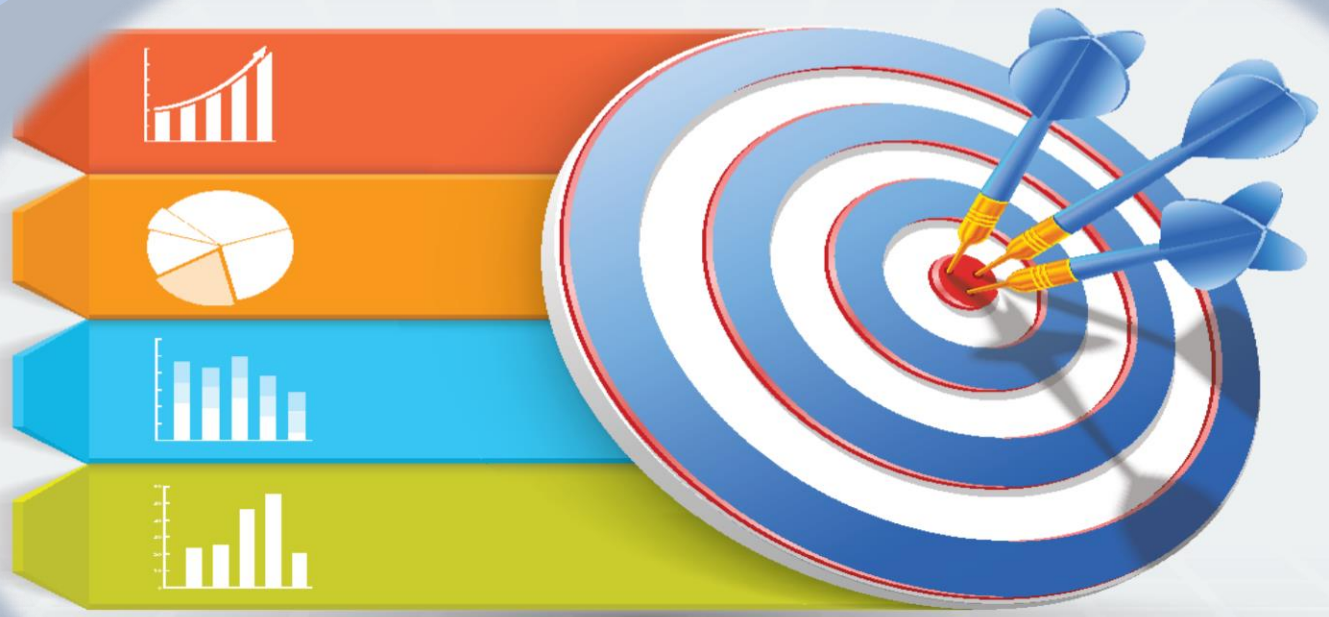

Technical Annex, Evaluating measures of value: Measuring the Value of Official Statistics

Contents page

Contents page

view in slide show and select tiles…

Evaluating measures framework

What are objective measures?

Objective measures

What are subjective measures?

Subjective measures

What are Monetary measures?

Monetary measures

Evaluating measures

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
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.
A number of related questions were asked in the template, including: whether they are currently producing the measure in their organization; for how long it has been produced; what it is used for, and whether this use is considered to contribute to increasing the value of the statistical product or service in question; for what audience the measure is intended; how important or useful the measure is considered to be and why; how the need for the measure was identified (if at all).

Completed templates from 18 countries and organizations provided a rich source of information. The findings were considered collectively by Task Force members, who then worked through each measure in turn by means of an online polling tool to decide whether the balance of opinion tended towards keeping, discarding or amending the measure. In reaching their decisions, the Task Force considered these detailed submissions from participating countries, the material shared in the case studies, as well as their own personal expertise. Decisions were guided by the principles described in section 4.2.
A key finding from the review was that, as discussed in Chapter 3 and Chapter 4, many of the measures proposed in the initial framework are grounded in a production-based mindset with respect to the perception of the value of official statistics. Such measures may carry great importance for operational and management processes within an NSO, and/or they may serve to gauge statistical quality. However, such indicators alone cannot offer a comprehensive assessment of value—they need to be complemented by measures grounded in a customer-based perspective.
With this distinction in mind, the evaluation of measures below attempts to distinguish between these two broad perspectives, albeit recognizing that the distinction is by no means clear-cut and that some measures may fall into both categories or be difficult to categorize.

Objective Measures

The Task Force reviewed the objective measures proposed in the framework plus additional ones that were put forward during a Task Force brainstorming (full list set out in Annex 1). Of these, fewer than half (15 from the original framework and two additional ones) were found to be potentially suitable as measures of the value of official statistics according to the country experiences and the guiding principles detailed in section 4.2.

Objective measures with potential as indicators of value

Production based

Consumer based

Punctuality of statistical releases (share of punctual/late/cancelled releases	Share of error-free statistical releases	Quick correction of errors (Average delay in correction of errors in releases)	Accuracy of statistics – average number of revisions required	Timeliness of statistical releases (weeks from the reference period)	Number of website visits
Downloads of statistical data by domain	Visits to digital library/publications webpage	Number of social media followers	Number of agreements to use microdata for research	Number of media citations	Number of citations in research/policy work
Most cited statistics	Most viewed/downloaded statistics	Number of retweets (comments, likes etc.)	Exploration of Digital Object Identifiers (DOI's) to monitor use of official data in online publications	Interoperability & link ability – how easy it is to work with, aggregate and join up data sets.	

Objective measures not currently recommended as indicators of value

Quality

Potential measure of value 1: ***Punctuality of statistical releases (share of punctual/late/cancelled releases).***

Rationale: Production based

Long-standing quality dimension and key performance indicator that allows for reporting and shows track records for quality management. It can help with value from the perspective of trustworthiness, reliability and satisfaction of the users. 'Users get on 'on-time' information, published according to a pre-release calendar and work programmes. For the user, all releases have equal weight with everyone getting it at the same time.

Consideration: Change to, "Share of scheduled releases issued on time/late/cancelled" (with accompanying reasons). Measure should be accompanied by a target

Quality

Potential measure of value 2: ***Share of error free statistical releases***

Rationale: Production based: Key performance indicator.

Consideration: Should define and combine with correction and accuracy. Measure should be accompanied by a target

Quality

Potential measure of value 3: ***Quick correction of errors (average delay in correction of errors in releases)***

Rationale: Production based Key performance indicator.

Consideration: Need to define and combine with correction and accuracy. Measure should be accompanied by a target

Quality

Potential measure of value 4: ***Accuracy of statistics – average number of revisions required***

Rationale: Production based and consumer based

Revisions are used for quality management. Important but interpretation is not the same everywhere and there are often planned and unplanned revisions - we need to be clear on these nuanced differences. Some countries have planned revisions which are mapped out in advance which allow timely delivery with an opportunity to add richer data when that is available at a later date. That is part of standard statistical production process that users are aware of.

Observations and considerations*:*

Suggest reword to 'Average number of revisions by type of revision over the x amount of time. To also include "Number of statistical breaches'.

