Office for Statistics Regulation

Office for Statistics Regulation Research Programme

How statistics can serve the public good: a think piece

February 2024

Contents

1 Executive summary

At the Office for Statistics Regulation (OSR) we believe that statistics should serve the public good and, to support this vision, we have invested in research to deepen our understanding of this topic. This think piece sets out OSR's current views on how official statistics can serve the public good, alongside the thought processes that led to these conclusions. We are publishing it in the knowledge that there is refinement needed: this think piece acts as an invitation for readers to join the discussion and help develop our position further.

At present, we propose that:

> Official statistics serve the public good as public assets that provide insight, which allows them to be used widely for informing understanding and shaping action.

There are several elements of this proposal that we would like to highlight:

1. We remain neutral to conceptions of what outcomes may be in the public good, but assert that statistics serve it through they insight they convey, which a range of potential users can harness to produce outcomes that benefit society.
2. Being a 'public asset' comes from the public funding and public ownership of government statistics and data. Treating statistics as a public asset means statistics must be accessible to all, produced in an efficient manner, and their production and regulation should be accountable to the public.
3. 'Providing insight' represents the role statistics play in what we know, which serves the public good by informing society and by facilitating transparency. It is about statistics that meet society's need for information, which can be through validating existing evidence or providing new information. Implicit in providing insight is a sufficient level of quality, neutrality and context in the statistics. Serving the public good in this way requires producers to understand user and potential user needs for information, which is supported by wide engagement.
4. In addition to providing insight, statistics being 'used widely' allows them to unlock new ways through which they can serve the public good: by 'informing understanding and shaping action'. There are a wide range of potential users and uses that may serve the public good, statistics must be useable to support this use. Uses in the public good benefit the public, whereas misuse of statistics undermines the public good that statistics may serve.

We invite you to digest the perspective presented in this think piece, and to share your thoughts on what it may mean for how official statistics serve the public good by e-mailing us at regulation@statistics.gov.uk. Through inviting your participation, we intend to develop a shared understanding that can guide both our regulation and the production of statistics in our statistical system.

2 Introduction

2.1 Our interest in statistics serving the public good

As the regulatory arm of the UK Statistics Authority, the Office for Statistics Regulation's (OSR's) statutory objective is 'promoting and safeguarding the production and publication of official statistics that serve the public good'. As such, our vision is that statistics serve the public good, a goal shared by the wider UK Statistics Authority and highlighted in its strategy Statistics for the Public Good.

In OSR we make judgements every day on whether statistics are collected, produced, published and used in a way that serves the public good. In doing so, we strive not only to serve the public good ourselves, but to support the wider statistical system to do so as well. But when we say statistics are serving the public good, what do we really mean? **How do official statistics serve the public good?**

Given statistics that serve the public good is a widely shared goal, answering this question will improve the ability of the statistical system to make practical decisions and as such collectively deliver on our vision. In OSR we have a Research Programme that seeks to bring us closer to answers, and since its inception in 2019 we have undertaken a range of projects, in addition to looking across the statistical system and beyond to identify other evidence.

In January 2023, we undertook an affinity diagramming exercise, which is where a team works together to collate a body of evidence and identify themes within it. This focused on official statistics serving the public good and included findings from our Research Programme alongside reflections from our regulators and evidence from outside sources.

2.2 An invitation to discussion

The evidence we collated through our affinity diagramming shaped our thinking around how statistics serve the public good. Since that original exercise, we have sought further evidence, discussing this topic with individuals and groups, and inviting feedback at events to progress our thinking. However, is important to us that this is not a closed discussion – everyone should have the opportunity for a seat at the table. Therefore, we have chosen to open the floor for anyone with an interest to share their opinions by publishing our initial thoughts and unresolved questions in this think piece. As a think piece, this document is exploratory by nature and aims to promote discussion rather than definitively answer the question we are exploring. It articulates our thought process so far about how statistics can serve the public good and acts as an invitation for readers to join the discussion.

The perspective in this think piece is underpinned by a range of sources such as research evidence, regulatory experiences and philosophical discussions. We hope that the feedback we receive is just as varied, which is why we are publishing this openly and seeking views from anyone with interest or expertise in the topic of statistics serving the public good.

We are eager to share our thinking with you and we invite you to participation in this work by sending your views to regulation@statistics.gov.uk.Your contribution will be greatly appreciated.

2.3 Structure of this think piece

This think piece sets out our views so far about how official statistics may serve the public good. There are two main sections:

- How we currently see statistics serving the public good; and
- Why we arrived at this conclusion.

For the section on how we currently see official statistics serving the public good, we explain our overall perspective – how we have tackled the question, and an initial answer.

For the section on why we arrived at this conclusion, we go into greater detail on our thought process and evidence-base for different aspects of the answer we are currently working with.

