications plan to prepare for risks and mobilize knowledge translators & thought leaders to engage with Canadians. To measure impact and progress they will be identifying key performance indicators to monitor StatCan's ability to measure ongoing public sentiment. Looking to earn social acceptance to collect and analyze data in new and innovative ways, turning risks into non-events and managing issues to avoid reputational crises. A key aim to deliver relevant, widely-used data to build a fairer, more inclusive Canada.
120. Slovenia sees trust as a key performance indicator as it combines the objectivity, impartiality, transparency and independence of the organization and the data an organization publishes. The Statistical Office of the Republic of Slovenia (SURS) has measured the trust of an organization and trust in the statistical data published continuously (in annual customer satisfaction survey) since 2019. Before that, it was measured occasionally. Results have given valued feedback on work and learn who trusts them less (user groups) and act on that if necessary (what are doing wrong, what should improve etc.). So far, trust is higher every year so they have never acted on the results. To try to reach as many users as possible, they publish invitations to participate on all dissemination platforms and achieve a high response and results and findings published on the website
121. The ONS is seeking a timely and relevant measure of its impact and value amongst public audiences through a 'Quarterly Brand Survey'. A question set of eight core questions is repeated in each wave of the survey in order to track change over time. The core question set was largely designed to measure public agreement with a series of reputationally relevant statements relating to qualities such as trust, relevance, and perceptions of the organization's impact at an individual, community, and societal level.
122. These measures have direct relevance to value to the customer and gathers public audience feedback. It is designed to understand and demonstrate perceptions of value amongst public audience groups. As an internal-only measure, the measurement is designed to be useful to decision makers within the organization to measure performance and inform action.

123. [Mexico’s ‘social perceptions of INEGI’ study](#)” looks at measuring three different indicators around “knowledge”, “use” and “trust”. It looks at social trust, trust in INEGI information, usefulness of information for the design of public policies, perception of veracity in INEGI information, perception regarding the independence of INEGI, transparency of information published by INEGI and evaluation of INEGI by other institutions. The Social Perceptions of INEGI study is carried out following the international recommendations of the OECD for statistical institutes on confidence in official statistics, which is carried out every four years by an agency other than INEGI, in order to give greater reliability to the data obtained. The results of the study have been used in the Institutional Strategic Program.

Case studies of countries’ experiences with customer satisfaction surveys and targeted surveys

124. Some of the Task Force members’ countries have undertaken targeted surveys and tested questions from the questionnaire, as per the 2018 Recommendations.
125. [Ireland has tested in depth the generic user satisfaction questionnaire](#) (as outlined by the framework). Mauritius conducted a user survey (2019) and produced their own shortened version of the questionnaire, focusing on quality aspects such as satisfaction with specific products and services up front, followed by trust and dissemination aspects. Armenia conducted a user satisfaction survey using the exact framework questionnaire and also include the ‘additional considerations outside these survey questions’ questions on page 70 of the framework that no other task force members have looked at explicitly. Also, with the support of Statistics Denmark, Armstat have developed another User Trust survey that is the light version of OECD Trust Survey and the last one was conducted in 2017. Hungary used an adapted version of the user satisfaction survey looking at use of statistics, access of statistics and user satisfaction. The UK no longer undertake a large user satisfaction survey (in the format as outlined in the framework), however, they outsource a survey on public views on trust in official statistics (based upon the OECD recommendations upon which the UNECE framework is also based) and further work has been undertaken by Economic Statistics Centre of Excellence UK who have also used the recommended questions from the framework.

4.4.3 Case studies of countries’ experiences with proposed monetary measures of value

126. Methodologies for monetizing value are far less explored than other fields proposed in the framework. Monetary measures can help to provide a simple and convincing story about the ‘value added’ of official statistics, enabling people to see in monetary terms what is made possible with official statistics. However, the complexity of such methods cannot be understated, since official statistics are a public good, financed by governments and (principally) ‘free at the point of delivery’, therefore, measuring value for money becomes complicated.
127. Many of the measures proposed in the framework are around the “production-based” approach and are not metrics on value but rather performance metrics that are used to provide information and set thresholds for budget allocation and operating costs and therefore limited in showing value.
128. In spite of the challenges, the appeal of monetary measures remains strong, and some countries have invested in developing and combining methodologies (Annex 3 contains full details of case studies). The few NSOs which have made attempts to produce such monetary measures of value

were, and still are, therefore, strongly encouraged to share their methodology and experiences with others to facilitate their further development.

129. Some of these include successfully undertaking cost-based approaches, including cost-benefit analyses and a that look to incorporate willingness to pay, revealed preferences and counterfactuals as well evaluated business cases, have been the most successful in showing value. New Zealand, Australia and the UK have presented these approaches to justify their censuses (See case studies in Annex 3). New Zealand has undertaken three such efforts and Australia provided a counterfactual against which to assess the value of the census (see Annex and case studies). Results showed quantifiable benefits to society and a return for the UK, Australia and New Zealand of 6:1 (though NZ most recent 2021 evaluation was 4:1). Though the most significant return has been seen in the New Zealand study showing a fiscal return of \$8 returned in the case of the Māori census. Armenia assessed the cost of conducting a census using traditional, combined and fully register-based methods and demonstrated a 55 per cent reduction in cost with the latter approach.
130. The UK is exploring the use of ‘conjoint analysis’ to explore its potential to value official statistics. This is a stated -preference approach, using a survey-based technique used in marketing studies which asks respondents to “consider jointly” their preference between products based on different attributes. A proof of concept was undertaken looking at the possibility to calculate a hypothetical ‘willingness to pay’. This work built on a previous study conducted for ONS (Deloitte 2016) which used indirect questioning of data users about price, but which had limited success. A second study has recently been undertaken.