Need to define and combine with correction and share of error free releases. Measure should be accompanied by a target

Quality

Potential measure of value 5: ***Timeliness of statistical releases (weeks from the reference period)***

Rationale: Production and consumer based :

Long-standing quality dimension and key performance indicator with legislative requirements to comply, used for assessing the speed of the dissemination of data and analysis to users. Particularly relevant during times for greater user need, as evidenced in the Covid-19 pandemic. If too much time passes some data can become irrelevant, however, it can still be valuable if the best available source.

Consideration: Suggest **r**eword to, 'Number of statistical releases published from the reference period?

Measure should be accompanied by a target.

Use

Potential measure of value 6: ***Number of website visits***
Rationale: Production based

Highly valued useful and tested key performance measure, used for many years by the majority of NSOs to track website and usage (user volume/behaviour and access to products). It shows a general gauge of interest in official statistics.
However, it does not tell us about the result of the site visit. This measure is an attempt to capture a conclusion about customer value but does not capture satisfaction, useability, benefit and impact.
When using the 'number' of website visits, there is a need to mitigate against poor web design which could drive up page views, users also vary, so there is a need to distinguish between them (Important especially when trying to serve multiple audiences with publications) and reach of users, for example, some specialist users may seek information once and quickly, whereby lay-users may need to hunt for their answers simply because they don't know where, or what, to look for.

Consideration: *Amend to '"Number of website views"* and Include as part of a suite of measures looking at the website. There is a need to be clear what is included (E.G. a sum of visits to the main page and all subpages and if include products & applications). It maybe beneficial to look at interaction/share than raw users/length of time viewing/active time on page/active readership.
Various organizations use different web traffic analytics' tools, so might not be directly comparable but it would be useful to share tools.

Measure should be accompanied by a target.

.

Use

Potential measure of value 7: Downloads of statistical data by domain
Rationale: Production based

Consideration: Key performance measure used routinely for many NSOs. Measures volume and shows the domains and data that are of interest and in demand by users. Information can help inform future releases and products.
However, communication strongly influences the outcome of this measure. Broadband companies can bundle IP addresses, so might not always get a clear indication of users.

Not all downloads are equal (and value realized depends upon the application of the data by the user). Seasonality might also be a factor and difference in definitions of statistical domain may complicate comparability.

Consideration: What is meant by data, does this include publications as well or only tables and datasets? What is meant by downloads: visits or exports? Maybe use "statistical products. Definition of domain should be clear and these can vary across NSOs.
To be part of a suite of questions and defined more precisely.
Measure should be accompanied by a target.

Use

Potential measure of value 8: ***Visits to digital library/publications webpage***
Rationale: Production based: This is a useful measure of website activity and provides a key piece of information on reach and volume of public use. Looking at page views and sessions are a transparent measure of use of a digital product, however, it does not tell us anything about satisfaction, useability or indeed impact.
Knowing the statistical domains that users are interested in is somewhat helpful in getting to know our audience, but it is not an essential part of establishing the content of the website.
Breaking down data by topic provides more useable information than the overall number or trend and entry point for understanding but important to distinguish between specialists (lower user) and generalists and in particular those publications trying to serve multiple audiences.

Consideration: This measure is similar to number of website visits consider merging ?). Measure should be accompanied by a target

Use

Potential measure of value 9:*Number of social media followers*
Rationale: Production based: Key measure around engagement with users and provides a good indicator of social media account growth. Important communication platform to reach a wider audience, (especially younger people). The interactions and interest around products are also helpful in getting to know our audience - especially pertinent for engagement when rolling out a census. Ensures the visibility and promotes official statistics among a wide circle of current and potential users. However, following on twitter for example does not necessarily equate to 'use' of statistics. It is a useful indicator as part of wider suite of metrics and combined with other measures of engagement.
Whilst this measure can indicate long-term interest as a proxy of implied value, there is a risk that this measure could change publication/communication channels simply to suggest increased value.

Consideration: Keep as is. Useful to look at in relation to major releases etc and how other organizations compare the use of different social media channels. Measure should be accompanied by a target

Use

Potential measure of value 10:*Number of agreements to use microdata for research*
Rationale: Production based: Reflects the interest and demand of a particular group of users (researchers, business and organizations) accessing our products and strengthens the collaboration with research community. However, demand can be variable and can be influenced by events such as Covid (as impacts on how people can work in a secure research environment) and low number of agreements might reflect difficulty with the process rather than lack interest.
Consideration: Recommendation: Keep as is or reword to, 'Number of accredited research projects that have access to microdata'. Measure should be accompanied by a target.

Relevance

Potential measure of value 11: ***Number of media citations***
Rationale: Consumer based
Media is a target group for NSOs, seen as an enabler, an important redistributor in presenting statistics and helps with raising visibility, interest and awareness in statistical information. This measure helps track the reach of publications and helps show how the media find our work valuable. Shows how statistics are used in reporting, a gauge for public interest and helps monitor public interest and impact on public debate. The media's reach is far greater than a NSO alone and offers an exponential increase in audience - therefore increasing value to the community. It brings valuable data for defining communication strategies**.**
However, this measure can often be at the mercy of the relationship between the NSO and the media. Additionally, looking at the number of citations is also not enough without knowing the relevance of the media in which they appeared.

Consideration: Need to define 'citation' and its use of a statistic or output (does it include a mention of a survey or census or the NSO). It is not apparent that this measure directly talks about the ‘relevance’ of the data set. In-depth analysis of media-coverage could be a complementary method. Example of tool – Media toolkit. Measure should be accompanied by a target.

Relevance

Potential measure of value 12 : *Number of citations in research/policy work*
Rationale: Consumer based: *Measure sho*ws the contribution of statistics to scientific output and policy making. This measure records actual use of data.

Consideration: Need to define ‘citation’. Need to separate out policy, public organization and research. All are important but the former is probably more important in the judgement of the value of statistics. Not a measure of relevance. Measure should be accompanied by a target

Relevance

Potential measure of value 13 : *Most cited statistics*
Rationale: Consumer based: Measure is great for maintaining visibility in the public eye and gives the overview of the orientation of the users towards different statistics and what is most important and of interest to media and society/public. However, caution in claiming that most-cited equals most valuable.
Consideration: Need to define ‘citations’ (does this mean media citations or in- case citations? Need to share tools. Quantitative information is not enough on this area In-depth analysis of media-coverage could be a complementary method.

Relevance

Potential measure of value 14 : ***Most viewed/downloaded statistics***
Rationale: Production based: Majority of NSOs use this as a measure or KPI for monitoring use of products and user volume and interest. Helps to plan releases and shows what statistics are in demand and used for further in-depth research. This measure references the most downloaded as a proxy of value, however, it is not a precise measure, it does not mean that least downloaded is the least valuable. It is a volume measure and does not say anything about the users' satisfaction**.** Additionally, website usability and users IP addresses do not always figure and users can prohibit their browsing, so only a sample provided.
Consideration: Split into 2 measures, 'Most viewed' on website and 'most downloaded' from website'. As above follow up with further website analytics. Measure should be accompanied by a target.