In some areas, the evidence we have encountered and discussions we have held to date point in a single direction. In these cases, we use this think piece to set out our stance. In other areas, however, the picture is less clear. To support the conversation around areas where there is less clarity, we have included ‘discussion boxes’ setting set out questions we are still debating. Although we are open to receiving views on any aspect of this think piece, these ‘discussion boxes’ highlight areas where new contributions would be most valued.

3 How do we currently think official statistics can serve the public good?

This think piece is about creating a shared understanding and collaboratively answering the question **how can official statistics serve the public good?** This section of our think piece describes the evolution of our overall thoughts on answering the question, and proposes an initial answer.

3.1 How to answer this question

In answering the question 'how can official statistics serve the public good?', we focus on what allows official statistics to serve the public good, rather than defining 'the public good'. We are aiming to collate information and add depth to known definitions rather than replacing existing understandings. Even the decision to address this specific question and from this perspective took time to develop.

The OSR research programme began our exploration Into statistics that serve the public good with a literature review on the topic of the public good of statistics. This literature review notes that the economic definition of a public good is 'a resource which is nonexcludable, meaning everyone can use it, and non-rivalrous, meaning one person using it does not take it away from anyone else'. Statistics produced by government could widely be considered public goods in that they meet this criteria, however some explorations of this in greater detail (such as Asle Rolland's paper on the concept and commodity of official statistics or section seven of Steve MacFeely's paper in search of the data revolution) show there is more nuance that could be debated here. A further (unpublished) consideration in this respect came from discussions with independent researcher Ken Roy, who challenged how closely official statistics align with the colloquial definition of public goods as well (produced by government for citizens). Here, he highlighted that as well as being produced for the public, official statistics are also produced for, and can sometimes prioritise, government use.

If we followed the public goods route, then perhaps 'statistics serving the public good' might mean statistics are produced by government, and can be accessed and utilised by everyone. However, we see a distinction between *the* public good and these public *goods,* which makes such a definition overly simplistic. The distinction is described in a book sharing perspectives from the Census of India, with authors Abhishek Jain and Varinder Kaur stating 'Census is both a public good and for public good. Census collects data, but this data is not an end in itself, but a means to promote human welfare'. This statement about censuses can be extrapolated out to official statistics more widely – official statistics themselves may be public goods, but they must benefit society if we are to say they serve the public good.

Noting the distinction between public goods and the public good means we can move away from existing definitions of public goods and focus on understanding the concept at hand. To define the public good, Neil Walker described in a paper on European public good and public goods that one requires 'a prior sense of who that

particular public is; which public, once identified, provides the reference point for what is good for it'. This statement might suggest that to understand statistics and the public good we ought to be identifying who the public is and what good is for them. Doing so would follow the approach of the National Data Guardian's explanation of what public benefit means for the context of public benefit evaluations, which set out who the public is and what benefits are in this context. In fact, we did explore this when we spoke to members of the public as part of a series of public dialogues. From this we heard that 'the public' could be a current or future public, and could include either everyone or a subset of the population. We also heard that the 'good' could be wide ranging, including tangible and less tangible, but overall ought to be about meeting real world needs.

We might have explored this further as our primary research question, because knowing who the public is and what the good is would allow us to define 'the public good' in the context of statistics. However, in OSR our remit is not to dictate what is 'good' or 'bad' for the public. Our role is to support public confidence in, and appropriate use of, statistics rather than to direct the ends to which they are used. As such, we chose to remain neutral with respect to conceptions of what is in the public good and instead focus on what the role of statistics may be in achieving it.

When exploring the public good of statistics from this angle, the most obvious piece of evidence comes from legislation. The Statistics and Registration Service Act (2007) states that 'serving the public good includes in particular informing the public about social and economic matters, and assisting in the development and evaluation of public policy'. This statement begins to explain what it means for statistics to serve the public good. However, it can be difficult to know what it means in practice: the high-level legislative definition still leaves statistics producers and OSR as a regulator with little information on how they can make decisions that support statistics to serve the public good.

This is the space where the present think piece sits – focussing on how official statistics can serve the public good, rather than looking at statistics as 'public goods' or stating what is 'good' for 'the public', and aiming to get to a point where understanding allows for practical decision making. We already provide practical guidance in the form of our Code of Practice for Statistics (the Code), which sets out a framework made up of pillars, principles and practices and states 'statistics will serve the public good if producers follow the principles and practices set out in the Code'. However, knowing that the public good is being served by following the Code is not the same as understanding how to serve it, in the same way that driving a car is not the same as understanding the underlying mechanics. Although most of the time this deeper understanding may not be needed, in new or complex situations it can be beneficial.

To provide the strongest possible support for our statistical system, we therefore have chosen to focus on the question **how can official statistics serve the public good?** While our question of interest is specifically about official statistics, much of the content within this think piece could be applied more widely as well; given our remit is tied to official statistics, however, they remain our primary focus.

