Statistical Journal of the IAOS 37 (2021) 85–105
DOI 10.3233/SJI-200667
IOS Press

The manipulation of official statistics as corruption and ways of understanding it

Andreas V Georgiou[a,b]
[a]*Visiting Lecturer and Visiting Scholar, Amherst College, USA*
[b]*Former President (2010–2015), Hellenic Statistical Authority, Greece*
E-mail: avgeorgiou83@amherst.edu

Abstract. We argue that the manipulation of official statistics is corruption and indeed it is grand corruption and political corruption. We note the steps that have been taken in recent decades, which inter alia help make corruption in official statistics more difficult, but argue that instances of corruption persist. To decisively address the problem, it is important to have an analytical tool for understanding it. In this context, we provide a typology of the various broad manifestations or phenomena of corruption in official statistics. We then proceed to identify the process or mechanism that gives rise to these manifestations/phenomena, the elements/component parts of this process/mechanism and the ways these parts interact. This constitutes the schematic model that we propose for understanding corruption in official statistics. In this context, we provide a discussion of the nature of the individual elements/component parts of the process of corruption of official statistics, i.e., we discuss the drivers of the phenomena, the enabling conditions of the phenomena, the modalities and methods used to arrive at the phenomena, and the vectors or agents that execute/propagate the phenomena. We believe all cases of corruption in official statistics can be analyzed using this schematic model. The benefit of having such a model is that it enables one to identify what institutional or legal setting, action, institution or person presents a problem, vulnerability or source of risk in a given system of production of a specific official statistical product in a given country, and address it. The model can also inform a discussion about what needs to be changed at the level of international/supranational arrangements, whether concerning institutional settings, processes, or legal and ethical frameworks, affecting the production of official statistics.

Keywords: Official statistics, corruption, manipulation, principles, ethics

1. Introduction

In this paper we aim to establish that the manipulation of official statistics is corruption and to offer an analytical framework for understanding it. This is an important step in effectively addressing corruption in official statistics.

In Section 2 of the paper, we explain that the manipulation of statistics is grand corruption and political corruption as it is very often a manipulation of the statistical processes and statistical institutions at the instruction of or in anticipation of the interests of politicians/policy makers and those they represent, and this is done so as to achieve a false picture of economic, social and environmental reality with a view to distorting policies and state functions. We discuss the ulterior motives in terms of distortion of allocation of economic resources of the state and of the economy as well as in terms of misappropriation of political power.

In Section 3 we discuss steps that have been taken in the area of official statistics in recent decades that inter alia help make corruption in official statistics more difficult. We take note of international and national codes of statistical ethics, supranational and national legislation, international statistical standards, and supranational and national institutions and processes. We argue that, while significant, these steps need to be supplemented by additional ones that would tackle persisting enabling conditions and modalities of corruption in official statistics.

In Section 4 we categorize the manifestations or phenomena of corruption in official statistics into about a dozen categories and allocate them under three broad

types (A, B and C) of phenomena of manipulation of official statistics, from outright 'crude' to most 'sophisticated'. We also discuss how phenomena of manipulation of official statistics can also be thought of in terms of their place in the timeline of the 'life' of an official statistics product. We further explore whether any state of economic development is immune to manipulation of official statistics and whether different levels of economic development are associated in some way with certain of the types of manipulation of official statistics we have identified.

In Section 5 we argue that the manifestations/phenomena of corruption in official statistics can be fully understood only by identifying the process or mechanism that gives rise to such phenomena, the elements or component parts of this process/mechanism, and the ways these parts affect each other. This is a schematic model that we offer for understanding corruption in official statistics. All cases of manipulation of official statistics can be analyzed using this model. In this context, in Section 5 we provide a discussion of the nature of the individual elements/component parts of the process of corruption in official statistics that lead to the manifestations/phenomena of corruption, i.e., we discuss the drivers of the phenomena; the enabling conditions of the phenomena; the modalities and methods used to arrive at the phenomena; and the vectors or agents that execute/propagate the phenomena. We argue that the benefit of having such a model is that it enables one to identify what institutional setting, action, person or institution presents a problem, vulnerability or source of risk in a given system of production of a specific official statistical product in a given country, and address it. Moreover, we submit that the model can also inform a discussion about what needs to be changed at the level of international/supranational arrangements – whether concerning institutional settings, processes, or legal and ethical frameworks – affecting the production of official statistics.

In Section 6 we summarize and offer some concluding remarks.

2. The manipulation of official statistics is corruption

The manipulation of official statistics is corruption and indeed it is grand corruption and political corruption.

Corruption can be defined as "dishonest or illegal behavior especially by powerful people (such as government officials or police officers) and as inducement to wrong by improper or unlawful means (such as bribery)" [1]. Grand corruption is said to consist "of acts committed at a high level of government that distort policies or the central functioning of the state, enabling leaders to benefit at the expense of the public good" [2]. Political corruption can be defined as "a manipulation of policies, institutions and rules of procedure in the allocation of resources and financing by political decision makers, who abuse their position to sustain their power, status and wealth" [2].