4.5 Conclusion: towards a refined measurement framework?

131. Information received from the 18 countries that completed the measures review template made it abundantly clear that a simple ‘dashboard’ of indicators that limits itself to a production-based perspective would not amount to a meaningful assessment of value, nor would it be suitable across different national contexts with different strategic goals and varying levels of organizational maturity. Some measures make sense only in groups, accompanied by extensive contextual information. Viewed in isolation, for example, measures of citations, social media interactions, data downloads, or publication orders have little meaning and do not provide helpful guides to future action. However, combined suites of such measures may be much more useful, especially if specific targets or trajectories are formulated for such suites of measures.
132. The combined findings of the review of the measurement framework and the experiences of countries detailed in their case studies illustrate clearly both the challenges and the importance of making continued efforts to assess and improve the value of our products and services. It is important to note that the measures reviewed here were proposed prior to the Covid-19 pandemic which, as described in section 2.2, catalyzed a gear shift in how users’ perceptions are sought and responded to. In light of this fact, and in view of the proposals which follow in Chapter 5, the Task Force cautions against interpreting this work as ‘a refined measurement framework’. If the recommendations made in this document are followed, it is likely that more measures will be set aside from any usable framework while additional new measures may be developed by following the process defined in the next chapter.

Chapter 5. A Paradigm shift in measuring value: measuring for results

5.1 Introduction

133. At the kickoff event of the Task force in September 2019, an important question was asked by a presenter: are we trying to measure the value of official statistics with a view to *proving* that value to others, or to help us to *improve* value? The consensus was that both are important goals, even though the two may lead us towards different types of measures. Improvement depends upon measures that can quantify starting points, define targets, help us formulate actions and track progress (a fact with which we statisticians are of course familiar). Proving value, meanwhile, calls for measures that have communicative power, that are simple and unambiguous.
134. Whichever the focus for any particular NSO, the endeavour of measuring value falls into the broader realm of performance measurement. This chapter argues that the metrics and methodologies discussed in Chapter 4 and Annex 2 are tools, and that those tools should be used within the framework of a performance measurement approach that helps to define the task to which these tools are put to work. That is, the measures of value are not the goal—they are means to help us achieve a goal. A performance measurement framework enables us to ensure that the outputs being measured are those which contribute to the desired outcome; where an outcome is a strategic goal being pursued, while an output is a transactional activity contributing to delivery of the outcome.

5.2 Turning value inside out

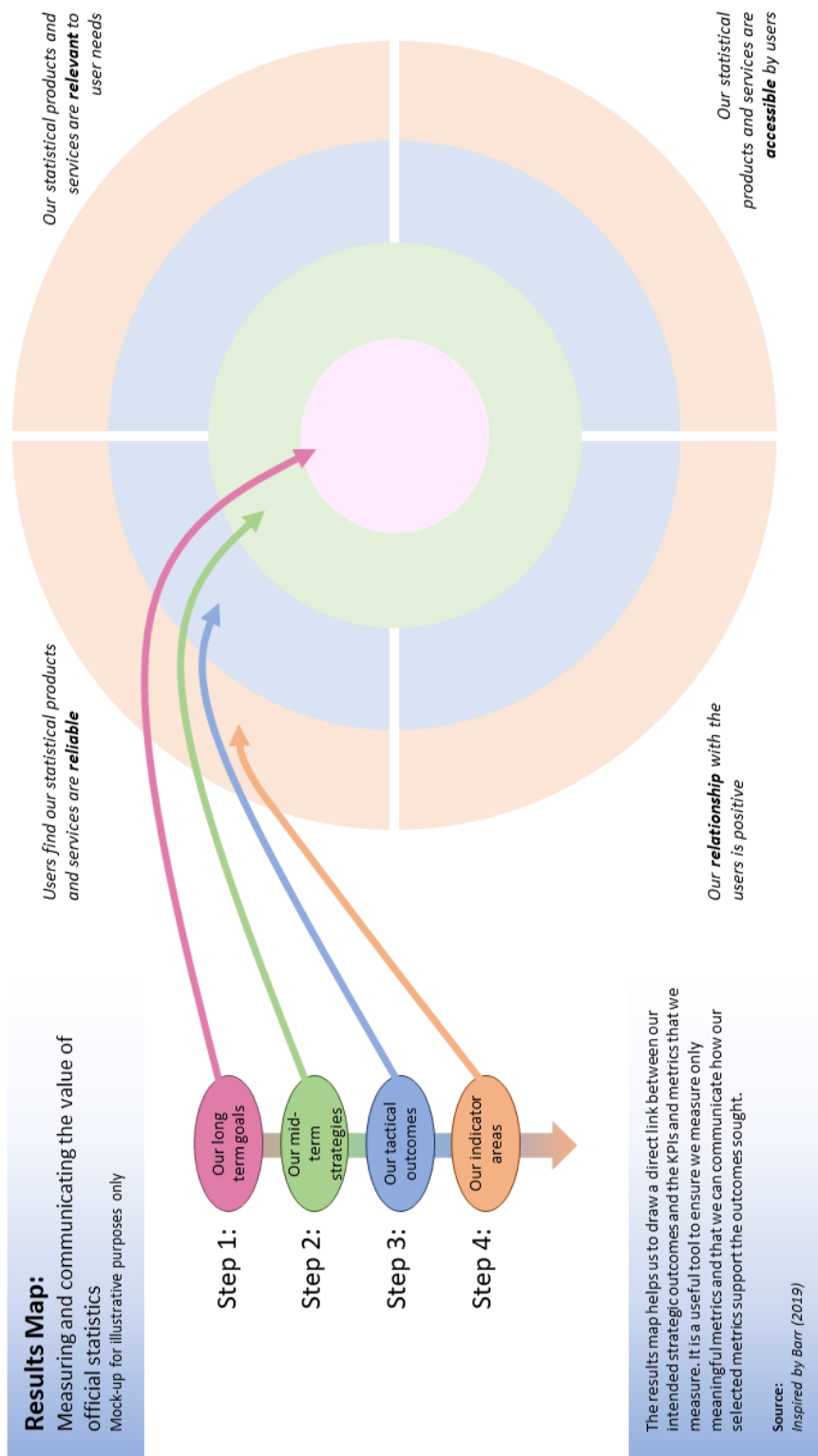
135. The measurement framework reviewed by the Task Force, as summarized in Chapter 4, was a preliminary set of ideas of ‘what components might there be to value, and how might we be able to measure these using currently available information?’ Such an approach follows a logic that is not best suited to devising an action-oriented performance measurement framework. Effectively measuring performance with a view to improving it requires that we must first define what constitutes good performance; we must have an agreed outcome as our goal. **Formulating the goal must be the very first step.** The previous framework had a *concept* (value), but not an explicit *goal* (improving value) at its centre.
136. In this respect, a measurement framework for measuring the value of official statistics has an important difference from many of the other measurement tasks that NSOs undertake. In other areas of quantification, the NSO provides statistical information to others who will use that information to inform their decisions. It is they, not we, who have the goal (be it improving economic growth, reducing child mortality, enhancing profits, or any other goal). In contrast, here we as NSOs are the ones who will use the measurements to shape decisions (actions we might take to improve our value). Hence, we are the ones who formulate the desired outcome, the long term goal that we are trying to achieve and which the measurement framework will serve.