3.2 An initial answer

The evidence we have identified and discussions we have held to date have informed our thinking and supported the development of an initial understanding on how official statistics can serve the public good. To communicate our thinking in a succinct manner, we have captured our initial answer in a statement:

> Official statistics serve the public good as public assets that provide insight, which allows them to be used widely for informing understanding and shaping action.

This statement compliments the Statistics and Registration Service Act (2007), which notes 'serving the public good includes in particular informing the public about social and economic matters, and assisting in the development and evaluation of public policy'. In this legal definition, 'informing the public' relates to being a public asset that provides insight, 'development of policy' sits under shaping action, and 'evaluation of policy' is one example of statistics informing understanding.

The most substantial difference between our current understanding (reflected in our initial statement) and legislation is that we propose that actions in service of the public good go beyond only the development and evaluation of public policy. For example, during the COVID-19 pandemic businesses and individuals were using statistics to make a variety of decisions that could be seen to benefit society. Despite this difference, we assert that proposing a broader definition in this way does not contradict legislation, as the Act describes the public good as *including* the specific public policy examples, rather than being *limited to* only them. In addition, statistics use beyond public policy would still fit with the legal definition, as for use of statistics beyond policy to happen the public must be informed, which the legal definition includes as serving the public good.

In the next section of this think piece, we will set out how we arrived at our current understanding, exploring why we include certain terms in our statement and what we see them as meaning. Before that, however, is the first discussion box of this think piece. These discussion boxes include specific questions we would like to hear your views on.

Is our high-level statement appropriate?
Before you read the full explanation of the statement in the rest of this think piece, to what degree does our current proposed definition make sense to you? Upon first reading, how appropriate do you see it as an answer to how official statistics serve the public good? Are there any words or phrases which you think are confusing or problematic?

4 Why did we come to these conclusions?

This think piece supports our goal of creating a shared understanding and collaboratively answering the question **how do official statistics serve the public good?** We are sharing our current thinking with the view that this will continue evolving, and we welcome evidence both in support and challenge of what we propose so that we can continue to develop our understanding.

This section is structured around our initial answer to the question of how do official statistics serve the public good: '*Official statistics serve the public good as public assets that provide insight, which allows them to be used widely for informing understanding and shaping action'*. Here we set out why we are including specific phrases in the statement, and what we propose they mean in this context.

Within our explanation of official statistics serving the public good, we use the term 'public' in multiple different ways. Three examples of this are:

- Public funding: This relates to a sector of the economy, where 'public' is about state-raised and distributed funding. It is essentially an accounting concept;
- Public availability: This is about the non-exclusive access to official statistics, where 'public' represents access for all. It is essentially an economic concept, and most closely related to the economic idea of "public goods"; and
- Public accountability: This ties into a democratic notion where 'the public' represents individual citizens with no formal representative role. It is a democratic concept.

While we recognise that using the same term to represent multiple concepts may be confusing, in parts it is unavoidable. As such, where not specified throughout, when we use the term 'public' we are referring to the democratic concept.

4.1 Public asset

4.1.1 Why we propose official statistics should be seen as public assets

As government departments are publicly funded ('public' in the accounting sense) we propose that the statistics government produces are a public asset. The 'public' we refer to in public asset covers multiple versions of the concept, for example official statistics were described by the United Nations Economic Commission for Europe (UNECE) work on valuing official statistics as a 'public good resourced by public money', covering both the economic and accounting concepts. The view that official statistics are public assets is further supported by UK government guidance on managing public money, which describes categories of public sector assets ('public sector' relating to the accounting concept of 'public') including intangible assets such as 'data and information'. Given 'data and information' are assets, we argue that statistics based on these data are too.

In addition to the accounting and economic senses of 'public', the phrase 'public asset' can also tie to the democratic meaning in that official statistics are an asset owned by members of the public. We anticipate that the public feels a sense of ownership over government statistics, especially when they are created using their

data. When we spoke to members of the public through a public dialogue project, participants expressed feeling ownership of their data, including feeling responsible for the outcomes of its use, with one participant saying 'irrespective of whether my name is attached to it, I hate to think that my data is contributing to something that harms someone'. As such, when an individual's survey or administrative data is processed into statistics, we propose that a sense of ownership will remain, supporting the view that official statistics are the public's asset.

Are official statistics always public assets?
We acknowledge that not all statistics come from data about individual members of the public, for example business or environmental statistics. Therefore, we are interested in evidence or views about whether the argument that official statistics are a public asset is weakened in such situations, or whether other evidence is strong enough to justify our perspective that official statistics are public assets?

4.1.2 What it means for statistics to be a public asset

In our understanding and statement, statistics being public assets includes:

- Ensuring statistics are accessible to the public (including many types of the public);
- Producing statistics in an efficient manner; and
- Being accountable to the public with this resource.