The manipulation of official statistics so as to distort the picture of reality that official statistics are supposed to record fits well the broad definition of "corruption".[1] It surely is dishonest **or** illegal behavior[2] by people in positions of power, such as the official statisticians themselves or the people – virtually always politicians/policy makers – on the basis of whose instructions or subtler inducement, or in anticipation of whose interest, statisticians manipulate official statistics.

The manipulation of official statistics is "political corruption" and "grand corruption" as it is very often a manipulation of the statistical procedures and statistical institutions at the instruction, or in anticipation of the interests, of politicians/policy makers and those they represent. And this is done so as to achieve a false picture of economic, social and environmental reality with a view to distorting policies and state functions. The ulterior motive is to allocate the economic resources of the state and of the economy in general in a way that these resources would not have been allocated otherwise; and this for the benefit of political/policy leaders and those they represent and at the expense of the interests of the broad public.

Types of manipulation of official statistics that involve acts directly undertaken by politicians/policy makers (e.g., as controllers of upstream data sources

[1] An observer might argue that defining corruption one necessarily has to take as a starting point and build on the United Nations Convention Against Corruption, something which this paper does not do. However, the UN Convention Against Corruption lacks even a definition of corruption and is confined to a limited set of specific acts, such as bribery, embezzlement, money laundering and obstruction of justice (something that probably points to its creation as a politically palatable construct to the more than 180 signatory governments). The present paper, instead, is explicit about the definition of corruption and then proceeds on that basis to construct a comprehensive system of conceptual elements, without artificial delimitations, tailored to official statistics and their specific contexts.

[2] It should be noted that for a behavior to be understood as corruption it would not have to be treated as illegal in existing laws. This applies also to corruption in official statistics.

necessary for the production of official statistics or as actors setting the standards to be followed in the production of official statistics) also fit the definition of "political corruption" and "grand corruption". We also submit that manipulation of official statistics by abusing institutions, policies and rules of procedures to delegitimize official statistics with a view to undermining their credibility also fits the definition of corruption.

The above noted misallocation of economic resources when there is manipulation of official statistics is most often accompanied by the misallocation of political power. The political, the political decision makers who are involved, directly or by proxy, in this corruption abuse their position to sustain and increase their political power and status. They often – by the same token – also get to sustain and increase the resources available to them, which helps them further skew the political playing field in their favor, reinforcing the misappropriation and misallocation of political power. Therefore, corruption in the form of manipulation of official statistics leads to a degradation of democracy and other forms of government.[3] Corruption in the form of manipulation of official statistics generally leads to and accompanies a degradation of social contracts.[4]

It is important to remind the reader and highlight the fact that *corruption is not the usual mode of official statistics; all national statistical systems do* ***not*** *suffer all the time from such phenomena*. Yet, corruption is possible, in isolated or pervasive form, in *any* statistical system *when the conditions are right*. And only by recognizing this, measures can be taken to maintain the integrity of official statistics (see Box 1 for a small sample of cases from around the world).

Some well-meaning stakeholders and observers may object to the characterization of the various types of manipulation of official statistics discussed in this paper as corruption. For example, some may feel that certain manipulations should be characterized as 'indirect political interference' and not as corruption. Some may want to delimit corruption in official statistics only to phenomena that are already criminalized in international and national legal frameworks. Finally, others may feel that some of the phenomena described in the present paper should not even be characterized as manipulations of official statistics at all – much less corruption – but instead be referred to as phenomena that 'just undermine the impartiality of official statistics.'

In our view, first, such approaches would overlook the fact that the phenomena described in this paper fit widely accepted definitions of corruption, political corruption and grand corruption as argued above. Second, such approaches would de facto tend to accommodate the *status quo* in politics and society vis-à-vis the integrity of official statistics. Following these approaches would not allow statisticians, policy makers and the broader society alike to push strongly enough to identify and address phenomena of manipulation – indeed corruption – in official statistics that continue to arise in today's world and often have staggering consequences for the welfare of individual nations and, in some cases, of the global community. In writing the present paper we made the decision to 'call things by their name' and not to soften the discussion by choosing to use what may be considered euphemisms that take away from the stark reality of the mechanisms behind manipulated statistics, which often have massive economic/social/environmental consequences.

3. Steps taken in recent decades that inter alia help make corruption in official statistics more difficult

In the area of statistical production that is official statistics the problem of potential manipulation has in recent decades been recognized as a threat along with other risks to the quality of official statistics. There have thus been efforts to create institutional environments and processes aimed at safeguarding the use of statistical principles and ethics in the production of this type of statistics. They inter alia include international and national codes of statistical ethics, supranational and national legislation, international statistical standards, and supranational and national institutions and processes. Below we discuss some of these major institutional modalities that have emerged in recent decades.

In respect of principles and ethics in the area of official statistics, the United Nations Fundamental Principles of Official Statistics (UNFP) [10] have been a major reference point for the practice of official statistics since 1994, when they were adopted by the United Nations Statistical Commission (UNSC). In 2014, the UNFP were enhanced with a useful preamble and were adopted by the General Assembly of the United Nations.