137. Performance measurement is an established professional discipline, yet many sectors and organizations struggle to apply it effectively. This often stems from the very first step, where an outcome is poorly defined or articulated in a way that is open to interpretation. If our mission is “to produce great statistics” we have to ask ourselves “what does ‘great’ mean?”; “what do we mean by statistics? Just data, or analyses, tools and services too?”; “who is the judge of greatness?”; “when are we going to do this?”; “do we want to maintain a given level of greatness, or improve it over time?”. Defining this goal can (and should) in itself be a time-consuming task, involving many people across any organization and inviting the views of stakeholders to ensure that the goal is clearly articulated. Such an approach is central to some strategic management systems, such as organization change management or a ‘Balanced Scorecard’ strategic planning and management system—but these are not in widespread use in NSOs.
138. The idea of ‘turning value inside out’, then, is that instead of beginning with a set of possible measures of value and assessing them for suitability, NSOs should begin with a central goal based around creating and improving value, and work from there towards the formulation of actions we need to take to achieve that goal, and then to ways we can measure the effectiveness of those actions. The measures, in this view, come last instead of first.
139. This may sound obvious, but the experiences of NSOs suggest that it is not. In fact, to take this approach on an institutional level would, for many, genuinely be a paradigm shift. Given the other major shifts currently taking place in official statistics (as mentioned in Chapter 2), now is an ideal time to consider adopting this major change of direction.

5.3 What is a Results Map?

140. The Task Force explored one approach to performance measurement, drawing inspiration from a number of sources, including Barr (2014, 2017, 2019). This utilizes a visual model to map the pathways from goal to measurement (see Figure 4). Known as a Results Map, this visualization has, for the current purposes, been superimposed on a representation of the DARE model presented in Chapter 3, enabling us to see how these various pathways are distributed across the four dimensions of value. It should be borne in mind that, as explained in Chapter 3, the dimensions of the DARE model are in fact overlapping and their visualization as quadrants here is for simplicity only.
141. The diagram shows a hierarchy, leading from the centre outwards. In the centre (coloured pink) is the goal, the intended impact. This is akin to an overall mission of an organization, and provides its direction as set by the executives of the organization. Moving progressively outwards from green to blue to orange are layers leading gradually from strategies and outcomes towards actions, then measures, then targets.

Figure 4: Structure of a Results Map



142. The layers are not absolute or separate. Showing them as separate layers is a visual aid, but in practice there is a gradual shift from areas closer to the centre, where strategies are emphasized, to those closer to the outside, where the emphasis is on means of achieving those strategies. Phrasing the strategic results areas as statements helps to formulate them in a

results-oriented manner. For example: “customers can find the right data for their needs” or “people trust us”. It is important that these statements refer to *results*. They are not actions.