Relevance

Potential measure of value 15 : ***Number of retweets (comments, likes etc.)***
Rationale: Production based: Social media is an important communication channel in promoting official statistics and used as KPI for some NSOs for measuring an online presence and interaction with the public. Social media engagement metrics can help towards understanding reach, impact and interaction with the public and specialist users. It can help us to understand relative importance/priority of our statistics and help to see a direct influence in public debate. However, if something is re-tweeted, does that mean it is relevant? Are all re-tweets done in a positive context? What if re-tweeted in the context of evidencing government wasted effort? Or any other negative context?
While social media has become more and more important as a communication channel, not all organizations are on social media.

Consideration: A useful measure when used as part of a wider suite of metrics as important to get holistic view. Not useful in just counting but combining different measures and classify types of engagement (for example, quoted tweets show more engagement than likes (also what about other platforms, such as Facebook, LinkedIn, Instagram, YouTube). Measure should be accompanied by a target. (link to case studies (BLS have done an engagement study involving machine learning and understanding the sentiment surrounding how BLS is portrayed in the Twitter verse) .

Use : New measure proposed in group brainstorming

Potential measure of value 16 :*Exploration of Digital Object Identifiers (DOIs) to monitor use and impact of official data in online publications*
Rationale: Consumer based: . This area of work has been suggested as a potential way to track outputs and their added value and impact. DOIs are seen as an important factor in the academic and scientific world and are already used by some NSOs. The US provide a case study and the UK are currently developing a pilot to test its potential for measuring impact of outputs.

Consideration: This should be considered as a potential measure. Suggested measure that could look something like: 'Number of re-uses of a dataset as revealed by DOIs.

Use : New measure proposed in group brainstorming

Potential measure of value 17 :*Interoperability & link ability - working with, aggregating and joining up data sets*

Rationale: Production based: . This area of work has been suggested on the Wiki and all NSOs rate work this as VERY important. This is also a complex area of work. The post-consumption value of statistics and analysis is not well understood and development of work in this area could help to fill that void.

Options suggested could include work around objective approaches, E.g. compliance with standard formats and common linkages and subjective approaches and a need to find out more from users what they need. Another option is to look at open data sources and standards.

Consideration: NSOs are really interested in work in this area. The measure should be more clearly defined for the user and the producer.

Objective measures not currently recommended as indicators of value.

Press on each slide to see the measures

Quality

Relevance

Transparency

Use

New measures proposed in brainstorming group

Objective measures not currently recommended as indicators of value – Quality

Not recommended as a measure: *Share of statistics released with quality assessments*
Rationale: Adherence to quality assurance is part of producing all official statistics.
Consideration: It is important to provide clear guidelines on what constitutes quality assessments (including standardization of reports using an agreed quality framework).

Not recommended as a measure: *Number of new visualization tools introduced*
Rationale: This is not an indicator of quality or improvements to outputs. Visualization tools' is vague and furthermore, what counts as a visualization tool? New ways of presenting statistics isn't just down to whether a tool is made or not, it's about the content and if it is written in a clear way.
Consideration: It could incentivize an unnecessary diversification of dissemination tools that have similar functionalities but with siloed databases powering them.
It's not clear that there is a monotonic relationship between the metric and the value to Official Statistics, making interpretation difficult. A lot of tools can mean that you are not basing development on your strategy and you are allowing your ecosystem to grow without order. The adequate number of tools should be determined by your strategy regarding the different users' needs that you are trying to satisfy (if external) or by the capabilities within your organization (if the users are internal).

Not recommended as a measure: Innovation or quality awards received
Rationale: The results of this metric do not necessarily translate to better quality of products nor of more statistical capacities within the organization. It could foster an unhealthy competitive environment within the organization and/or create the incentive to workers to seek awards with lesser quality and difficulty. Not clear that there is a 'monotonic' relationship between the metric and the value to Official Statistics, making interpretation difficult. Difficult to measure and difficult to compare (what would be the objective criteria) what awards would be used and are they all treated equally). Frequency could be low too for measurement.
Consideration: May be more effective to have presence of an innovation awards program

Objective measures not currently recommended as indicators of value – Transparency

Not recommended as a measure: *Share of statistics released with metadata*
Rationale: Metadata a vital element of improving statistical literacy and assisting users in getting what numbers are about and how they should be interpreted. Metadata users lead to trust of user and thus increase its value. Not clear that there is a 'monotonic' relationship between the metric and the value to Official Statistics, making interpretation difficult:
Consideration: As with average age of metadata ''periodically reviewing metadata is accurate and of high quality' or 'Metadata are presented, and archived, in a form that facilitates proper interpretation and meaningful comparisons'. The mere existence of metadata is not an exhaustive measure. The quality (fulness, timeliness, comprehensibility etc.) of metadata descriptions is an even more important.

Not recommended as a measure: *Number of blog posts by official statisticians*
Rationale: Counting blogs or other self-authored opinion pieces do not tell us about value. Awarding self-authored blogposts with no review processes in place could be a risk to the Organization's reputation. If a process is in place that assures the quality of the blogposts themselves, resources have to be allocated to it. Furthermore, any creation of a measure needs to fall within the strategical communication goals of the organization. Though feedback and usefulness of organizational blogs could be linked to communication metrics such as web and social media measures.
Consideration: Blogs can be a useful tool for communication.

Not recommended as a measure: ***Number of users trained***
Rationale: Statistical programmes need to be in place for users to be defined and trained.
Training users has value but resources are scarce and the determination of who to train and how has to be linked to strategic directives. The critique for the inclusion of this metric is that it is not a measure that quantifies the value created by training users.
To note also, many users are trained through many sources (schools/colleagues/universities/their employers). The problem with this metric is that is assumes that all users are equally important to achieve the organizations goals. For example, we could reach X elementary students or X policymakers: we are reaching the same number of user's but the value created regarding the use of official statistics and NSO's mandates will probably be very different. Not clear that there is a monotonic relationship between the metric and the value to Official Statistics, making interpretation difficult: Denormalized measures such as this should always be regarded with care: X numbers of users are reached out of how many you intended to reach.
Consideration: A level of training is required for users to access datasets, work could be done around looking at 'number of researchers given accreditation'. (Also linked to accessibility).

Not recommended as a measure: ***Number of journalists trained***
Rationale: Statistical programmes need to be in place for users to be trained, this could include journalists. Journalists are an important link between the NSO and general public - they help communicate information and will do so more accurately if they know more about methods and tools. They can help with building trust and improve value of the production of statistics.
The critique for this metric is that it is not a measure that quantifies the value created by training journalists. It creates perverse incentives: If your strategic directive is to reach journalists because of the indirect reach to users, then not all journalists are equal for that strategic goal.
Consideration: Journalists of national mediums will have bigger audiences but will also have more resources and maybe wouldn't even need the training. Local journalist could potentially benefit more of a training.

Not recommended as a measure: *Average 'age' of metadata on website*
Rationale: Metadata a vital element of improving statistical literacy and assisting users in getting what numbers are about and how they should be interpreted. Not clear that there is a 'monotonic' relationship between the metric and the value to Official Statistics, making interpretation difficult. Metadata is required to produce the outputs, but does age of the metadata directly impinge upon the value? If seeking data from 3-5 years ago, the data must reflect that period, data with a younger 'age' is not relevant. More useful perhaps to have something about updating metadata and specifying how often etc. Similarly, if the situation is stable in a particular domain for many years, it may be perfectly normal for the metadata to be "old". This does not mean there is an issue with it. In judging the age of metadata, it is also important to take into consideration the frequency of publication (e. g. a three-yearly data collection does not have to have an annually refreshed metadata file).
Consideration: Needs refining, suggestion, 'Presence of metadata' and or metadata policy/strategy/system (that sets standards/ guidance on what constitutes acceptable metadata that many NSOs have).