[3] Under other forms of government could be included those that claim to operate in a rational and benevolent way in the interest of the people or society, such as autocracies.

[4] Social contract is "an actual or hypothetical compact, or agreement, between the ruled and their rulers, defining the rights and duties of each" [3].

Box 1. A small sample of instances of manipulation of official statistics

Argentina – The misreporting of the consumer price index during 2007–2016. The March 2016 Report of the Independent Evaluation Office of the International Monetary Fund (IMF) [4] stated: "In January 2007, the Argentine government changed the personnel in charge of producing the consumer price index (CPI) at the National Statistics and Census Institute (INDEC). Concerns about the integrity of the CPI started to be voiced soon thereafter. These concerns led several private sector entities and provincial governments to compute their own indices that showed consumer prices growing at significantly higher rates than those reported by INDEC. . . . The apparent underreporting of CPI has implications for other key variables of significant importance for economic analysis. Inasmuch as the official CPI enters their calculation, measures of poverty or of the real effective exchange rate would be underestimated while the real growth of the economy would be overestimated. Underreporting would also have notable financial implications given Argentina's issuance of inflation-linked peso bonds."

Greece – The misreporting of certain interest payments in Greek public finance statistics prior to 2010. The Annex to the January 2010 Eurostat report on Greek government deficit and debt Statistics [5] stated: ". . . the revision of interest recorded between April and October 2009 turned out being a case of misreporting of a figure rather than a methodological issue . . . the Public Debt Directorate of the GAO [General Accounting Office of the Ministry of Finance] communicated already in April 2009 the correct figure to its hierarchy, but that for some reason the figure communicated to the National Statistical Service of Greece was changed by the hierarchy of the GAO to 45 mn € from 495 mn €, reducing the deficit for 2008 . . . "

United States – The 2005 suppression of statistics on traffic stops and racial profiling. According to press reports [6], political supervisors within the US Department of Justice of which the Bureau of Justice Statistics (BJS) is part ordered the head of BJS to delete certain references on the more aggressive police treatment of black and Hispanic drivers, compared to white drivers, in traffic stops, from a news release. The head of the BJS refused to delete the racial references, arguing to his supervisors that the omissions would make the public announcement incomplete and misleading. Instead, the Justice Department opted not to issue the news release on the findings and posted the report online (which could not be found by the journalists), thus effectively suppressing to some significant extent indefinitely the statistical results. The head of BJS was removed and he was demoted to a lower position in another institution for the last six months of his civil service career.

Central African Republic – The misreporting of census data in the 1970s. The then leader of the country ordered the increase of the total population figure in the statistical release of the results of the population census. Accounts of the events [7] have it that the total population figure was increased by 1 million in a country that the census had found that it had about 2 million people. The latter was deemed by the then political leader to not conform to his intended image and role for the country.

European Union – The persistence of a standard precluding the recording of trade credits/accounts payable as 'Maastricht' government debt. The government finance statistics standard used in the EU's Excessive Deficit Procedure specifies that trade credits and other accounts payable are not classified as government debt, i.e., as 'Maastricht debt' of EU Member states. This is an instance of a standard, where its application produces biased statistical results. In 2013 official statisticians wanted to change the standard so as to more accurately reflect government indebtedness, consistently with the European Statistics Code of Practice which under principle 12 requires that "statistics accurately and reliably portray reality". "However", according to one press account [8] "the representatives of Treasuries of EU member states refused to modify the EU legal text. The effect can be impressive: In the case of Belgium, payable obligations exceeded 5 billion euro since 2014, while before 2006, they oscillated between 1 and 1.5 billion. Thus, the debt may be understated further by 3–4 billion euro, and this is effected by simply withholding payments to suppliers and not recording it as debt."

Jordan – The misreporting of GDP and public finance statistics in 1996–97. A 2016 Background Paper of the Independent Evaluation Office of the IMF [9] stated: "During Jordan's 1996–98 extended arrangement from the Fund, the authorities provided [IMF] staff with erroneous information on national accounts and fiscal data. These data were subsequently revised in mid-1998. The revisions indicated that GDP growth had been substantially lower than initially reported – around 1 percent per year instead of 5 percent – . . . the budget deficits in both 1996 and 1997 had actually been substantially higher than reported." IMF staff reports from that time noted that "these revisions made clear that the data set that was available until May 1998 had portrayed a fundamentally distorted picture of the state of the Jordanian economy and performance under the extended arrangement."

Currently the implementation of the UNFP is monitored via a survey conducted every ten years by the United Nations Statistical Division (UNSD). The survey sent to country authorities is processed by the UNSD and its aggregated results are presented at the UNSC. The current system of monitoring implementation of the UNFP is thus based on potentially incomplete or potentially biased evidence provided through self-assessment.

Moreover, while the UNFP have been useful as a reference point, it became clear overtime that more was needed in terms of coverage and specificity of statistical principles/ethics, as well as detail regarding best practices. For example, more was needed regarding naming and directly addressing the statistical principle of 'professional independence' in the production of official statistics, about which the UNFP are completely silent.