143. All too often organizations attempt to make a direct link between the inner and outer layers, bypassing these intermediate steps, often brainstorming key performance indicators (KPIs) from among things that are already commonly and easily measured, and thereby missing the logical link between the goal and the proposed indicators. The result is often that the indicators do not accurately measure things that would lead to progress towards the goal. Indeed, they may actually foster behaviours that deflect attention from or even detract from the goal. This can happen when an indicator becomes a target—something which happens frequently, whether or not it is intended as such. An example would be a measure of social media interaction as an indicator of public engagement with an NSO, presumed to be indicative of people finding our work valuable enough to want to engage with it in this way. If the measure is simply a numerical count of interactions, this could inadvertently foster over-emphasis on obtaining such interactions, without due regard to their nature or content and the impact that they are having on the brand. Social media managers, knowing that their performance will be judged on these numbers, may be incentivized to post more potentially popular material, perhaps at the expense of more ‘serious’ statistical material that would better inform users but be less susceptible to being retweeted. Another example: if error detection time is taken as a measure of accuracy and targets are built up around this, might this lead to statistical teams taking longer to check and release their figures, and in so doing reduce timeliness?

5.4 Experimenting with Results Maps for measuring the value of official statistics

144. With the benefit of hindsight after extensive debate and discussion in the Task Force, the group recognized that the process that had been followed in developing and reviewing the framework for measuring the value of official statistics was exactly as described above: indicators had been defined before goals, rather than vice-versa, and had been selected from a pool of mostly already-available information rather than on the basis of asking what information would be the most needed for measuring value.
145. The group therefore decided to experiment instead with this approach in which the usual direction is reversed. It is on the very edge of the planned scope of the Task Force (to test *and develop* the framework). Hence it has not yet been explored in depth, and there remains much work to do if the official statistics community wishes to pursue this approach.

5.4.1 Process followed by the Task Force

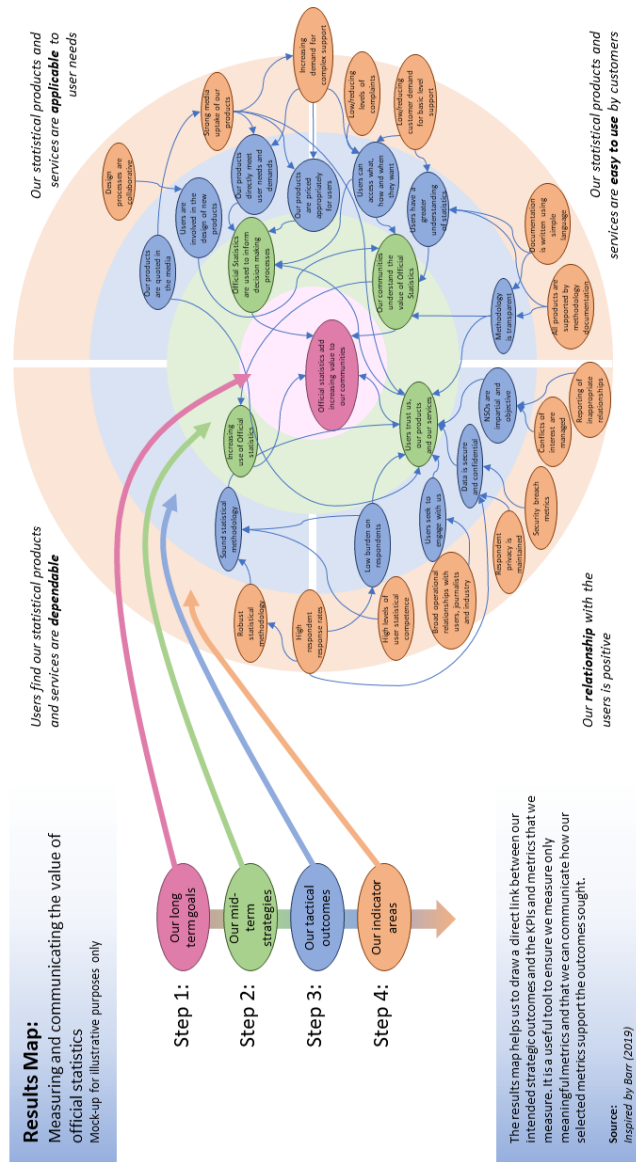
146. The group held two online workshop sessions, with work conducted by volunteers in-between. The aim was to explore whether using a Results Map could help to define a goal that encapsulated the idea of ‘the value of official statistics’ and then to map out some example pathways from this goal to possible quantifiable indicators of value. As a generic goal that attempted to reflect the commonalities of different NSOs’ missions, the group used ‘official statistics add value to our communities’.

147. The emphasis is on whether the process works, rather than on the specific outcome of these workshops. The group was interested in seeing the challenges and benefits of using the approach, as well as the similarities and differences between the value measures suggested through this method as compared with those in the original measurement framework.

5.4.2 Preliminary findings

148. The experiment was not intended to produce a finished Results Map that can be applied as a measurement framework. Indeed, the process of developing one must be country-specific; with the central goal being one that is defined at the highest level in an individual organization. The pathways that track outwards to measures then need to be defined by thorough consultation across large parts of the organization, and outside it, among stakeholders. Senior management and strategists would lead the development of the inner layers, while technical experts and communications specialists would play a major role in formulating the outer layers. This process would take a great deal of time and many iterations before a final Results Map would be agreed.
149. Having been produced in only a few hours by individuals from a range of organizations and positions within those organizations, the Task Force's mock-up is therefore not intended to be studied for the detail of its content, but rather an experiment to communicate some key findings:
- a) The exercise is intensely challenging, but enlightening. Participants noted that it is difficult to resist the ingrained tendency to start from the outside and work in.
 - b) If there is no clear pathway to trace a 'favourite' indicator back to the goal, perhaps it does not belong in a framework for measuring achievement of that goal. Participants also noted how challenging it is to accept this, as there is an urge to try and squeeze familiar measures into the figure.
 - c) A complete, well-developed Results Map could look very messy—pathways are not all unidirectional and one-to-one. There may be multiple different possible ways to measure any given intended outcome. Equally, one individual measure could potentially speak to multiple different outcomes, making it an especially useful measure.
 - d) Superimposing the DARE framework onto the Results Map is useful as it sheds light on sparse areas in need of further development in our battery of measures of assessing value, as well as revealing possible redundancies where multiple indicators map to the same outcomes.