Objective measures not currently recommended as indicators of value – Use

Not recommended as a measure:*Number of partnership agreements*
Rationale: Agreements and memoranda of understanding, on data exchange are important coordination tools for NSOs within the NSS. Greater numbers of such agreements could be reflective of many positive aspects, such as harmonization within the NSS, increased access to administrative data, etc. However, the absolute number of agreements is less relevant than the nature and purpose of the partnership.
Consideration: If using this as a measure, types of partnerships would need to be defined and considered separately. Also, effective partnerships may be created without formal agreements.
There is no direct and unequivocal link between the increased efficiency within the office and the value to the user—hence, this is not a good measure of value.

Not recommended as a measure: *Number of data cells in online statistical databases*
Rationale: The sheer number of data cells does not imply value. Quarterly GDP is only one cell. Is it then less valuable than an administrative register with thousands of them? The value of the data cell lies within its relevance as defined by the Fundamental Principles of Statistics. If decisions are based on the number of data cells in an output it would be difficult to reconcile the relevance behind data points that imply a lot of effort and money to create.
Consideration:

Not recommended as a measure:*Number of newsfeed subscribers*
Rationale: Covered by better metric
Consideration: Statistics generated by web metrics and social platforms (media relations and social media) should be sufficient to capture this.

Not recommended as a measure: *Number of statistics app downloads*
Rationale: While a few NSOs have a statistical app game most do not. The number of downloads does not tell you about value and use.
Consideration: Active use of an app can be measured by the number of requests that app is generating explicitly from databases.

Not recommended as a measure: *Number of chat contacts*
Rationale: It shows a level of interest of users and level of communication, however, not clear that there is a 'monotonic' relationship between the metric and the value to official statistics, making interpretation difficult: What does the number reported mean? If the number of chats rise is that because more people are interested, or a symptom for other reasons, such as un-navigable website? or are the users becoming more stats savvy and asking trickier questions of the products.
Consideration: More discussion needed on the link between chatting with an NSO and use. Chat contacts also needs to be defined.

Not recommended as a measure:*Number of agreements for chargeable services/sales of products or services*
Rationale: Measure is only applicable if the NSO is charging any fee for a product or service. NSOs are state institutions implementing functions aimed at public interest so this measure would generally depend on national legislation. However, most products are available and can be downloaded free of charge. Most do not sell publications and those that do the demand is decreasing continuously as opposed to demand for data and visualized products.
As a public service data free at point of delivery and charging less important- number of agreements may be also be low and could be viewed as negative.
Consideration: This cuts across to monetary as sales of products is revenue and also indicates willingness to pay.

Objective measures not currently recommended as indicators of value – Relevance

Not recommended as a measure: ***Number of tailored services by user groups (linked to number of agreements to use microdata for research).***
Rationale/Consideration: It could be included in an annual reporting as it improves and strengthens partnerships with users, however, measure is ambiguous, what is meant by tailored service, what would 'a number of mean? Does it mean a growing depth of interest and high demand for official statistics or an indication of gaps and maybe not doing a good job with its publications (could highlighting gaps be a positive and more tailored services imply confidence of users. It is important to know the user groups that require types of services, as well as the frequency of these requirements.

Not recommended as a measure: ***Number of new end-products or services***
Rationale/Consideration: It could be included in annual reporting but only helpful in concert with other measures. It could show level of innovations in the organization and awareness of changing user needs.
Pace of creating new products can be indicative of value IF we can show that we are responding to new demands and/or new capabilities to meet demands. Does the act of releasing a new service make it relevant, or of value? Assuming no, then why count them?
Defining what is a product is difficult and number alone not sufficient. Is a data point a new product? Or a table? If we transform an existing table, will it be a new product and if yes from what extent of transformation?
NSOs with a lower degree of organizational maturity will have a higher propensity to produce new products/services that advanced ones.

New measures proposed in brainstorming group…

Press on each slide to see the measures

Quality

Relevance

Use

New measures proposed in group brainstorming - USE

Not recommended as a measure: ***Number of memberships in international expert groups***
Rationale: Participation in international forums does reflect the level of institutional and international engagement. However, number of memberships does not say anything about the value of those groups and is limited in terms of output and impact.
Is the NSO part of a committee/bureau/board which influences work and decisions; or is it just a member (i.e. observers). Some NSOs can be members of EVERYTHING but don't influence much. Increasing the number does not automatically lead to a better overall value.
The measure could also create a perverse incentive. If this metric is used it implies that having a high number is a good thing and then the behaviour promoted is to join more groups, regardless of strategic objectives.
Consideration: The type of membership would also need to be defined.

Not recommended as a measure: ***Number of papers/presentations/inputs that contribute to international statistical work***
Rationale: International engagement is good and strengthens relationships but hard to extract value added purely from number. It is how they are used that are important. Some topics will invite more papers/speakers but each marginal paper will add small value; whereas in some areas a paper may progress work greatly. This measure encourages big numbers rather than promoting good practice.
Consideration: Could potentially be high-level measure for influence in an area. Better insight may be drawn from follow-up.

Not recommended as a measure:***Number of international study visits hosted***
Rationale: Visits help to generate an exchange of experience and good practice, however, measuring number of visits does not provide insight on quality of visit and what was achieved.
Consideration: Feedback on visits would allow for better distinction of visits

Not recommended as a measure: ***Number of participants in national and international statistical competitions in schools***
Rationale: Not clear if talking about use of 'own' products specifically or just stats in general. It might show an increased value in data analytics or related subjects more generally, but it does not mean that our products and services are relevant to our customers
Consideration: More work needed on defining this measure. The area provides an opportunity to promote statistics, increase statistical literacy and boost engagement but need to draw a logical link (that says 'students are doing stats (competitions) because we have somehow encouraged it') and whether we're talking about use of our own products specifically, or just stats in general.
It is of strategic importance for NSOs to work for better statistical literacy and sufficient education on the use of statistics and partnerships with schools to increase statistical awareness and literacy. Collaboration in this area is a fast-growing in NSOs

New measures proposed in group brainstorming -Quality

Not recommended as a measure: ***How much data collection is automated?***
Rationale: Greater automation in the collection of data tends to imply higher quality statistics. Lots of other factors intervene in relation quality-automation. It is not necessarily true that 100 per cent automation is desirable or achievable; certain field conditions need to be taken into account.
Consideration: Important area but measure needs defining. Needs may vary on how much is/should be automated how what is meant by automated? It could be an indication of the amount of registers and our goals to minimize the surveys. So it should be combined with the number of surveys and an explanation why we still use them.

New measures proposed in group brainstorming -Relevance

Not recommended as a measure: ***Working time used for development***
Rationale/Consideration: Measure important for planning, modernizing and innovation, however, spending time on something does not make it relevant. Would need to show value in relation to time spent developing & producing products. Spending time in building internal capabilities is also important and this could disincentivize adequate time allocation for it. Not clear that there is a 'monotonic' relationship between the metric and the value to official statistics, making interpretation difficult: In order for this metric to mean something it should be normalized by the value generated by the products that are being developed. For software specifically, because it evolves constantly, better frameworks to build it arise that in most cases reduces the time for development. Less hours could mean better software.