In this context, and spurred by the crisis in European Union official statistics – European statistics – in 2004, which was triggered by the uncovering of the misreporting of Greece's deficit and debt statistics used for Greece to adopt the euro, the European Statistical System adopted in 2005 the European Statistics Code of Practice (and amended/improved it twice since then) [11]. This was a significant evolution of the UNFP in terms of coverage, specificity as well as boldness with regard to statistical principles/ethics for official statistics production. For example, principle 1 of this Code of Practice is about 'professional independence' and goes to some length in providing 8 specific criteria that have to be met to effectively support professional independence. Nevertheless, a close reading of the European Statistics Code of Practice reveals that in some areas discounts vis-à-vis available best practices were

adopted and maintained over time for seemingly political reasons. For example, in the area of privileged pre-release access to official statistics, policymakers and others are still allowed to potentially have such access. Codes of practice similar to the European Statistics Code of Practice were also adopted by other national and supranational entities around the world, such as the UK's Code of Practice for Statistics, which was adopted in 2009 [12].

Regarding additional international codifications of principles of official statistics, there were some notable developments that built on and were inspired by the UNFP, such as the Principles Governing International Statistical Activities, which were endorsed by the Committee for the Coordination of Statistical Activities and which concern the Chief Statisticians or coordinators of statistical activities of United Nations agencies and related organizations.

Many countries also adopted statistical quality assurance frameworks (SQAFs) to promote the use of statistical principles and best practices. To support the development of SQAFs the UNSD produced a generic National Quality Assurance Framework (NQAF), that was endorsed in 2012 by the UNSC and comprised a Template and Guidelines, and which was designed to assist countries in developing their particular SQAFs so that they can engage in *self-assessment*. Subsequently, in 2019, the template and guidelines were replaced by the United Nations National Quality Assurance Frameworks Manual for Official Statistics (UN NQAF Manual), which was also adopted by the UNSC [13].

Improved legislation has also been adopted in many countries in recent decades that provides for the production of official statistics. These modern national statistical laws can inter alia provide the legal basis and the mandate for official statistical production; define the national statistical system (NSS); provide for the institutional setting of the national statistical office and its leadership, provide for the governance institutions of the NSS, and set the responsibilities of the national statistics office and other producers of official statistics. Very importantly, national statistical laws often also provide for the statistical principles/ethics that are to be observed regarding official statistics production. Statistical laws can do this by direct discussion of such statistical principles/ethics in the law itself and/or by reference in the law to existing international or national codes of such principles/ethics, such as those discussed above.

In a supranational setting, for example in the case of the EU, there can also exist such a statistical law. In the case of the EU, its statistical law is Regulation 223/2009 of the European Parliament and of the Council, which was amended and improved in relevant-to-the-present-discussion areas in 2015 [14]. According to the precedence principle, this European law is superior to the national statistical laws of member states of the EU and should be observed in every member state of the Union.

There may also be specialized statistical legislation at the national or supranational level that provides for specific areas in official statistics. For example, in the EU there is a large body of sectoral statistical legislation that provides rules for the compilation of official statistics in specific statistical domains. Such legislation elevates to law methodological standards for the technical aspects of the production of statistics. Thus, it can be instrumental in promoting the production of accurate and harmonized statistics and, by the same token, help make some types of manipulation of statistics more difficult. For example, Regulation (EU) No 549/2013 of the European Parliament and of the Council on the European system of national and regional accounts in the European Union provides the rules for the compilation of national accounts in EU member states and it is an important reference point that can help defend against some types of manipulation of national accounts [15]. Beyond those laws, in the EU there are manuals and guidelines that are provided by the statistical office of the EU, Eurostat, to deal with specific statistical compilation issues and have an enhanced legal status. They also serve similar purposes, including in helping defend against some manipulations of official statistics.

There are international statistical standards set by the international community to be followed in the compilation and dissemination of official statistics in various domains. These international standards provide important frameworks within which supranational and national statistical laws and manuals are developed. For example, the current System of National Accounts, the 2008 SNA [16] adopted by the UNSC, constitutes the framework within which the EU has developed and adopted the European System of Accounts (ESA 2010) in the form of the above noted Regulation (EU) No 549/2013. An example of a dissemination standard for some specific macroeconomic statistics is the Special Data Dissemination Standard of the International Monetary Fund (IMF),[5] which specifies periodicity and timeliness standards for these official statistics [17]. As

[5] There are also additional IMF dissemination standards, such as the e-GDDS and the SDDS Plus [17].

indicated above, international standards – when they are adequate – are instrumental in promoting accuracy and harmonization in the production and dissemination of official statistics, and thus help make more difficult the manipulation of official statistics.

There can also exist institutions and processes to provide (some) oversight of whether the rules of production of official statistics are followed. These rules can concern either statistical ethics/principles or the specific methodological rules of statistical production in a given statistical domain.