Figure 5: Illustrative mock-up of a Results Map



150. To illustrate point b) above: the indicator ‘number of new visualization tools introduced’ was included in the 2018 framework under the quality subdimension. Many NSOs have invested significantly in improving their visualization capabilities, recognizing that easily understandable graphics contribute to better understanding of statistics, and that attractive formats aid uptake in a competitive market. But this does not mean that the use of new tools is necessarily an indication of value, nor that the number of such new tools is important. User engagement, user understanding and re-use of our products in the media are all tactical outcomes (that could feature in the blue layer of a Results Map) that might be served by high-quality visualizations— but none of these outcomes is best measured by the number of new visualization tools employed.
151. Similarly, the proposed indicator ‘innovation or quality awards received’ could be a reflection of sound statistical methodology, or of excellence in meeting the needs and demands of users— both tactical outcomes that might in the blue layer of a country’s Results Map—but it is not an

obvious measure of either of these things, and could be influenced by external factors such as the existence and criteria for such awards. It could also lead to misdirected attention in terms of a focus on winning awards.

5.5 Conclusions and next steps

152. The value of this Results Map is more than purely providing visual clarity. It serves to ensure that only meaningful metrics are included in any proposed measurement framework, justifying each by the fact that a path can be traced from the goal to the measure.
153. It also serves to communicate this fact to NSO staff and to external stakeholders who need to know that we are actively assessing our value with a view to improving it.
154. A Results Map could enable NSOs to monitor and refine their own measurement frameworks constantly without needing to start completely from scratch each time: the central long-term goal and mid-term strategies will evolve more slowly than the outer layers, so if an individual measure is found to no longer be workable it could be replaced by another simply by slotting in a new one that maps to the same tactical outcome, without having to re-evaluate the entire framework. Constant feedback could be built into the process to ensure the continued relevance and utility of any given tactical outcome or measure.
155. Given all of these benefits of the Results Map approach, the Task Force suggests that any NSO embarking on a plan to assess the value of their work should consider adopting or taking inspiration from this tool, adapting it where necessary to their own national circumstances. Each country of course has their own limitations in terms of resources and capacity, and there is no suggestion that any country should exceed their capacities in an attempt to assess the value of their work. Nevertheless, the hope is that adopting this approach could in fact enable NSOs to make more efficient use of resources—a hope that should be tested by countries willing to act as pathbreakers, and the results shared with the international community.
156. It may be valuable for the international community to collectively refine a generic version of a Results Map to help guide such national adaptation. To do this effectively, the participation of senior managers and strategists from a range of NSOs would be needed.

Chapter 6. Conclusions, recommendations and further work

6.1 Conclusions

157. The work of the Task Force did not conclude with recommending a new or revised set of measures to use to show value. Rather, in reviewing the framework of proposed value measures, the work brought to light a number of key points which bear stating explicitly here.

6.1.1 We cannot take value for granted

158. The point of departure of this work, indeed, the opening sentence of the terms of reference of the Task Force was “Official statisticians know that their products, underpinned by the Fundamental Principles of official statistics, are uniquely valuable...” The deliberations of this group over two years revealed that starting from a more self-critical perspective would be helpful—that is, starting out by asking ‘are we adding value?’ (and endeavouring to determine the nature and extent of that value), rather than assuming from the outset that we definitely are.

6.1.2 We must not conflate value with quality nor with adherence to our values

159. Both statistical quality and adherence to central values of the official statistics community are fundamentally important and help to define and drive us and shape what we do. But they are not synonymous with the value of our work to society, and we must not treat them as such. Measuring value entails measuring something other than the degree of conformity with existing quality dimensions or compliance with agreed community values.

6.1.3 Measuring value helps us both to prove and to improve that value

160. In reviewing the measurement framework it became clear that some countries have been producing certain measures routinely for some time—sometimes for many years. Several countries reported during the review exercise that the task itself helped them to reflect on why they were producing these measures and what purpose they were serving. In some cases, they were being produced simply because they were easy to produce. But the introspection entailed by this exercise revealed that some of the most commonly-produced measures were not actually considered to be highly important, while some less-commonly-produced ones were ranked as very important. The importance of a measure derives from what is or could be done with that measure. There is no point in producing a value measure if it does not inform any resulting action. Such actions can include internal efforts to improve value (e.g. a measure of user satisfaction may be used as a benchmark against which to assess efforts to increase user satisfaction); or to prove value (e.g. a positive measure of public trust in official statistics can be used as a marketing tool)

6.1.4 Distinguishing between production-based and consumer-based concepts of value leads to different approaches to assessing value

161. When we view the value of our products in terms defined by the NSO and its production processes, we end up with measures of value that are also defined in these terms, based around what we make and what we do. If we switch to a consumer-based conceptualization of value, the range of possible measures is different, broader and potentially more helpful in guiding us towards targeted endeavours to improve our value.

