

Trustworthiness of statistics in a data-abundant age

By: Ed Humpherson

We live, we are told, in a data age. This is said to be an era of almost untold possibilities from the analysis and application of abundant data. Data are everywhere, from sensors and satellites, smart phones and social media, and the organisations that access this abundant, non-exhaustible resource are and will be, we hear, the corporate titans and the drivers of growth and productivity.

The sense of data abundance, and its liberating potential for humanity, calls to mind the optimism of the first wave of atomic energy stations after the Second World War, ushering in an era of huge potential supported by non-exhaustible energy that was “too cheap to meter”.

The “living in a data age” also spawns a series of anxieties: perhaps the nuclear waste of the data world. We hear of fake news and post truth. This dystopian vision of the data age comes with it a sense of threat to professions, including broadcast and print journalism; expertise and science; and my own profession of statistics. And even more profoundly, it has triggered a new set of debates and concerns about the ownership of data, the rights of the individual and the meanings of confidentiality, consent and privacy in this connected, data-abundant world.

Waves of technological optimism have often had a mixed record of realisation: after all, nuclear power stations were indeed built, but it would be hard to argue that they completely fulfilled all the expectations that people had for them.

This article is about one aspect of data – official statistics. It describes how my organisation, the Office for Statistics Regulation, thinks about statistics in the data age and responds both to the opportunities and anxieties of this age. It explains how we have come to place trustworthiness, based on the ideas of Onora O’Neill, at the heart of our work.

The Office for Statistics Regulation

The phrase “official statistics” refers to statistics published by government bodies. Official statistics have a long history – the UK’s first population census took place in 1801. Official statistics encompass a vast range of topics. They include statistics on: the size and structure of the economy; labour markets; the size and location of the population and its demographic characteristics; the health of citizens and the performance of the health system; crime and the criminal justice system; the environment; the transport system; and many other areas of society.

It is one of the core responsibilities of any government to publish information on the nature of the society for which it is responsible. This information, captured in statistical outputs and releases, enables citizens, communities and businesses to form judgements and make decisions. The decisions range from democratic choices (for example, who to vote for) to community choices (for example, do we need a new school) and economic decisions (for example, whether and where a business should expand).

The UK has a governance regime with independent scrutiny and assessment of official statistics, based on the Statistics and Registration Services Act 2007¹. The Act gave the UK Statistics Authority the power to set a Code of Practice, assess compliance with the Code, and monitor how far statistics are serving the public interest. The regulatory role is carried out by the Authority’s Office for Statistics Regulation.

The Office for Statistics Regulation’s purpose statement is:

Statistics are an essential public asset. We enhance public confidence in the trustworthiness, quality and value of statistics produced by government. We set the standards they must meet. We celebrate when the standards are met, and

¹ https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-statisticsandregistrationserviceact2007_tcm97-18260.pdf

challenge publicly when they are not. We are champions of relevant statistics in a changing public world.²

The Code of Practice for Statistics

The Code of Practice for Statistics is central to this model and has been at the heart of setting high standards for official statistics. This first edition of the Code was published by the UK Statistics Authority in January 2009 as one of the Statistics Authority's statutory responsibilities under the UK's Statistics and Registration Service Act³.

Since then, the Code has been adopted widely across the UK public sector, within over 150 bodies who publish official statistics. The Code sets out the standards the Government must adhere to when it publishes official statistics. The Code also forms the basis of regulatory judgements – whether statistics meet appropriate standards (if they do they are designated “National Statistics”); and whether the Government's publication and use of statistics breaches any element of the Code.

In 2015, we decided to update the Code. It was through this review process that we came to realise the central role played by trustworthiness.

Value: from a mechanical to a social view of statistics

We wanted to update the Code to reflect the new data environment. This environment is one of abundance and opportunity, but also one of greater anxiety – around post truth, privacy and misuse of statistics. All of these factors, both positive and negative, have an impact on the publication of official statistics by government bodies.