In the EU the institution that oversees the implementation of statistical ethics/principles in the production of European statistics is the European Statistical Governance Advisory Board (ESGAB) [18]. There can also be national level bodies, such as the Good Practice Advisory Committee of the Greek statistical system that was set up in 2013 [19], that have as a goal the oversight of implementation of some basic statistical principles/ethics in the national statistical system. The work of such institutions can be supplemented by processes whereby the production of official statistics in a country may be scrutinized by international teams of experts regarding whether the statistics follow statistical principles. An example of such processes is peer reviews. There have been two sets of peer reviews in the European Statistical System [20], in 2006–8 and in 2013–15 and all EU member states (and EFTA countries) as well as Eurostat had to participate in them. There have also been peer reviews in a number of African countries, in certain countries of the Latin American and Caribbean Region, and in a few other countries around the world, but they have been voluntary [21,22]. Another example of such processes is the International Monetary Fund's Reviews of Standards and Codes (ROSCs) regarding statistical data [23]. The ROSCs are also voluntary and usually get carried out once for a country with a possibility for an update some years down the road. However, these assessments reached their peak in the first decade of the 21 century and have virtually seized being carried out by the IMF in recent years.

For countries in the EU there tend to be internal and external quality assessment procedures for *specific official statistics data sets* they produce. However, only a small subset of all official statistics produced by EU member states are subjected to rigorous quality assessment by the statistical office of the European Union, Eurostat. These more rigorous and regular quality assessments are tied to the explicit use of the concerned statistics in the governance of the EU, such as the assessment of the quality of public finance statistics provided for in Council Regulation (EC) No 479/2009 on the application of the Protocol on the excessive deficit procedure annexed to the Treaty establishing the European Community [24]. There are also some other vehicles in the international area for assessing the quality of specific statistical data sets or specific aspects of them. One has been offered by the IMF to its member countries, in the context of its Reviews of Standards and Codes (ROSCs). However, as noted above, these assessments are voluntary, usually get carried out once for a country, and have virtually ceased being provided in recent years. The IMF also monitors observance to its dissemination standards – such as the above noted Special Data Dissemination Standard – for certain macroeconomic statistics of countries that subscribe to those standards.

Finally, one might also mention the existence in a handful of countries of a certification process of specific data sets of official statistics. The process is aimed at providing a certification that a certain level of quality is achieved by the official statistics produced by a statistical producer in the country. For example, in the case of the UK, such certification (conferring National Statistics status) is carried out by the Office of Statistics Regulation of the UK Statistics Authority [25]. In other countries (such as Greece), certification is carried out by the national statistical office itself and it concerns the statistics of the other producers of official statistics within the national statistical system.

We have aimed to provide a broad sense of currently available arrangements for assessing and supporting the quality of official statistics, and in particular the implementation of statistical principles/ethics. We have noted the existence of codes of principles and ethics, statistical laws, sectoral statistical legislation, international and supranational standards and manuals, and bodies and processes aimed at assessing adherence to statistical principles and high statistical quality in general. These arrangements can make manipulation of official statistics more difficult.

While the area of official statistics has displayed a significant evolution in recent decades, especially since the 1990s, by building the arrangements described above, what we have described should **not** convey the impression of solidity and adequacy. Specifically, this 'web' of arrangements is very far from covering every country of the world and doing so effectively. Not every country has a dedicated statistical law and not every such law, when it exists, properly provides for the implementation of statistical principles and ethics in the national statistical system. Relatively few countries

have national level institutions with the mandate to oversee the implementation of statistical principles/ethics. Quality assessment of statistical output provided by supranational or international entities does not apply to all countries, is not frequent and effective enough when it takes place, and – even if some domains of official statistics are assessed – many statistical domains are not subject to such quality assessment. Moreover, only a small subset of countries in the world has its implementation of statistical principles/ethics reviewed by supranational entities and processes.[6]

The 'web' of arrangements is relatively dense in some parts of the world, e.g., the European Union, while in other parts of the world the 'web' is quite sparse. One might then be tempted to believe that the area of official statistics is on its way to address the issue of principles and ethics, and thus of corruption, in the production of its statistics and that what is needed is just the spreading of best practices from some parts of the world, such as the European Union, to the rest of the world – the spreading and thickening of the 'web', so to speak, so that it effectively covers the global community of official statistical production.

However, the difficulties in addressing the issues of ethics and quality in official statistics are more formidable than that. There are significant challenges beyond the absence of global coverage/application of the arrangements we have described above. There are problems with the *effectiveness* of the arrangements themselves even when they are applicable to the official statistics of a given country. Enabling conditions of manipulation of official statistics can be pervasive. To the extent that enabling conditions and modalities (see Section 5 below) for manipulation/corruption in official statistics are not addressed, difficulties in this area are bound to continue, which means that corruption in official statistics would not be eradicated or at least minimized.