Not recommended as a measure: Number ***of students choosing Statistics as their profile***
Rationale/Consideration: Does not show value.

Not recommended as a measure: Number ***of participants in statistical events (stands during knowledge fairs, night of Museums etc.)***
Rationale/Consideration: Events can help increase statistical awareness and promote statistics but there needs to be greater detail to show the value of just collecting numbers (similar to the number of people in competitions). Additionally, the number of participants could be affected to its communication strategy efficacy, not by the quality of the event. Not clear that there is a 'monotonic' relationship between the metric and the value to Official Statistics, making interpretation difficult: What counts as a statistical event and how it is delivered and attended?

Not recommended as a measure: ***Number of candidates for the best doctor thesis in statistics***
Rationale/Consideration: It reflects the interest in statistics and maybe a university's ability to sell their programmes but it is hard to draw anything more than a tentative link between doctoral work and the work of the NSO.

Subjective Measures

User satisfaction surveys and targeted surveys

Some countries have undertaken targeted surveys and tested questions from the questionnaire, as per the 2018 Recommendations. Some of these include:

Mauritius conducted a user survey (2019) and produced their own shortened version of the questionnaire. They focused on questioning quality aspects such as satisfaction with specific products and services up front, followed by trust and dissemination aspects, with information about the respondent towards the end, as opposed to the order in the original framework. They cover largely the same questions, with very similar wording, but slightly simplified response options.

Armenia who conduct 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 questionnaire that is the light version of OECD Trust Survey and the last one was conducted in 2017.

Hungary uses an adapted version of the user satisfaction survey. They ask many of the same questions, though use response options adapted to their own user typology based on sectors. The personal information section is optimised so that it can also be adapted to other systems such as users in the contact centre. The questions asked include use of statistics, access of statistics, user satisfaction.

The UK no longer undertakes a large user satisfaction survey (in the format as outlined in the framework). The UKSA now outsource a survey on trust to NATCEN an independent organization who undertake a survey on public views on trust in official statistics. Questions were developed based upon the OECD recommendations upon which the UNECE framework is also based, so the questions overlap considerably. ESCoE have also undertaken some testing of the recommended questions in the valuing of economic statistics baseline report and a follow up study has also recently been undertaken.

Ireland has tested in depth the generic user satisfaction questionnaire as outlined by the framework.

The Task Force reviewed the subjective measures proposed in the framework plus additional ones that were put forward during Task Force brainstorming sessions (full list set out in Annex 1). The subjective measures with potential as indicators of value and those measures not recommended are shown in the following section.

Subjective Measures with potential as indicators of value

- Average reported 'ease' of users finding what they were looking for (accessibility)
- Average share of users who found what they were looking for (or those who couldn't find what they were looking for)
- Average share of users whose information needs were met
- Average rating of overall quality of website
- Average rating of ease of navigation on website
- Average reported ease of finding what looking for on website
- Average ease of locating data in database /data warehouse
- Average ease of extracting/downloading/manipulating/visualising data from database/Data warehouse
- Average satisfaction with visualisations /interactive tools/maps/graphs/etc
- Average ease of locating relevant metadata
- Average satisfaction with available metadata
- Degree of trust in official statistics
- Degree of belief in the political independence of NSO
- Degree of understanding by the remit of the NSO
- Future proofing of official statistics
- Increasing statistical literacy through school engagement and resources
- Share of users who recognise the corporate branding of the NSO

Subjective measures not currently recommended as indicators of value

Satisfaction with products and services

Potential measure of value 1: ***Average reported 'ease' of users finding what they were looking for (accessibility)***

Production based

Rationale: End users feature heavily in this measure and is crucial to better meet user's needs. The easier it is for users to find what they are looking for, the more use and value there will be. Useful indicator as part of a wider suite of digital metrics and helps with understanding **user satisfaction**. It shows the **website 'accessibility'** and how well the design of the navigation system is. Helps to show areas to improve website & other communication services.

Consideration: NSOs should differentiate between both products and user groups when compiling this indicator (experts might find stuff easily but maybe not for a general user or even a non-regular user). Also important is users' ability to find what they seek quickly/efficiently and importantly ease of understanding products and their suitability to gain feedback for improving products to ensure user's needs (on topics and themes) are met. Not all NSOs have tools. Measure should be accompanied by a target (for outputs and the site at large).

Satisfaction with products and services

Potential measure of value 2: ***Average share of users who found what they were looking for (or those who couldn't find what they were looking for)***
Production based
Rationale: Accessibility should be a primary goal to help ensure better meet user's needs. Value can only be achieved through the end user. Although could be difficult to fully measure as those who have not found what they were looking for might not be engaged
Consideration: Directly linked to above so could come out of that measure. (Covered under design, communication & metadata (website) below customer experience improvement. CONSIDER MERGING WITH ABOVE OR DROPPING). Measure should be accompanied by a target.

Satisfaction with products and services

Potential measure of value 3: ***Average share of users whose information needs were met***
Production based
Rationale: Crucial in understanding user satisfaction and identifying user's needs, in order to make improvements to better meet those needs. Also used in efforts to improve website and services.
Consideration: Important to emphasize difference with the one above- linked to 'Ease of users finding what they were looking for' Ease of understanding products. Again indicator is somewhat similar to the one above. Covered under design, communication & metadata (website).

Satisfaction with products and services

Potential measure of value 4: ***Average rating of overall quality of website***
Production based
Rationale: If the website is the most important dissemination tool, it's a good measure. It helps to improve the website that is most important channel for dissemination and communication of statistics. Talks to the value of the service, not that of any single official statistic.
Consideration: Some caution should be exercised. probably best within a suite of measures about website navigability. Maybe cross-reference with relevance and accessibility since for many, the website is the principal channel for finding & using most stat products. Measure should be accompanied by a target.

Satisfaction with products and services

Potential measure of value 5: ***Average rating of ease of navigation on website***
Production based
Rationale/Consideration: Similar to previous one. It is an important indicator as it helps to increase the use of statistics.
A useful indicator as part of a wider suite of digital metrics to understand the health of an output. Again, talks to the value of the service, not that of any single official statistic.

Satisfaction with products and services

Potential measure of value 6: ***Average reported ease of finding what they were looking for on website***
Production based
Rationale: It is an important indicator and relates directly to accessibility and its role in value. Helps to increase the use of statistics and helps to understand if we're helping meet user's needs.
Consideration: From the point of view of respondent, hard to distinguish between navigability and ease of finding what you're looking for. (maybe the former include things like ease of returning to homepage etc? but how much does that matter?). With website-related indicators, maybe it would be good to have one CORE indicator - an umbrella one – as part of a wider suite of digital measures

Satisfaction with products and services

Potential measure of value 7: ***Average ease of locating data in database/data warehouse***
Production based
Rationale: ***This is similar to above and*** **covered by previous measures** - *(A*lso relevance & accessibility). This could also be indirectly measured by things like download rates, bounce rates etc (i.e. instead of asking people directly how easy it was, look at their behaviour once they arrive on database page- do they pursue or give up?). It is important measure and can help to improve databases. It talks directly to accessibility and its role in value, however, there is a need to specify what is meant by database?
Consideration: Could also be indirectly measured by things like download rates, bounce rates etc (i.e. instead of asking people directly how easy was it, look at their behaviour once they arrive on database page- do they pursue or give up?).