6.1.5 No single value measurement framework will be suitable for all countries

162. Different levels of organizational maturity and capabilities, different core missions and strategic goals, and different kinds of statistical products and services across countries mean that it would be impossible to produce a single framework of value measures and recommend all countries to produce them all. Cross-country comparability in value measures is therefore not a realistic prospect, nor one to which countries aspire since their attempts to measure value are principally intended for domestic purposes. Nevertheless, the review of countries' experiences did reveal commonalities, allowing for a common direction in broad scope if not in detail. Furthermore, the review revealed that value measures may lack meaning if treated in isolation: they are often more useful when looked at in groups of related measures (such as a series of measures of use of statistics). The measurement framework should not therefore be viewed as a menu or wish list from which to select individual indicators at will, but rather as a guide to dimensions and sub-dimensions on which countries may elect to focus.

6.2 Recommendations

163. In conjunction with the framework put forward for measuring the value of official statistics, the Task Force recommends that NSOs consider taking the following actions to support such measurement:

6.2.1 Start with a goal and build a Results Map

164. While we have proposed here a generic common goal that applies to all countries ('official statistics add increasing value to our communities'), NSOs may wish to adapt this goal to align with their own mission statements or strategies. Building outwards from their goal, NSOs should determine their own mid-term strategies and tactical outcomes, using the Results Map presented here as a model or as inspiration if they wish. While the present work proposes a measurement framework in general terms, the actual measures selected by any given country to measure the value of their official statistics should fit into their own Results Map, with every measure used being justified by a clear connection to the outcomes they are aiming for.

6.2.2 Find out from users what is valuable to them

165. Do not assume that we know what constitutes value in the eyes of our users. Seek evidence of the criteria determining value, the relative weights applied to them, and the ways that these differ among users groups and uses. We must acknowledge that asking people 'what makes statistics valuable to you?' is not the best way to do this. More subtle methods of determining people's preferences must be employed, since we are rarely conscious of and able to articulate our own values.

6.2.3 Keep asking

166. Conduct research on user perspectives (what criteria they use to determine value, as well as how well those criteria are met by our products) on a continuous basis. User needs are not static, as we have seen from section 2.2 (pandemic impacts on value), so the methods we use to assess how well those needs are met also cannot be static.

6.2.4 Ask everyone

167. We cannot assume that all components of value are equally important to all users. We must segment our user (and passive user and non-user) base and investigate their needs and preferences separately.

6.2.5 Consider consequences of a measure becoming a target

168. Pay careful attention to the actions that might be promoted if inclusion of a measure in a value measurement framework leads to efforts to affect that measure. Such unintended consequences could be positive but are often negative.

6.2.6 Review and revise measures and methods

169. Deciding upon a suite of measures by which to assess the value of our work is not a one-time undertaking. Users' needs will evolve; our products and services will evolve; the strategic goals, strategies and outcomes of our organizations will evolve. Hence the appropriate set of indicators for measuring value will also change. Available methods will also improve over time. An NSO may wish to refer back to their organization's Results Map to see which aspects have changed and should be updated, rather than undertaking the entire exercise from scratch each time. This kind of review should be integrated into the regular processes of the organization.

6.2.7 Make use of the value measures produced

170. Official statisticians love to repeat the adage that statistics are not useful unless they are used. Only then can their benefits be delivered. This applies just as much to measures of our own value as it does to classical statistical measures. We should not produce 'value metrics' merely for the sake of doing so, but should endeavour to act upon the findings, to improve what we

do. Relatedly, if we find that a value measure is not routinely used to drive change, we should question its importance and, if we find it not to be justified, discontinue its production.

6.3 Further work

6.3.1 Continue gathering national examples

171. The case studies included in the current version of this work were collected in tandem with the effort to review the measurement framework. Experiences from countries informed the decisions about retaining or setting aside measures from the framework. Some of the case studies, therefore, refer to experience with measures which, ultimately, have not been retained as potential measures of value. An important exercise will be to gather a wider selection of case studies from across the CES countries, with a focus on those which have been marked as having potential as indicators of value. Countries could make use of the existing agreed case study format to enable them to share experiences and lessons learned. Country case studies could be gathered to demonstrate efforts at measuring aspects of value not yet touched upon among the existing ones, such as attempts to evaluate the social welfare and well-being impacts of statistics; as well as examples of how information gathered from user engagement and impact assessment is actually turned into actionable information to change behaviour.

6.3.2 Transform this work, especially national examples, into a living online tool

172. As agreed early on in the current work, and reiterated by the CES Bureau when reviewing an interim progress report in October 2020, the usefulness of national case studies will be greatest if they can be housed in an online format, which can be added to and changed as experiences accumulate. Further work could therefore examine possibilities to develop this entire report into an online environment, with a particular focus on building a living repository of country experiences. This would need to be accompanied by a plan for maintaining such a repository on an ongoing basis.

6.3.3 Continue international collaboration to share and improve

173. The wealth of ideas, knowledge and experience among NSOs is immense. The present Task Force has found enormous benefit in regular exchange on everything from high-level conceptual questions (defining value) to detailed methodological discussions (techniques for measuring some of the indicators, techniques for ascertaining user preferences, etc.). The imperative to ask ourselves difficult questions about why we are doing what we do, whether we are doing it well, how we can improve and how we can demonstrate our worth is common to all countries.
174. It is essential that this does not end with the present report but is extended into the future through continued sharing of experience on the successes and failures of different value measures, methods for determining users' perceptions of value, and ways in which this information is used to inform actions designed to improve the value of official statistics.