4. Phenomena of corruption in official statistics

Corruption in official statistics appears as *manifestations or phenomena*[7] *of manipulation of official statistics*. These manifestations or phenomena are what official statistics producers and other actors that interact with them and with the statistics appear to do as they engage in the manipulation of official statistics so as to distort the picture of reality official statistics are supposed to record. These are the outwardly and directly perceptible acts or instruments that lead to the distortion of the picture. However, they do not explain adequately how the manipulation of official statistics got there – what is involved in the manipulation. They are akin to the 'tip of the iceberg'. To reach an understanding of how official statistics are manipulated one needs to understand (i) the *drivers*, (ii) the *modalities*, and (iii) the *enabling conditions* leading to the phenomena of manipulation of official statistics. (For explanations of drivers, modalities, vectors and enabling conditions, see Section 5 below.) Naturally, to address corruption in official statistics one has to address the phenomena by addressing their drivers, enabling conditions and modalities.

Below we first provide an indicative list of phenomena of corruption in official statistics. It is important to note that there are various such phenomena,[8] as there are various instruments that can be used.

4.1. Typology of phenomena of corruption in statistics

The phenomena can usefully be classified in three broad types:

Type A phenomena include the use of the '*crudest' instruments* in the manipulation of official statistics, such as direct alteration of produced statistical results; alteration of upstream data necessary for the production of official statistics; lack of application of standard methods and principles that countries have agreed to apply in the compilation of their official statistics; and suppression, for the short term or the long term, of official statistics that have already been produced.

Type B phenomena include the use of what may be considered as *more 'sophisticated' instruments* in the manipulation of statistics. They involve carrying out the manipulation while intentionally avoiding to formally and very openly infringe on the existing rules of production and dissemination of statistics. The infringement on the rules is there, but it can hide behind what is and what is not in the 'letter' of the rules. Type B phenomena include 'managing' the agreed-to standards to produce statistics so as to misrepresent reality; de-

[6] A detailed empirical analysis of arrangements and practices around the world, such as those referred to in this paragraph, is not within the purview of the present paper.

[7] The term manifestation is used here to denote "a perceptible, outward, or visible expression" [26] of something else. The term phenomenon is used to indicate "an object or aspect known through the senses rather than by thought or intuition" [27].

[8] Based on empirical evidence collected by the author.

manding from statistical producers specific statistical information to be produced (implying that certain information needs to be collected), which in turn affects the quality of a broader group of statistical outputs; choosing (as statistical producer) to use a certain upstream data source instead of another so as to affect the downstream statistical results; choosing (as owner of an upstream data source) to withhold certain upstream data with a view to affecting downstream statistical results; distorting the form of statistical release; suppressing the production of certain statistics by, for example, intentionally withholding funding for their production; undermining the credibility of official statistics either by undermining the quality of official statistics themselves through e.g., weakening the production of the statistics or by undermining the perception of the quality of official statistics.

Type C phenomena include the use of the *most 'sophisticated' instruments* in the manipulation of statistics. This is because they involve the setting itself of the standard methodologies of specific statistics (e.g., the coverage/scope of general government debt) and the codification of ethical principles that are supposed to be followed in the production and dissemination of statistics. By influencing the setting of standard methods and ethical principles to be followed, these phenomena provide the ultimate 'cover' under the agreed standards themselves to intentionally misrepresent reality. The setting of standard methods includes besides the setting of inappropriate new standards with biased motives also maintaining of – or resisting to amend – inadequate existing standards with biased motives. Type C phenomena arise when powerful actors participating in the international and national processes that set (in some legal form) the standard methods and ethical principles undertake such initiatives. For example, in the sphere of international statistical standards, the actors would have to be powerful countries acting alone or with the cooperation or tolerance of other powerful countries. They can also be groups of countries acting as an effective majority 'block' in terms of influence on the setting of new – or on maintaining existing specific elements of – international statistical standards and codified ethical principles.

It should be noted that the goal of manipulation of official statistics is to alter the *perception of the reality that these statistics are supposed to depict* so as to serve certain political interests. Thus, the manipulation of statistics can take the form of altering the statistical results themselves **or** it can take the form of altering *the perception of the statistics*.[9] Thus corruption in official statistics can appear with both these forms of manipulation, and the phenomena that are included in the present typology include phenomena from both these forms and not just from the 'expected' one – i.e., distorting the numerical figures themselves.

Below we provide a listing of phenomena of manipulation of official statistics with a very brief explanation of each (see also Fig. 1 for a snapshot of such phenomena).

Type A

1. **Direct alteration of statistical results**. This involves the substitution of figures with fabricated ones. This is a crude method, which however can be and is used when the enabling conditions are in place. The alteration of final statistical results can take the form of **a modification of an existing final result or the alteration of an input data point** that leads to a modified final result. These changes/fabrications may have something to do with the real figure or may be dictated completely by the need to achieve the desired result of the released (usually aggregate) statistic, i.e., the change is 'reverse engineered'. The alteration of the statistical results is often **centrally sanctioned by statistical producers**, i.e., the statistical producers responsible for the final results are fully aware of and participate in this manipulation. However, in some cases, the statistical producers may not be fully aware, but instead may follow an explicit approach of **neglect and acceptance of whatever is provided to them by upstream data owners**, who are usually (fully controlled by) policy makers. As the downstream, final statistics producers have to be responsible for the quality of the statistical product they release, this approach is effectively equivalent to collusion in the manipulation of the official statistics, especially when it is repeated over time. Finally, there can also be situations whereby the alteration of statistical results is the outcome of **altered upstream or**

[9] Examples of phenomena of manipulation of official statistics aimed at altering the perception of the statistics include distorting the form of statistical releases; undermining the credibility of statistics by undermining the perception of their quality; and undermining the credibility of statistics by weakening production and thus actual quality. These types of phenomena would arguably also include various sorts of suppression of statistics (e.g., on account of the effects and consequences of the absence of the (suppressed) statistics from the public's perception).