If we state that official statistics are a public asset, then they should be something the public can access. Our position on this began with 'official statistics should be available'. This stance is in line with the UN fundamental principles of official statistics, which says that official statistics should be made available to honour citizens' entitlement to public information, and the Statistics and Registration Service Act (2007) which talks about serving the public good by informing the public. We anticipated that public availability would support transparency of government, and allow for wider uses of statistics compared to if they were accessible solely to government.

However, it soon became clear that stating official statistics should be available was not as strong a position as we had imagined. We heard concerns that some outputs which meet the criteria to be labelled official statistics in our official statistics policy may be branded as 'analysis' or 'monitoring information' rather than official statistics. Some expressed that this was a purposeful attempt to take such analysis outside of the remit of the Code and other official statistics governance. As such, stating that official statistics should be available could in some situations be meaningless, as if an actor did not want the statistics to be published they might simply re-brand them as a different type of output. In acknowledging this challenge, we moved away from saying 'official statistics should be available' and towards the view that if official statistics are to serve the public good, then all government statistics must be made available, regardless of whether they are referred to as 'official statistics' or not. By stating all government statistics should be available, we aimed to support public confidence that regardless of whether or not statistics shone a favourable light on those in power they would still be published.

When we first proposed that government statistics should be available, we received challenge on how it would work in practice in terms of national security and personal privacy. Such a challenge is described in the Exploring Data as and in Service of the Public Good report, which explains how for some datasets full open access may be at odds with public benefit. As the line between government statistics and other data or analysis can be blurred, our stance could be interpreted as stating that we want all data and analysis by government to be published regardless of privacy or security. Given the importance of privacy and security, we briefly moved to the position that 'data quoted publicly, for example in parliament or the media, should be made available to all in a transparent way', which comes from our regulatory guidance for the transparent release and use of statistics and data.

However, the intelligent transparency perspective may be taken to mean statistics should *only* be made public when they are quoted publicly, which could limit access rather than enhance it. Because of this, we reverted back to our earlier stance and sought more information to develop our position, moving us to conclude that government statistics should be open by default. Our proposal is informed by the paper Open data for official statistics: History, principles, and implementation, which explains that open by default means publishing all statistics *unless* there are specific reasons for limiting access. The paper describes reasons such as protection of security, privacy, confidentiality, and intellectual property. This does not mean openly sharing all data behind statistics – privacy is still paramount as we heard in our public dialogue project, but it does mean the aggregate statistics themselves should be available. We therefore propose that to serve the public good, government statistics should be publicly available (open) by default, and there should be a presumption that they will be shared unless there is a justifiable reason not to. This is in contrast to only publishing government statistics when they have been formally labelled as official statistics.

How open to the public should government statistics be?
We propose government statistics should be open by default, but have we got the balance right? Is it fair to look at government statistics rather than official statistics? Should we be discussing government data as a whole? What conditions for restricting publication are appropriate? Should there be additional conditions, such as proportionality, or does this risk weakening the message? And how might this impact our statistical system, especially in terms of adding strain during a time of resource challenges?

Once we settled on the stance of open by default, we recognised that our focus here had been on *availability* of statistics rather than *accessibility,* and that availability without accessibility would not be sufficient. Accessibility is more than statistics being published - they would also need to be in the right format (such as open data formats rather than files which can only be opened with proprietary software) and presented with the right information (including clarity about licencing, and appropriate metadata) or else they risk still only being truly accessible to a subset of the public. In OSR, we recognise that 'the public' may make up everyone outside of government, but it is not a homogenous group. The public comprises individuals, who may be organised as communities or organisations (including businesses,

charities etc), or may stand on their own as individual citizens; any of these groups could be seen as publics in themselves, and will have their own accessibility needs.

One part of accessibility is statistical literacy. The relationship between statistical literacy and accessibility is explored by Steve MacFeely in the context of data and statistics as public goods, where he notes that using data properly 'requires significant expertise and contextual knowledge', and claiming statistics as public goods that everyone can use 'presupposes a certain level of statistical literacy'. We have seen through a literature review that there is no uniform definition or level of statistical literacy, therefore in OSR our position on statistical literacy is that 'rather than address deficits in skills or abilities, we recommend that producers of statistics focus on how best to publish and communicate statistics that can be understood by audiences with varying skill levels and abilities'. Therefore, considering access of statistics means communicating statistics in ways that are understandable to both technical and non-technical audiences.

Accessibility goes beyond accounting for varied technical skill level though – even more fundamental is considering access needs. According to gov.uk guidance on accessibility requirements for the public sector, being accessible means ensuring something can be used by as many people as possible, including those with 'impaired vision, motor difficulties, cognitive impairments or learning disabilities, deafness or impaired hearing'. The guidance states that accessibility includes 'making your content and design clear and simple enough so that most people can use it without needing to adapt it, while supporting those who do need to adapt things'. Without considering how official statistics are communicated in relation to visual, auditory, motor and/or cognitive variations across the public, disabled users may be disproportionately burdened or even deprived of access, meaning statistics cannot be seen as available to all.