Design, communication and metadata

Potential measure of value 8: ***Average ease of extracting/ downloading/ manipulating/ visualizing data from database/data warehouse***

Production based

Rationale/Consideration: : There are too many questions in one here. Would it consist of sub-indicators? 'ease' might not be the only valuable feature. You might be 'easily' able to draw a graph or visualization but it might be a rubbish one (example: tools that let you draw a choropleth map of absolute values of something). It helps to introduce new technologies. The same remark as to the website - maybe one core and others should be supporting?, i.e. developing the idea laid down in the core one? Talks directly to accessibility and its role in value. divide into series of indicators on the different aspects. Ease of downloading a dataset in csv format is not the same thing as ease of drawing a graph on the interface.

Design, communication and metadata

Potential measure of value 9: ***Average satisfaction with visualizations/interactive tools/maps/graphs/etc***
Production based
Rationale/Consideration: It helps to improve visualization and interactive tools. Talks directly to accessibility and its role in value.

Design, communication and metadata

Potential measure of value 10: ***Average ease of locating relevant metadata***
Production based
Rationale/Consideration: Talks directly to accessibility and its role in value. It is important measure that helps to improve metadata structure.

Design, communication and metadata

Potential measure of value 11: ***Average satisfaction with available metadata***
Production based
Rationale: It is an important measure that helps to improve a metadata structure and for establishing a metadata system in line with international metadata standards. Talks directly to accessibility and its role in value.
Consideration: There can be a number of dimensions to this indicator, so clarity is needed on features of metadata. Do we mean satisfaction with how much metadata is made available, or satisfaction with its actual content (e.g. age of dataset, understandability, accessibility, completeness etc).

Relevance, responsiveness and innovation: new measures proposed in group brainstorming

Potential measure of value 12: ***Future proofing of official statistics , e.g. data linkage across governments, analytical and coding capabilities of staff***

Producer based and consumer based

Rationale: A lot in this measure. Requires extensive knowledge of the internal processes and initiatives and does risk being rather introspective (how does it impact on perceived value among users that an NSO is 'future-proof'? Unless they know about it and it enhances their trust in our adaptability).

Consideration: Highly rated as an area that NSOs SHOULD focus on. Not a measure in the way it's currently expressed so needs to be defined with a target.

Relevance, responsiveness and innovation: new measures proposed in group brainstorming

Potential measure of value 13: ***Increasing statistical literacy through school engagement and resources***
Producer based and Consumer based
Rationale: Need to clarify whether the measure being proposed is a measure of statistical literacy, or a measure of INCREASE in it due to our efforts, or a measure of THE STRENGTH OF OUR EFFORTS to increase it. The 3 are quite different.
Consideration: Efforts to increase statistical literacy are no doubt important but are they indicative of value? The fact that we go to schools and help people understand 'stuff' might add to our value, but it isn't a measure of it. Needs reworking and refining but NSOs see this area as important and an area that warrants some form of measuring.

Awareness of brand and message

Potential measure of value 14: ***Degree of trust in official statistics***
Consumer based
Rationale: Measure used as a KPI and significant part in the value of official statistics.
Consideration: Could reword to ‘Level of trust’. Trust and/or Trustworthiness could also be broken down into sub elements of trust as this is such a crucial element (such as trust in the office, trust in the chief statistician, trust in the numbers themselves, trust in the independence from politicians…)

Awareness of brand and message

Potential measure of value 15: ***Degree of belief in the political independence of NSO***
Consumer based
Rationale: Crucially important measure for NSOs that can also be very time and sensitive. Has been especially pertinent with the Covid-19 pandemic.
Consideration: Could reword to '***degree believe free from political interference***' and have another measure covering 'awareness' of independence as many people don't even know that official stats are meant to be independent.

Awareness of brand and message

Potential measure of value 16: ***Degree of belief in the political independence of chief statistician***
Rationale:
Consideration:

Awareness of brand and message

Potential measure of value 17: ***Degree of understanding of the remit of the NSO***
Consumer based
Rationale/Consideration: The ‘remit' of the NSO needs to be defined and further need needed. Do we just mean 'do people understand the term official statistics?' or are we trying to get at something more specific, like topics and products?

Awareness of brand and message: new measures proposed in group brainstorming

Potential measure of value 18: ***Share of users who recognize the corporate branding of the NSO***

Consumer based

Rationale/Consideration: Not the most important part but it can become more important as society more often calls for some kind of quality assurance on data. Could possibly be divided into visual identity and mission statement & services.

Subjective measures not currently recommended as indicators of value.

Press on each slide to see the measures

Satisfaction with products and services

User support

Design, communication and metadata

Relevance, responsiveness and innovation

Specific products and services (headline statistics and collections produced by all countries e.g. census, GDP)

Specific products & services – headline stats produced by all (census, GDP)

Not recommended as a measure: ***Share of users who have heard of a specific product***
Consumer based
Rationale: This measure is seen as quite a significant element in a communications strategy and an organization as a whole. Knowing that something exists doesn't tell you about its quality, but it could be a 'prerequisite' for finding it valuable. Knowing about specific products are less likely to be known to a general audience. Some products may attract the interest of specific group based on related domains though the scope of beneficiaries may differ based on the scope of domains covered by the product.
On the downside, it's not easy to make comparable time series as products change over time. Statistics are not always seen as actual products, so could this lead to bias and a perverse incentive? Often these are not based upon statistical samples and satisfied customers may be less likely to respond.
Consideration: There could be sub-components of satisfaction. (i.e. you could be satisfied/dissatisfied with it as a respondent, or as a user, or as a potential user, etc).
Would need to be used alongside qualitative product studies. (As before, arguably could go under one of the previous headings instead of a separate heading for specific products. This one could be under user satisfaction).

Not recommended as a measure: ***Share of users who have used public use microdata***
Rationale:
Consideration:

Not recommended as a measure:***Degree to which users think they or their organizations would benefit from increased availability of anonymized microdata files (or other specific products)***
Rationale: Suggestive in its aim and the description and very similar to one above, could combine
Consideration: Needs redefining

Satisfaction with products and services

Not recommended as a measure: ***Relative importance ratings given by users to quality dimensions (timeliness, accuracy, trustworthiness, national comparability, international comparability, etc.)***
Rationale/Consideration:
This measure could help to understand the concept of quality to better meet user's needs, however, in general, abstract conceptual names of the dimensions are not known to the general audience and their use does set a 'statistician tone', not an 'end user tone'. We need to look beyond internal quality language and be more inclusive with how we express and consider value. Some NSOs have reported that where they have used abstract quality dimensions they have faced unfounded evaluation with a large number of respondents quitting the survey.