What is the problem we were trying to solve? First, we knew that we wanted to move from a mechanical to a social view of statistics. This meant we needed to recognise the social value of the Government's role in publishing statistics. This emphasis on value sends a very important signal: it is not enough to produce high quality statistics in a well-managed way. Statistics must also be useful to users. They must be accessible, meet user needs, provide insight and answer user questions. This raises the stakes for

² https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/04/Business_Plan_201819.pdf

³ Section of the Statistics and Registration Services Act 2007, Section 10

government statisticians – they cannot simply publish a set of numbers and think their job is done.

This is ever-more important in the increasingly crowded space of data abundance. In this crowded space statistics must provide insight, by connecting to the questions and interests of their audiences. The focus on value de-emphasises the mechanical and prioritises the social communication and social impact aspects of statistics – a profound change in perspective.

Second, we needed to bolster the Code in an environment of concerns about post truth. The features of the post-truth world noted above came together to create the case for change. We became concerned that widespread debate about the reliability of political statements might infect public confidence in official statistics. We wanted to issue a strong response that defended the value of official statistics. We also needed to recognise the impact of data abundance, and to reflect much greater concern about the reliability of organisations who hold private information on individuals.

In short, we wanted to update the Code for the age of data abundance.

The walled garden

There was another, more endogenous, factor: the culture created by the first version of the Code itself. While this Code had been very successful in setting standards across the UK public sector, and in providing the basis to protect statistics from political interference, in some ways it may have done its job too well. The first version of the Code came to be seen as a barrier to innovation in some Government departments. It led us to lament that official statistics were starting to resemble a “walled garden” – beautiful, well looked after, but elitist and cut off from the world outside, and not consistent with public confidence. This elitist walled garden could not survive in a data-abundant world, and the Code needed to reflect that.

Our approach: focus on trustworthiness

In responding to these factors, we were not attracted to a focus on trust as an organizing concept. Why was “trust” not appealing to us?

Trust is reflected in people's stated beliefs. Most of the assertions that societal trust is declining is based on surveys, which ask people who or what they trust and do not trust. These sorts of surveys typically show a low rating to politicians and Government and to journalists; and a higher rating to scientists.

There are three problems with this trust-focused approach for the Code of Practice for Statistics. First, there's a well-known difference between what people state and what their behaviour reveals. They may claim to trust someone or something and then act as if they do not – and vice versa. For us, it would be a problem if our Code required Government to measure and increase levels of *stated* trust in statistics if stated levels of trust were not associated with people using statistics in their decision-making (their revealed preference). Focusing on the stated trust might be to address the wrong target. Second, the stated belief of trust ("do you trust government statistics?") may reflect a very wide range of factors: political leaning; perceptions of individuals; influences and so on. It would be hard, if not impossible, for our Code to require Government to target all these factors when publishing statistics.

The human factor

Instead of trust, we wanted to put public confidence to *use and rely on statistics* at the heart of the new Code.

But what does this mean in practice? We started to identify the general factors that underpinned confidence in any exchange of information, and it gradually became clearer that in an exchange of information between human beings, the quality of the information itself is only one of the things that supports confidence. The identity and reliability of the provider of the information is equally important, and in many situations, precedes any focus on the information's quality: a person is hardly likely to inspect in detail the quality of a statement made by someone who is known to be a liar.

In turning away from "trust" as a unit of analysis and focusing instead on what we came to think of as the human factor, we realised that we were following in the footsteps of the great philosopher, Onora O'Neill. She had laid out the basis for thinking about trust

and trustworthiness in a series of lectures and articles over the last decade. Her ideas provided us with the key to unlock these questions and to refresh the Code of Practice.

Onora O'Neill's Reith lectures, articles and TED talks starts with the same insight that we were concerned with: "trust" is an unwieldy and unhelpful concept, and particularly problematic because people appear to place *active* trust in organisations and individuals they profess not to trust. Her thinking moved on from this starting point. Something else, O'Neill argues, is clearly going on. She concludes that the aim for an organisation should not be to pursue trust, since it has an inchoate quality, but instead to seek to demonstrate trustworthiness. This demonstration involves providing evidence that demonstrates integrity, honesty and reliability in a form that allows others to check whether the organisation is indeed living up to these values. David Spiegelhalter, President of the Royal Statistical Society, put it succinctly thus: no-one can just expect to be trusted. An organisation must provide testable evidence to demonstrate that they have the interests of the public at heart, by demonstrating competence, honesty and openness. Organisations, therefore, should not pursue trust, but demonstrate that they are trustworthy.