The next aspect of public assets we moved to was statistics being efficient. This was seen in UK government guidance on managing public money and also in the Statistics and Registration Service Act (2007), which states that the Authority must 'exercise its functions efficiently and cost-effectively'. In addition to these formal references, there is also a logical argument behind stating that efficiency serves the public good. In our public dialogue project, a common view that was expressed was that the public good was about the most possible good within a given context. If statistics are produced efficiently that will mean that limited public funds (in the accounting sense) can stretch further, therefore allowing for more 'good' to be produced.

In practice, considering efficiency may include creating Reproducible Analytical Pipelines, or re-using and linking existing datasets to gain new insights from existing data (linking datasets was described as in the public good in researcher applications to access public data). Efficiency is also seen in principle V5 of the Code (efficiency and proportionality), which suggests data sharing and linkage, alongside other practices such as using recognised standards, classifications, definitions and methods. If we accept that being a public asset is important for statistics to serve the

public good, then these approaches to improving efficiency ought to maximise public good.

After considering efficiency, the final part of being a public asset that we moved to was accountability. This was a logical conclusion from the proposal that the public funded and felt ownership over statistics – if statistics are the public's property, then producers and OSR as a regulator ought to be publicly accountable in what happens with the statistics. Accountability can be considered throughout the full process of statistical production and dissemination, for example by considering ethics and quality at each stage. We do not suggest accountability to the public in the same way as internal governance structures, but we do see that the public should know what is happening with the statistics they fund. This could be achieved by producers through explaining decision on statistical production and publishing future development plans, and by us as a regulator clearly and publicly explaining our judgements.

4.2 Provide insight

4.2.1 Why we include providing insight

When we first considered what it meant for statistics to serve the public good, we took inspiration from participants in our public dialogue project who associated the public good with positive impact. We took this to mean that there had to be an impact to serve the public good, which we proposed came out of statistics being used.

However, this idea that statistics must be used if they are to serve the public good sat uncomfortably. From discussions both within OSR and in the international statistical community we began to hear more and more that those working in statistics see an intrinsic value in the insights statistics provide, even when they are not used.

Because of this, we revisited our original interpretation of evidence to see how the views of the statistical community could be reconciled with perspectives expressed to us by members of the public. In doing so, we saw the conclusion we had drawn was biased by the purpose of the research – the project asked about public good *uses* of data for research and statistics, thereby presupposing that *use* had to happen. In the same project we also heard reference 'knowledge' serving the public good, which implied that participants may have accepted insight itself as serving the public good.

Further support of the idea that people may see value without use came from a brief review of OSR's casework issues log for the current financial year. This log reflected multiple instances where we had been contacted about the representation of statistics and how accurate or complete they were perceived to be, even where no specific use case was mentioned. We interpreted this to mean that people care about statistics reflecting society, even when there are no plans for them to be used beyond this. As such, we began to solidify our position that the knowledge official statistics convey in itself serves the public good.

Other evidence supporting this perspective comes from legislation, where 'informing the public' is a way for statistics to serve the public good, and the first UN fundamental principle of official statistics which states (in part) that 'official statistics

provide an indispensable element in the information system of a democratic society'. These two quotations reinforced our view that statistics can serve the public good through the insights they provide and transparency they promote without having specific uses and broader impacts.

Is providing insight enough?
We have wrestled with both sides of the argument on whether statistics do indeed serve the public good by providing insight or whether the public good comes from their use, and would be interested in hearing your views. How strong or weak do you think our position is on this area? Do you have any examples where you see statistics serving the public good when they are not used? Is there a risk in saying that statistics don't always need to be used as it may disincentivise actions being taken to promote their use?

4.2.2 What we mean by providing insight

In our understanding and statement, providing insight includes:

- Providing value by meeting society's need for information;
- Striving for neutrality;
- Both validating existing evidence and providing new knowledge;
- Supporting transparency and democracy; and
- Representing the experience of people across society.

As we steered towards the position that statistics can serve the public good even when they are not used, we considered settling on the view that official statistics are intrinsically valuable. However, this is not the stance we currently hold. Our position was tested by the UNECE work on valuing official statistics, which challenged the international statistical community to consider what their users valued in statistics, and ask 'are we adding value?... rather than assuming from the outset that we definitely are'.

Rather than reverting to our previous view that use is necessary to serve the public good, we explored instances where official statistics can provide value when they are not used. We had evidence that some people appeared to value the knowledge statistics provided (from reviewing our casework evidence log), which suggested in some circumstances knowledge itself can be valuable to users. We also saw this in our 2022 update of lessons learned for health and social care statistics from the COVID-19 pandemic, which states that statistics can serve the public good through 'providing the information [people] need about the issues they care about'.