Not recommended as a measure: ***Distribution of preferences for different types of access (online, phone, publications, in-person)***
Rationale: It is important to show accessibility whether the website should be usable via mobiles and various devices and important to know users and adapt to their needs and constantly improve. For a product to be valued the customer needs to be able to access it and it can help with supporting the delivery of that. However, the link between this and value is tenuous and will vary depending on the country.
Consideration: An alternative is for 'used' rather than 'preferences'. Covered by another measure - we know which devices are used by looking at web analytics (covered by objective measure).

Not recommended as a measure: ***Distribution on access types (online, phone, in person etc). Preferred devices types used***
Rationale: As above, it can help to inform better meet users needs and might be useful when contemplating the introduction of new modes of dissemination or developing existing ones.
Consideration: Needs redefining

Not recommended as a measure: ***Ratings given to national stats on each of the X dimensions***
Rationale:
Consideration:

User support

Not recommended as a measure: ***Compliance with data protection legislation***
Rationale: It's implicit that NSOs adhere to code of practice. Though compliance should to data protection legislation should be made clear it helps build trust. Furthermore, whilst compliance may reflect on the value of the service as a whole, it does not specifically impact the value of any single official statistic.
Consideration: Doesn't quantify value

Not recommended as a measure: ***Engagement and training with other government departments***
Rationale: This talks to the value of a service. Training is an important coordination instrument and helps to strengthen the coordination role of NSOs. However, it is hard to see a clear quantitative relationship between number of 'trainings' and value. More training might be because more people have expressed confusion about our work, which is a signal that we are doing something wrong, not right! Engagement is vague and engagement and training are two different things.
Consideration: Needs more work in defining and precision.

Not recommended as a measure: ***Average rating from user on how well the NSO is doing***
Rationale: It is important to know user's opinion about the NSO, however, this is not specific enough and too general. It's not clear what activity the user is being asked to rate- 'doing well' what does this mean? A given rating from this measure wouldn't help in anyway to improve, a person might think the NSO is doing great stuff but still they don't use it or engage with it.
Consideration: Needs rewording and further defining and breaking down (could say something like ''how well the NSO is meeting your needs'?

Not recommended as a measure: ***Average rating from user on what NSO could do better***
Rationale: Important to know user's opinion about NSO, however, not specific enough and too general. It's not clear what activity the user is being asked to rate- 'doing well' what does this mean? A given rating from this measure wouldn't help in any way to improve. A person might think the NSO is doing great stuff but still they don't use it or engage with it.
Consideration: Needs rewording and further defining

Not recommended as a measure: ***Average rating of customer support services***
Rationale: This is a KPI for some NSOs as it provides quick feedback that shows the degree of satisfaction with services, without burdening respondents heavily. It is used to assess if meeting user needs though caution as only a small number of actual users expresses the level of their satisfaction.
Consideration: Needs redefining. Make clear what is meant by support services. Implies 'all' customer support services (which could include IT, user problem etc) rather than the value of an official statistic.

Relevance, responsiveness and innovation

Not recommended as a measure: ***Average rating of effectiveness in responding to user needs for statistical topics***
Rationale: What are statistical topics? Where do they stem from? Again, we need to make clear that users' perspective on this is subjective, especially if we are asking them about our responsiveness to ALL users' needs (i.e. in general, rather than specifically their OWN needs). Important to demonstrate user views about responsiveness but might want to supplement with 'objective' measures of responsiveness (some sort of quantitative measures of such response)
Consideration:

Not recommended as a measure: ***Average rating of effectiveness in responding to user needs for different kinds of products***
Rationale: Similar to above. Statistical products are designed for different user groups. Also presumably this comes from user surveys. Some, maybe most, users could have no answer to this as they would have no experience of expressing a need for a new product and seeing it it's responded to or not.
Consideration: Needs redefining.

Not recommended as a measure: ***Average rating of degree of innovation in how NSO works (technologies, methods, data sources)***
Rationale: Similar thoughts to preceding measures. Don't use it as the only measure of innovation. Also don't assume (unless we have evidence to prove it) that innovation necessarily equates to value.
Consideration: Needs redefining.

Not recommended as a measure: ***Average rating of effectiveness of NSO in informing public debate on current issues in country***
Rationale: The indicator seems to be dependent on the levels of an NSO's proactiveness in the public debate. It may be hardly comparable because of heterogeneity of statistical systems across the globe. note that users' perspectives on our effectiveness at informing public debate are only one of many ways to look at it. They are subjective (we might have a huge role but they rate us as having none, or vice versa).
Consideration: Needs redefining.

Design, communication and metadata

Not recommended as a measure: ***Average rating of ease of navigation on website***
Rationale: Very similar to previous one. It is important indicator as it helps to increase the use of statistics. A useful indicator as part of a wider suite of digital metrics to understand the health of an output. Again, talks to the value of the service, not that of any single official statistic.
Consideration: Needs redefining.

Not recommended as a measure:***Average reported ease of finding what they were looking for on website***
Rationale/Consideration: Covered previously.

Measures and methods for monetizing the value of official statistics

Methodologies for monetizing value have become more and more developed in other fields. Such methods have great communicative power, speaking the ‘language of money’ and enabling people to think of value in terms they may understand more readily than with the more abstract terms of some of the other approaches to value measurement. Monetary measures 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 and how they can contribute to income generation, effective use of public resources, etc.

The initial framework for measuring the value of official statistics therefore recommended developing this approach, given that one of the principal aims of measuring value is to communicate this value to users and stakeholders.

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’, measuring value for money becomes complicated.

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.

A number of approaches were proposed in the framework:

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

Monetary measures

Cost-based approaches

Stated preference methods

Revealed preference methods

Monetary measures not currently recommended as indicators of value

Impact assessments

Cost-based approaches

Potential measure of value 1: ***Cost benefit analysis and ratios***
Consumer based.
Rationale: A 'ratio ' can be used to promote the value and awareness of official statistics. Useful in business cases though hard to evidence but a valuable measure if **benefits** are clearly defined. Benefits are what are realized or projected to be realized as a result of producing a product. These can change over time and can depend on stakeholder views.
This is a quantifiable measure that helps to provide different levels of value for different statistics. It provides an opportunity to quantify products and helps to increase the effectiveness and efficiency of value/worth. This approach in general works by tallying up all the costs of producing the output and subtracting that amount from the total projected benefits – this 'value' is then presented as a ratio.
Census case studies from New Zealand, UK and Australia provide excellent examples and can be used as a benchmark. One (of the two) New Zealand case studies was able to derive an eight times return on investment for the Māori community, helping to counter invisibility issues enabling value-added decisions to be made.
Armenia have also undertaken a cost benefit analysis to move from a traditional census to a combined census approach to create a fully register based census. Costs have been compared and initial calculations have shown combined methods reduce the budget by 55%. The Census is due to take place in 2021.

Consideration: The benefits of producing this measure are low cost but a lot of NSOs are not in a position to carry out the approach and would benefit from help to facilitate the method and calculations.
Case studies around the methodology are provided by the UK, New Zealand and Australia and include a combination of techniques that also include willingness to pay, counterfactuals and revealed preference.