175. Possible means of achieving this might include

- a) organizing international workshops for the presentation of country case studies
- b) organizing a workshop to develop a more refined generic form of the Results Map, leading to development of guidance for countries on employing this technique
- c) inviting countries to share experience with employing the approaches proposed in this document to develop their own national frameworks for assessing value.

6.3.4 Communicate the results of this work

176. Countries that have succeeded in producing measures of value may have individually communicated their results on a national level. It may also be helpful for the official statistics community as a whole to ensure that the collective effort is communicated widely. This would require interested NSOs to champion the cause as well as being done at the international level.

References

- Allin, P., & Hand, D. J. (2021). "Setting the Scene" *From GDP to Sustainable Wellbeing*, pp. 1-23. Available from: https://doi.org/10.1007/978-3-030-53085-3_1
- Allin, P., & Hand, D. J. (2021). "Statistics and Public Policy" *From GDP to Sustainable Wellbeing*, pp. 49-81. Available from: https://doi.org/10.1007/978-3-030-53085-3_3
- Allin, P., & Hand, D. J. (2021). "Using Statistics to Assess Progress" *From GDP to Sustainable Wellbeing*, pp. 25-48. Available from: https://doi.org/10.1007/978-3-030-53085-3_2
- Allin, P., & Hand, D. J. (2021). "Wider Audiences for New Measures of Progress" *From GDP to Sustainable Wellbeing*, pp. 83-109. Available from: https://doi.org/10.1007/978-3-030-53085-3_4
- Atkinson, A. (2005). "Atkinson review". Basingstoke: Palgrave Macmillan.
- Bakker, C. (2014). "Valuing the Census". *Statistics New Zealand*. Available at: www.stats.govt.nz
- Bakker, C. (2019). "Value of the Census for Māori" *Stats NZ Tatauranga Aotearoa*. Available from: www.stats.govt.nz.
- Barr, S. (2014). "Practical Performance Management: Using the PuMP Blueprint for Fast, Easy and Engaging KPIs" *Samford: The PuMp Press*.
- Barr, S. (2017). "Prove It! How to create a high-performance culture and measurable success", *New York: John Wiley & Sons*.
- Barr, S. (2019). "How to Measure what Matters: How to end your KPI and performance measurement struggles and reach your goals sooner with less effort" Available from: <https://staceybarr.com/downloads/WhitePaperMeasureWhatMatters.pdf>
- Chang, C., & Dibb, S. (2012). "Reviewing and Conceptualising customer-perceived value" *Marketing Review*, 12(3) pp. 253-274. Available from: <https://doi.org/10.1362/146934712X13420906885395>
- Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., Tennison, J. (2020). "The Value of Data" *Bennett Institute for Public Policy, Cambridge*.
- Deloitte. (2016). "Measuring the Value of Statistics" *A report to ONS*. Deloitte Analytics.
- Heeks, M., Reed, S., Tafisiri, M., Prince, S. (2018). "The economic and social costs of crime Second edition" *Research Report 99*.
- HM Treasury. (2011). "The Magenta Book Guidance for Evaluation" *UK government guidelines*.
- HM Treasury. (2019). "The Public Value Framework: with supplementary guidance" *UK government guidelines*.
- Kara, A., Lennard, J. (2020). "Valuing Economic Statistics: A Case Study". *ESCoE Occasional Paper No. 2*.
- King, J. E., & McLure, M. (2014). "History of the Concept of Value," *Economics Discussion / Working Papers 14-06*, The University of Western Australia, Department of Economics.

- Lateral Economics (2019). "Valuing the Australian Census". *Australian Bureau of Statistics*. Available at: <https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Value+of+the+Australian+Census>
- Loane, S. S., Webster, C. M., D'Alessandro, S. (2014). "Identifying Consumer Value Co-created through Social Support within Online Health Communities" *Journal of Macromarketing*. 2014;35(3):353-367. Available from: <https://doi.org/10.1177/0276146714538055>
- Office for National Statistics. (2011). "2011 Census benefits evaluation report". Available at: <https://www.ons.gov.uk/census/2011census/2011censusbenefits/2011censusbenefitsevaluationreport>
- Oliver, R., Alexander, B., Roe, S., Wlasny, M. (2019). "The economic and social costs of domestic abuse" *Research Report 107*.
- Rolfe, H., Runge, J., Srinivasan, S. (2020). "The Value of Economic Statistics: Baseline report". *ESCoE Technical Report No.05*.
- Simpson, I. (2016). "Public confidence in official statistics" *NatCen Social Research*.
- Statistics New Zealand (2019). "Stats NZ's Strategic Intentions" Available from: www.stats.govt.nz.
- Upadhyaya, S. (2017). "Cost of Conducting industrial surveys in developing countries" *61st World Statistics Congress, CCSA Special Topic Session 074*. Available from: www.unido.org/statistics.
- Williams, S. (2021). "Valuing official statistics with conjoint analysis: April 2021" *Office for National Statistics*. Available at: <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/valuingofficialstatisticswithconjointanalysisapril2021>
- Wray, L. R. (1999). "Theories of Values and the Monetary Theory of Production" *Explorations in Theory and Empirical Analysis / Working Paper No. 261*, The Levy Economics Institute of Bard College.
- Chang, Connie and Dibb, Sally (2012). "Reviewing and conceptualising customer-perceived value". *Marketing Review*, 12(3) pp. 253–274.
- Wray, Randall L. (1999). "Theories of Value and the Monetary Theory of Production", *The Jerome Levy Economics Institute*, Working Paper No. 261.
- King, J.E. and McLure, M. (2014). "History of the concept of Value" La Trobe University, Victoria and University of Western Australia, Discussion Paper 14.06.
- Stewart Loane Susan, Webster M. Cynthia and D'Alessandro Steven (2015). "Identifying Consumer Value Co-created through Social Support within Online Health Communities". *Journal of Macromarketing*, 35(5) 353-367.
- Prychitko L. David, Marxism (2002). "Marxism, The Concise Encyclopedia of Economics." Available at: <https://www.econlib.org/library/Enc/Marxism.html>