From this, we concluded that statistics on a wide range of topics can serve the public good without being used, provided they meet society's need for information in line with the pillar of Value in the Code. In addition to being relevant, accuracy clearly influences how well or poorly information needs are met, and other dimensions of quality (such as coherence, clarity and timeliness) will also impact the insight which statistics provide to those who consume them. As such, to serve the public good statistics must reflect relevant topics and be at an appropriate degree of quality.

Related to maintaining an appropriate quality profile is the importance of official statistics remaining neutral if the insight they convey is to be appropriate and trusted. Neutrality is complex, as every decision from sampling to question design will introduce some bias, and the government-funding of official statistics may influence which topics are seen as priorities for production. There is a tension here that remains unresolved, as we assert that neutrality is both congruent with serving the public good, and unrealistic in its purest form. We anticipate that steps can be taken to protect and promote neutrality, such as considering and explaining biases in outputs, and maintaining professional independence.

How should we manage neutrality?
What is the best way to promote neutrality while acknowledging idealism in such a goal? Is it fair to state that statistics should be neutral if they are to serve the public good, or would it be more appropriate to say they should be politically neutral while acknowledging other biases? Where should the line be drawn in terms of acceptable bias?

It may be obvious that providing insight includes statistics in telling society new information, however in our public dialogue project we also heard from participants that public good uses of data included validating existing evidence. As such, we have expanded our thinking and propose that providing insight can include confirming or challenging existing knowledge, as well as revealing new information.

In some instances, it may be completely irrelevant whether the insights statistics provide is new or not; we currently posit that the public good may sometimes be served simply because insights are available, no matter what they tell us. This perspective comes from discussions about the role of official statistics in democracy with those who work in statistics. They expressed to us the importance of the transparency official statistics provide, even when they are not directly used to understand or challenge government decisions. This speaks to the legitimacy of government, as regardless if action is taken, a ruler who can be challenged has greater legitimacy in a democracy than one who cannot.

Statistics would still need to be on a topic that society cares about for them to serve the public good in this manner (investing resources in a public asset where there is no conceivable interest cannot be an appropriate use of limited funds). However, from our discussions we anticipate that knowing statistics on topics of great importance are available to be used if needed may provide reassurance to society as a whole, even when individuals never personally choose to use them.

Statistics may also serve the public good for a different group as well - those whose data is included in them, be they respondents to surveys or data subjects in administrative data. We assert that this group may place importance on their experiences being captured and reflected, both because they wish their experiences to be accounted for in decisions made by others, and also because they may place value on being seen. We interpret requests for numerous ethnic groups to be added as tick-boxes to the England and Wales 2021 Census as support for this assertion.

While respondents to this census were able to write-in any ethnicity that was not listed in tick-boxes, the Jewish Chronicle raised that without a Jewish tick-box 'citizens will not feel fully counted', implying that for some there is value in both being and feeling visible in statistics. This highlights the social role statistics can play for those whose data feeds into them, and raises the importance of considering the experience of data providers as well as of data users.

4.3 Used widely to inform understanding and shape action

4.3.1 Why we include wide statistics use to inform understanding and shape action

Whilst we acknowledge that statistics can serve the public good by providing insights, we also recognise that using these insights can allow for greater impact. As described in the book From GDP to Sustainable Wellbeing , 'lives will only be changed if those who should make use of the statistics do in fact make use of them'. This sentiment was echoed in our public dialogue project, where participants expressed that they wanted to experience tangible change from data being used for research and statistics, for example through local service provision, national policy making, and research with clear applications. This led us to conclude that on top of providing insight, using statistics unlocks additional outcomes that are in the public good.

When considering what uses may serve the public good, we reviewed evidence from our public dialogue project where we heard participants express that to them, serving the public good was about meeting real-world needs, sharing both tangible and less tangible examples of needs being met. The examples we heard sat into two categories – statistics being used to inform understanding and to shape action.

Do we need to talk about understanding and action?
Does reference to understanding and action add anything beyond stating that wide use serves the public good? Is this too broad, focused on the public good in general being served rather than the specific role official statistics play? Would it be useful to share information on when members of the public see understanding and action as serving the public good, or does that stray too far from impartiality?

When we propose that statistics play a part in both action and understanding, it would be an impossible burden to consider them the sole driver of behaviours or beliefs. This position is congruous with the Statistics and Registration Service Act (2007), which puts the role of statistics as *assisting* rather than driving action and understanding. For this reason, we have chosen to refer to *informing* understanding and *shaping* action, to acknowledge that use of statistic will not be the only influence, and statistics will play a part in a wider landscape.