Stated preference methods

Potential measure of value 2:*Willingness to pay for official statistics*
Consumer based. .
Rationale: NSOs can often receive requests with a willingness to pay for data, indeed there are bespoke services provided that meet this requirement, however, on the whole official data is financed by governments and free at the point of delivery and there for the public good so trying to gauge its value using this method can be somewhat problematic. Whilst the method is seen as a useful 'pricing' tool in the commercial world, there is an element that could be used as a tool to estimate and set pricing of official products, though a focus on willingness to pay could then risk accusations of revenue seeking and profit making, rather than delivering products to benefit society: Which could lead to losing public good will and support and misinterpreted as an intention of the NSO to start charging for everything. Other issues are around difficulty of determining the number of 'actual' users of respective statistics. It equally reflects an ability to pay and therefore not an accurate measure of value but affordability
Consideration: Some research explored 'stated preference' and 'revealed preference' techniques but did not come up with any recommendations. However, it is suggested as an area to be retained as there is potential for calculating willingness to pay by looking at using this in combination with other of methods. A pilot study undertaken in the UK has looked at developing an innovative approach using conjoint analysis methodology.
Conjoint analysis technique is widely used in marketing studies and is a survey-based approach that asks respondents to "consider jointly" their preference between a number of products described by underlying attributes. This approach attempts to quantify value by using a revealed preference approach (see below) and derives willingness to pay from the underlying attributes. Therefore, the approach used in conjunction with other methods has the potential to show value. A second phase is currently being planned and as a result of the work, other work is being developed.

Revealed preference methods

Potential measure of value 3: ***Conjoint analysis - experimental approach***
Consumer based.
Rationale: A revealed preference approach called 'conjoint analysis' is a technique that is widely used in marketing studies and is a survey-based approach that asks respondents to "consider jointly" their preference between a number of products described by underlying attributes which is based on utility theory (a concept used to model value or worth). A pilot study in the UK undertaken by ONS has attempted to look at the potential of this to inform on the value of official statistics. Using the approach, it was possible to calculate 'willingness to pay' which was derived from the underlying attributes.
The approach takes attributes and asks a series of questions to respondents who each time choose their preference at varying levels. These questions are typically called "trade-off" questions where attributes and levels are "considered jointly". These preferences can be used to infer relative "utilities" that represent the relative value an individual has for each attribute and its levels. Adding these utilities for each attribute leads to the utility for the overall product. For example, a television might be described using three attributes: "brand", "screen-size" and "price". A specific television might have a brand level of "Panasonic", a screen-size level of 54 inches and a price level of £1,000. The utility a customer has for this television will simply be the sum of their utilities for each appropriate level of these attributes. The main use of relative utilities is to derive "preference shares" where each respondent's preference for a specific product, among a set of products, is modelled. With appropriate weighting it is possible to create preference shares representing the population. With certain assumptions these preference shares are used to denote actual market shares that should be expected given a set of products defining the market.
Consideration: Methods such as conjoint analysis have been widely used in marketing, most typically in product positioning studies where markets are well defined, so this is experimental. The model works well and other studies are now being carried as a result of this work.
A measure in this area needs further defining and development and the Taskforce welcomes further work to explore in this area.

Impact assessments

Potential measure of value 3: ***Impact assessment***
Consumer based.
Rationale: ***Impact and evaluation are crucially important and is where the value of our work is realized.***
Consideration: Measures around impact should cover all measures around monetary, subjective and objective.
The framework reported that 'statisticians have not been good at assessing and measuring such impact' and that 'this has hampered ability to promote the value of official statistics'. We need useful insights on how actual users use our data and how this influences their decisions and results not only to promote our value but for us to know what that value is and for us to improve on that value. This needs to be a key area of focus for value metrics, though it is difficult to measure as impacts can emerge on a wide scale, NSOs should be looking to seek investment in this. This is also useful information for users.
There is a need for case studies and we need to share and learn and stay abreast of best practices on how best to evaluate impact across the statistical community.
An Impact Framework new approach proposed in group brainstorming

Monetary measures not currently recommended as indicators of value.

Press on each slide to see the measures

Cost based approaches

Market equivalent pricing

Stated preference methods

Cost based approaches

Not recommended as a measure: ***Total actual costs of producing official statistics***
Rationale: Budget allocation is not a performance metric, it's more of a threshold. Essential for consideration in planning cycle for production. Costs cover a wide range of overheads and would need to be broken down. Base cost may be better though might exclude major programmes (like Census) though organizations may differ.
Consideration: As a public good, the total cost of operating as an NSO is already in the public domain. Though it would be beneficial to show the exact cost of producing official statistics and the elements used in calculations, budgetary allocations are not the best way to do that and trying to grasp human capacity involvement is also rough to calculate (E.g. approaching this only from data based on time spent is not sufficient). Service outcome cost could possibly be a more effective measure.

Not recommended as a measure: ***Cost (per capita per year as % of GDP?) of producing official statistics***
Rationale: This measure suggests an inclusion to measure cost by 'person'. Similar to the above measure, a simple transformation rather than a measure in its own right (the cost of the above divided by the population).
Consideration: Looking at 'per capita' has better communication power than the absolute figure and maybe more comparable against other organizations. NSOs could produce 'time spent' on activities (most use codes) which could be broken down by per capita, so rough estimates could be provided but not clear what additional value this would have. As above, approaching this from only data based on time spent is not sufficient. Few people would have any sense of whether the total operating cost of an NSO, is a lot or a little when you tell them the number.

Not recommended as a measure: ***Cost (per capita per year as % of GDP?) of specific products, e.g. census***
Rationale/Consideration: This measure is moving to a 'unit cost' approach and is only really feasible with big stand-alone products or projects (E.g. Census, Labour Force Surveys etc). However, it is hard to separate out costs of most products as difficult to quantify the time need and labour costs as some products are interdependent (E.g. surveys whose sampling frame comes from census, topic-based stats drawn from multiple sources.) NSOs usually keep track of their most important products so depending on the approach of an NSO, it may or may not be comparable.

Not recommended as a measure: ***Time trends in each of above, adjusted for price changes/not adjusted***
Rationale/Consideration: This proposed approach suffers from the same problems as above. The measurement around an output has to have regard for how the outcome has been improved and it is hard to capture a quality change of outputs. The Atkinson Review (in UK) made some recommendations around measuring the 'value added' of a public service output which was based on user need. There has been some exploratory work looking at measuring productivity of a statistical organization, described more as a *Time series of unit values'* (case study provided)

Market equivalent pricing

Not recommended as a measure: ***Market equivalent pricing***
Rationale: Gauging the pricing of statistical products in the market would provide a relevant reference for the value of official statistics, NSOs are already competing for customers, however, this is difficult when 'official statistics are a public good and free at the point of delivery.
If NSOs are to remain relevant and secure future funding, comparative pricing provides a key performance metric in an increasingly commercial environment. Some key products should be selected for which a market equivalent could be defined, but we should not aim for a total comparison.
Consideration: This method might be more useable for services.

Stated preference methods

Not recommended as a measure: ***Willingness to accept compensation for the removal of access to official statistics***

Rationale: There is little knowledge of NSOs looking to develop work in this area and openly communicating the very idea to users is viewed as risky on a number of levels with many implications. This approach does little to promote a positive perception of NSOs value.

Consideration: With the need for funding of services and resources, there does need to be meaningful conversations about a value proposition of data and the consequences of their removal.