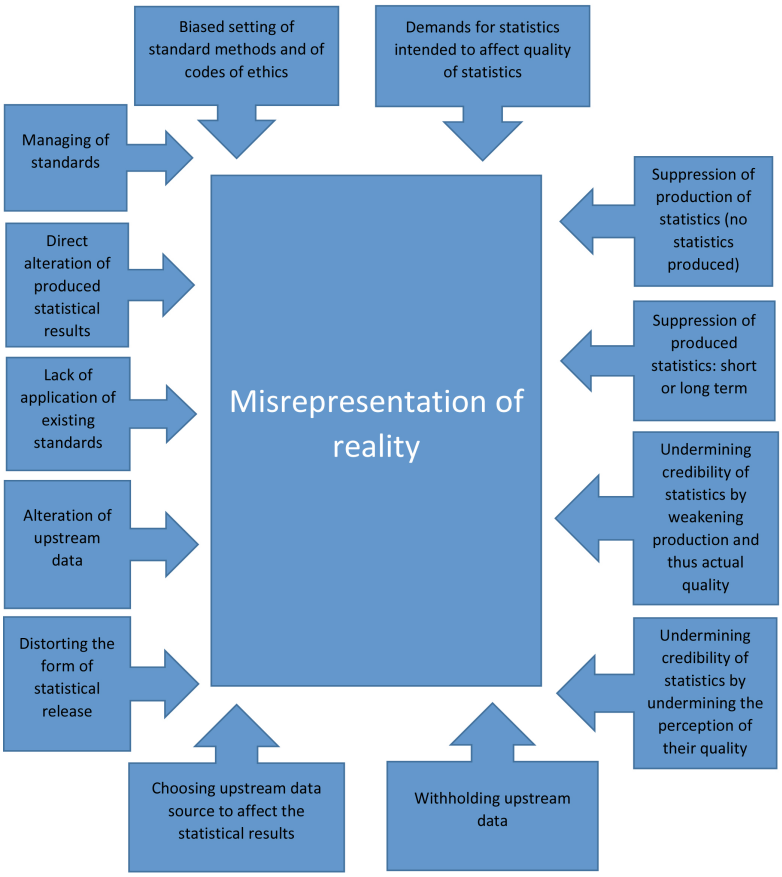

Fig. 1. Phenomena of manipulation of official statistics.

collected data despite the stated intentions of the statistical producers responsible for the final results. This can occur because of powerful personal incentives that may affect the upstream providers of the information, e.g., needing to meet plan targets for economic production allocated to them or plans for environmental or health conditions. Such phenomena have to do with the intentional manipulation of upstream data, which downstream statistical producers may have no way of correcting. However, they would have the possibility and ethical responsibility of refusing to use such data sources, even at the expense of not producing the final statistics. For these reasons, such a phenomenon belongs to Type A.

2. **Not applying existing standards**, whether international or national. In this class of phenomena, international or national standards exist and statistical and policy authorities in the country have made explicit commitments to implement them. For example, these can be standards for statistical recording such as rules regarding definitions/concepts, classification, accrual accounting, and consolidation. They can also be standards regarding principles/ethics, such as professional independence, provision of adequate resources, statistical confidentiality, and timeliness and punctuality. The phenomenon of manipulation of official statistics here takes the crude form of ignoring these standards and statistical production and dissemination taking place as if these standards do not apply.
3. **Suppression of statistics** is a major and time-honored manifestation of manipulation of official statistics. It has been used when official statistics results cannot be changed for one reason or another but powerful interests desire to avoid statistics affecting the perception of reality. There are various versions of this phenomenon. The crudest version of the phenomenon (classified as Type A) involves **suppressing the statis-**

tics *after* they have been produced when the results do not conform to nonstatistical needs and interests. This suppression can be applied for the long run. Another version of this phenomenon is the **temporary suppression of official statistics** that have been produced – basically affecting the date of their release on account of nonstatistical considerations.