Annexes

Annex 1. Original (2018) measurement framework

Objective indicators included in original framework

Quality	Transparency
punctuality of statistical releases (share of punctual/late /cancelled releases)	timeliness of metadata (average “age” of metadata on the website)
share of error free statistical releases	share of statistics released with metadata
quick correction of errors (average delay of corrections in days)	number of blog posts by official statisticians
accuracy of statistics (average revisions)	number of users/journalists trained
timeliness of statistical releases (weeks from the reference period)	number of articles explaining statistics
number of new visualization tools introduced	number of open data solutions featuring statistics
innovation or quality awards received	number of partnership agreements
availability of quality descriptions (share of statistics released)	number of data cells in online statistical databases
Use	Relevance
number of website visits	number of citations by media
downloads of statistical data by domain	number of citations in research/policy work
visits to the digital library/website of publications	most cited statistics
number of followers in social media	most used/downloaded statistics
number of news feed subscribers	number of retweets
number of stats apps downloads	number of tailored services by user groups
number of chat contacts	number of new end-products/services
number of agreements to use microdata for research	working time used for development
number of agreements for chargeable services /sales of products/services	number of papers/presentations/inputs that contribute to international statistical work
sales/number of publications requested	number of international study visits hosted
number of responses to international requests	number of memberships in international expert groups

Subjective indicators included in original framework

The section on subjective indicators proposed conducting a regular user survey (a generic formulation was offered in the publication as an example). The sub-components and areas in which it suggested to produce indicators were:

- Satisfaction with products and services.
- whether the user found an answer to his/her question(s),
- whether or not the relevant information was easy to find (e.g. accessibility of statistical databases)
- to what extent the information needs were met.
- what the user considers the most important characteristics of official statistics or statistics more generally (e.g. timeliness, accuracy, trustworthiness, (inter)national comparability, etc.), and
- how he/she rates official statistics on each of these characteristics.
- preferences for type of access (online, phone, in person), device type(s) used and preferred media
- User support:
 - general perception of user on whether or not we are doing well
 - what the user thinks we could do better in serving users.
- Design, communication and metadata:
 - design of the official statistics website in general, and the statistical warehouse in particular
 - how easy/difficult it is to navigate and find the relevant information,
 - how satisfied the user is about the visualization of official statistics (videos, infographics, maps, graphs, indicator sets), etc.
- Relevance, responsiveness and innovation:
 - How effectively does the statistical office inform the public debate on current issues affecting our country
 - to what extent do you think that we are innovative in the way we work (e.g. using new technologies, methods and data sources)
 - how important are official statistics in helping to understand societal developments.
- Awareness of brand and message:
 - trust in official statistics
 - perceived lack of political interference

- overall satisfaction with the statistical office
- understanding of its remit
- Specific products and services:
 - Have you heard of a particular statistical product
 - how satisfied are you with the quality of the product or service
 - have you used public use files or anonymized micro data and for which domains would you need them mostly
 - do you think there are benefits for you or your organization from increased sharing of anonymized data, etc.

Categories of monetary indicators included in the original framework

- cost-based approaches
- market (equivalent) pricing
- stated preference methods
- revealed preference methods
- impact assessments.

Measuring the Value of Official Statistics: testing and developing a measurement framework

Any attempt to quantify how valuable official statistics are must begin by asking what is meant by value, and whose perceived value matters.

This report reviews potential ways of measuring value, supported by case studies demonstrating their use and suitability (or non-suitability) for assessing the value of official statistics, available in a separate annex to the report. The overarching conclusions are:

- Consumer-based approaches to value result in a different set of potential indicators than those arising from a more traditional, 'production-based' approach to value.
- While production-based indicators can be very useful for operational and management purposes (including areas like quality and budgeting etc), they do not necessarily reflect the value of our outputs in the sense understood here.

Moving towards a consumer-centred approach to measuring value - and, by extension, to creating, maintaining and improving that value, which is the ultimate goal - necessitates a fundamental shift in direction.

This document offers a suggested method to navigate this change, which entails using a 'Results Map' to define a clear path to achieving the central goals of official statistics; working outwards from core strategic goals, to measurable outcomes, to quantitative indicators of value.

The report is the work of an international task force of experts from national and international statistical organizations, and was endorsed by the Conference of European Statisticians in 2022.

Information Service
United Nations Economic Commission for Europe

Palais des Nations
CH - 1211 Geneva 10, Switzerland
Telephone: +41(0)22 917 12 34
E-mail: unece_info@un.org
Website: <http://www.unece.org>