For us this was an absolutely crucial insight, and one we might not have arrived at on our own.

Our task was to take this concept and convert it into a set of practices that could be articulated in the Code; practices which could be easily checked by the public, by us as regulators or by the organisation itself.

To achieve this, we started by identifying commitments that an organisation publishing statistics could make. We were looking for commitments that would truly demonstrate trustworthiness even if those commitments were hard to make and harder to fulfil or may appear to run against the short-run interests of the organisation.

This then led us to a series of expectations of organisations that publish statistics, including:

- Government organisations should pre-announce when statistics will be published, and stick to that date. This is a commitment to sometimes publish statistics on a day that may turn out to be politically inconvenient – that may not fit with the news grids beloved of Government political communicators.
- The relevant Chief Statistician, not a Minister, must have the final say on the form and content of statistics published by government Departments
- The contact name on a statistical output should be the responsible statistician, not a press officer

These expectations, and a range of others, all relate to trustworthiness, and we organised them as the first “pillar” of the new Code of Practice. The other pillars are:

- The quality of the information
- The value of the information to the user

These pillars are intended to endure over time, and to be relevant and resilient to the data abundant age. Indeed, our view is that they become increasingly important in showing the grounds for confident use of and reliance on statistics: after all, in a crowded space, it may be that demonstrating provenance becomes ever more important.

Through an iterative process of consultation, we refined our definition of these three pillars. In the final text, we described them thus:

- **Trustworthiness:** Trustworthiness is about the processes, people and systems of organisations. An organisation must provide testable evidence to demonstrate that they have the interests of the public at heart, by demonstrating competence, honesty and openness. The practices under the Trustworthiness pillar set out the key commitments that must be made to support independent statistics production.
- **Quality** is about the data, and how they are processed into statistics. The Code recognises that process and system integrity is not, on its own, enough to guarantee worthwhile statistics. The statistics must be the best available estimate of what they aim to measure, and should not mislead.

To achieve this the data must be relevant, the methods must be sound and the assurance around the outputs must be clear. These aspects of statistical production are at the heart of the practices in the Quality pillar.

- **Value** follows the emphasis in the UN Fundamental Principles of Official Statistics on statistics that “meet the test of practical utility”⁴. It defines what statistics must provide for the public. This includes a coherent picture, a focus on users, an emphasis on what questions the statistics answer and a focus on innovation as the world changes. Trustworthy processes to create high-quality data may not be useful to the public if the statistics are not accessible, do not address key questions, are inefficiently produced, and do not add value or provide insight.

These pillars are conceptually distinct. But they support each other. A producer of official statistics is more likely to be perceived as trustworthy where the data they provide are clearly of high quality. High-quality statistics are much more likely to provide useful answers to key questions than lower-quality statistics.

We published the new Code on 22 February 2018⁵. It was built around the three pillars, supported by 14 principles (each supported by detailed practices), as shown in Figure 1:

Figure 1: The three pillars and 14 principles of the UK’s new Code of Practice



⁴ <https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf>

⁵ Code of Practice for Statistics, UK Statistics Authority, February 2018, (<https://www.statisticsauthority.gov.uk/code-of-practice/>)

Conclusion

In the data abundant world, it is a delusion to think that data is like the 1950s rhetoric about electricity from nuclear power stations – easy, abundant, too cheap to meter. Like nuclear power, there are social and economic impacts and there are also downsides and risks to manage. These opportunities and downsides apply in particular to the subset of the data world that is represented by official statistics published by government, and left unchecked could undermine the usefulness and integrity of this essential public information.

The revised Code is founded on the view that provenance matters in this data abundant world. Official statistics are never “just numbers”. They emerge from an organisational context. To command confidence, it’s important that as much attention is paid to the nature of the organisational context as it is to the technical characteristics of the numbers. For us trustworthiness is the essential starting point.

But the concluding message of this article is not really about official statistics, nor the Code of Practice, nor even the “data age”. It’s about the power of ideas - a celebration of the benefits of paying attention to the lucidity and clarity of a trail-blazing philosopher.