We have since evolved our perspective on statistics use to now emphasise the breadth of use by saying 'used widely' rather than just 'used'. This is for two reasons. Firstly, OSR's stance has long been that 'statistics serving the public good means the widest possible usage of statistics', as described in an OSR blog post on

exploring the value of statistics for the public. Secondly, our public dialogue project illuminated negative perceptions around missed opportunities for using data from members of the public who participated. We therefore found it important to emphasise wide use in our statement.

4.3.2 What we mean by statistics being used widely to inform understanding and shape action

As referenced at the start of this paper, it is important to us in OSR that we uphold our independence, including through remaining neutral to conceptions of what outcomes are in the public good. In some ways this makes it difficult to explore the use of statistics, as evidence we have on what uses serve the public good or not reflect the worldview of the people it comes from and may not always be neutral. To uphold our values, we have reflected upon our evidence and only present information that we anticipate would be largely unchallenged irrespective of political affiliation. As such, in our understanding and statement, statistics being used widely includes:

- Being used rather than misused;
- Being used at any point in time;
- Having a wide range of (direct or passive) users;
- Being usable; and
- Benefitting the public.

While it is not in our place to dictate to what ends statistics are used for, it is still important to make the distinction between use and misuse of statistics. When we talk about misuse, we are referring to incorrect or inappropriate use of statistics, and as described in our Annual Review of UK Statistics Authority Casework 2021/22 this can be misuse 'whether deliberately or not'. Our 2022 update on COVID-19 lessons learned identified a risk to health and social care statistics serving the public good if they are misused. We anticipate that misuse may be problematic for two reasons: firstly, it can lead to erroneous conclusions; and secondly repeated or high-profile misuses may erode public confidence in official statistics more widely, thereby impacting the public good that other statistics could serve as well. Therefore, in our understanding, statistics being *used* can serve the public good, but this does not include statistics being *misused.*

A further area that we see as appropriate to comment upon is the timescale in which use may occur. We propose that use can be at any point in time – it does not need to be immediate, and can be far off in the future. This stance originated from our public dialogue project, where members of the public stated that the impacts of something (such as statistics use) could be immediate, incremental or in the future and still serve the public good. Participant discussion here was specifically about impacts, however when we considered this further, we expected that the use itself may be in the future as well. For example, during stakeholder discussions we heard that looking back on statistics from the future as historical records could serve the public good.

How to understand future uses?

We see that spending money on statistics which are never used cannot serve the public good, but also that uses in the future can serve the public good. We cannot know what topics will be of interest or use in the future, so how can we reconcile this? Is greater public good served by spreading limited resources across a broad portfolio that may become relevant in the future, or by focussing on a narrower portfolio that directly aligns to user needs at present? How might statistics producers begin to capture the needs of future users?

As we seek the widest possible use of statistics, we anticipate that more use will happen if there are more users. Unlike producers of private statistics, 'official statisticians do not work for a specific consumer or a small group of users; they receive public funds to be at the service of the society at large' (Bodin, 2011). This underpins our stance that there is no single user sufficient to say government statistics serve the public good, and a range of individuals and organisations should be considered as potential users in serving the public good. For example:

- The user could be government, such as using statistics in the 'development and evaluation of public policy' as described in the Statistics and Registration Service Act (2007),which is an example of both informing understanding (evaluation) and shaping action (development).
- The user could be individual citizens, who may utilise statistics to hold government accountable for their actions. This is described in a parliamentary report on government transparency and accountability during COVID-19, noting that it is central for democracy for members of the public go be able to understand government decisions (informing understanding) and hold government to account (shaping action).
- The user could be academics, undertaking research or providing a new evidence-base for decision making. Providing an evidence-base for public policy (informing understanding) was the main aim we saw by academics when we analysed applications to access public data, and this report explores other uses academics have for data that may also apply to statistics.
- The user could be the media, with ESCoE research finding that the media see themselves as having a role in translating statistics for the public, including contextualising them and using accessible language (action to inform public understanding).
- The user could be other members of civic society, for example Full Fact who is a registered charity that uses government statistics to fact check claims made in society (shaping action).
- The user could be the private sector, for example businesses adapting their pricing based on inflation statistics (shaping action), which in turn influences levels of inflation and impacts society.

What about artificial intelligence (AI)?
In exploring who uses statistics in serving the public good, it was suggested to us that we might consider the role that AI and bots play. Bots use statistics in a way

that impacts society, for example automated trading of stocks influences the economy we live in, so should we include them as something that can use statistics in service of the public good? Or is the role held by those who programme them? And will this change as AI becomes increasingly sophisticated?

For each of these users, 'statistics are only useful if they can be used' (Statistics for the Public Good). In this quote, we interpret 'if they can be used' to refer to the idea that statistics must be usable. But being usable is a broad statement – what is usable by one person may not be by another. Usability is explored in the section of this think piece on statistics as public assets, in terms of accessibility. In addition to accessibility, we propose that other dimensions of statistical quality also play into usability. For example, being available at the moment needed for use (timeliness), and having appropriate information explaining the statistic (clarity) both influence how usable statistics are.

Implicit in our assertion that statistics must be useable is the perspective that users will directly use statistics themselves. However, we anticipate that this is not always the case. In addition to discussing traditional users (those who directly use statistics) the UNECE work on valuing official statistics also explores the concept of 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'. We see this concept of passive use as an important way in which statistics can be used to serve the public good – each individual need not directly interact with statistics for the public good to be served.

There is, perhaps, an obvious aspect of statistics use in service of the public good that we have not touched on so far: an assumption that to serve the public good these uses are somehow 'good' - we heard from participants in our public dialogue project that uses in service of the public good were about creating the 'greatest good' in any given context. Even this apparently straightforward idea is fraught with complexity – who is this public we are referring to, and what would be 'good' for them? These questions bring us back to the introduction of our think piece, where we noted that defining the public good in a specific context requires defining who the public is, and what the good is for them. Here, we acknowledged that it is not within our remit to dictate what is 'good'.

What is 'good'?
Is it appropriate to prioritise impartiality and remain neutral to what is 'good', or does this undermine our position in that malicious or ill intended uses are not excluded from it? Is it even our place to label some uses as malicious or ill intended? Would it be useful to explore how benefits of use must outweigh harms, or is this equally unhelpful as setting out what 'benefits' might be straying beyond neutrality? Would providing information on 'good' even support the statistical system in serving it, or does this stray beyond producer remit anyway?

However, avoiding reference to what the public good is impinges upon our goal of providing practical support – there are instances where such a definition would influence what actions producers may take if they wish to serve the public good (for example which topics to prioritise or how to disseminate statistics). As such, while avoiding perspectives on what 'good' might be, we have considered who the 'public' may be that must benefit for understanding and actions to serve the public good. In some instances, even evidence in this space reflects opinions, and may not be seen as an absolute. Where we have identified this risk, we have not ventured to propose a stance.

There are two areas of who the public are where we have reached a position. Firstly, we propose that the public good does not necessarily require each individual within society to directly benefit, in line with the National Data Guardian's explanation of what public benefit means for the context of public benefit evaluations and views expressed in the follow-up workshop in our public dialogue project. As well as aligning to this evidence, we endorse this perspective because it is difficult to imagine an instance where every individual could directly benefit from a statistic being used. Secondly, we anticipate that 'the public' in this instance may be broad, encompassing private citizens, commercial enterprises, the third sector and government. Useful evidence for arriving at this stance was the Public Administration and Constitutional Affairs Committee report from 2019 into the governance of official statistics, which states 'serving the public good is understood to mean that users of statistics have the data they need to make informed decisions, whether they be in government, business, the third sector or the general public'.

Despite two areas of clarity, our position on who should benefit for the use of a statistic to serve the public good remains high level. We instinctively suppose that uses of statistics which serve the public good must be something beyond personal utility, creating community benefit above the sum of transactional gains for individuals, however on the specifics of this and how to best articulate it we are unclear.

Who benefits?
How can we better describe the public benefitting from statistics uses that serve the public good? How do we reflect that it has to be more than individual gain, and that the public good has some collective element? At what point does clarifying who benefits stray away from impartiality? Is this discussion even of use to our statistical system, or is it too far beyond what our system can influence?

While we have further evidence from our research on perspectives around who might need to benefit to state a statistic is serving the public good, this often reflects the values of the individuals who expressed it, and therefore could be seen as leaning towards a specific political affiliation. In addition, implementation guidance for the Fundamental Principles of Official Statistics describes impartiality as including all users being given equal treatment and access to statistical information, therefore any evidence that could be seen to encourage one user over another would be

incongruent with the values of official statistics. As such, to ensure impartiality in our work we have chosen not to include further information in our understanding of how statistics can serve the public good that could be used to prioritise users.

5 Next steps

We propose that a shared understanding of the public good of government statistics will help those across the statistical system work towards a common goal: statistics that serve the public good. On a practical note, deep understanding of this topic will inform our regulatory approach, allowing our guidance to support statistics producers in making decisions that best serve the public good.

From our work to date, there are some things we *know* about statistics serving the public good (for example legal definitions), some things we have *concluded* to be true based on our experiences to date (for example the role of statistics in providing insight), and some things we have an *idea* about but are still trying to refine our position (for example who needs to benefit for us to say that statistics serve the public good). From this, we recognise that there is still more work to be done before a shared understanding is established.

To get to a shared understanding, we will continue developing our thinking as our evidence-base evolves and will update our stance as we learn more. We would like you to be a part of this work; we invite you to contact us at regulation@statistics.gov.uk to share your views on this think piece, and we look forward to working with you.