Type B

4. **Selective use of upstream data** in order to affect the statistical result is a phenomenon of manipulation of statistics that involves the statistical producer making choices about data sources, not solely on the basis of statistical considerations, such as their quality and contribution to the accuracy and overall quality of final statistical results,[10] but on the basis of achieving nonstatistical goals regarding downstream statistical outcomes. This phenomenon is crude enough that it could be argued that it belongs to Type A phenomena.
5. **Withholding upstream data** by the upstream data provider to serve nonstatistical purposes is another phenomenon of manipulation of official statistics. Withholding of upstream data in order to affect the downstream statistical result and its quality (see below), including timeliness, is a phenomenon that is somewhat more sophisticated than the alteration of upstream data. However, it is crude enough that it could be argued that it belongs to Type A phenomena
6. **Indirect modification of methodology on ad hoc basis** to serve nonstatistical, political goals. This phenomenon of manipulation of statistics is relatively sophisticated as it does not amount to a direct alteration of methodology in order to manipulate the statistics and achieve nonstatistical ends. It often involves making **demands for the production of specific statistical information with a view to affecting the accuracy and quality of these statistics in general and potentially giving rise to biased statistical results.** The quality of statistics can be affected as the interference amounts to an effective intervention in the collection or processing of statistics (the 'how") by making demands about a – only outwardly benign – matter of the information content of the statistics (the 'what'). Asking specific questions in a survey that predictably bias the results of the entire survey (or some important results) would be an example of such a phenomenon. Another version of this phenomenon is when politicians/policy makers impose a **requirement that certain statistics be produced using a specific methodology** (i.e., they explicitly want to determine the 'how' along with the 'what'), thus affecting the quality of the statistics, either with a view to biasing them in a certain way or undermining their credibility (see below). A form of this phenomenon is when the statistical producers are **prevented from using the appropriate methodology** to optimize the quality of the statistical results, with the purpose of serving nonstatistical goals. Here, those that interfere with statistical decisions are interfering with how statistical production takes place, not by imposing an element on the production of the statistics that affects their quality, but by not allowing the use of the most appropriate method in that production. This can affect the quality of the statistics, biasing them in a certain way. For example, detailed (as opposed to abbreviated) forms used with certain groups of respondents in a survey or census may be prohibited by policy makers or policy makers may decide that certain subpopulations (strata) should not be surveyed at all
7. **Managing existing standards**, whether international or national is another important manifestation of manipulation of official statistics. Here, the effort by the perpetrators of the manipulation could take a couple of forms: One version is to find any 'gray areas' in the rules and then to proceed with an **interpretation of the rules in these areas so as to fit a nonstatistical objective**. This appears as the use of 'flexibility' implicitly or explicitly allowed in the rules, as when, for example, there is more than one method available to carry out a recording and a choice is made in order to satisfy nonstatistical goals ('cherry picking'). The manipulation can also appear as interpreting the rules in a way that stretches the intended meaning of the rule (e.g., using an unduly broad interpretation of the rule) by taking advantage of existing or alleged verbal cues in the text in which the standard is laid out. This is done by ignoring the intent – i.e.,

[10] Bona fide considerations of the availability of resources and of avoiding excessive burden on respondents would also be part of such statistical considerations.

'spirit' – of the rule and instead focusing on how to satisfy the nonstatistical objective driving the exercise. Another version of this phenomenon of manipulation of statistics is circumventing the existing rules by **exploiting situations where the rules are not being explicitly exhaustive** about what can be done and what cannot be done, relying instead to a greater or lesser extent on the professionalism and 'goodwill' of national official statistics producers to try to depict the reality as intended by the rule. The phenomenon of managing existing standards often goes hand in hand with the phenomenon of biasing standards at the time these standards are adopted (see below).

8. Manipulation of official statistics can take the form of **distorting the form of the release** of these statistics by the statistical producer so that their perception and interpretation can be affected instead of the statistical figures themselves. Such manipulation of official statistics is not easy to describe a priori but it is usually easy to identify when one sees it, especially in the context of comparisons over time and with other international practices. The phenomena include adding or modifying text, tables, graphs etc. in the written statistical press release of the statistical producer with a view to affecting and distorting the interpretation of the statistics made available or announced through this press release. The phenomena also include excising textual, tabular and graphical elements that ought to be in the press release with the same intent. One version of this phenomenon of manipulation is the suppression of the written part of the press release altogether (i.e., the tables being published without any explanatory text). The effect can under certain circumstances resemble the effect of outright suppression of produced statistics (see above). Moreover, manipulation of statistics by distorting the form of the release does not need to take the form only of intervention in a written press release, but it can also take the form of **distorting a press conference** or other oral public statement by the statistical producer at or around the time of the release of the statistics. A press conference, for example, can be used in many ways to influence the comprehension and interpretation of the statistical results by the broad public/users for the purpose of serving political interests. The ad hoc elimination of a press conference from the statistics office when there was one before or the ad hoc introduction of such a press conference when there was none before can also potentially help bring about this kind of manipulation of the statistics. Such actions by the statistical side are usually complemented by actions taken by the policy/political side to influence in a politically biased way the interpretation of statistics or in some cases suppress knowledge of them. Even an effort to reduce or eliminate any existing embargo on government commentary on statistical releases (e.g., an embargo of one hour after the release by the statistics office) when the statistics are released by statistical producers embedded in policy making bodies can amount to an effort to manipulate the statistics by exerting greater influence on their interpretation by means of *conflating* (in the mind of the public) statistical and policy statements. In summary, the varieties of abuse of statistical release can vary a lot and cannot be subject to an exhaustive description.
9. A more sophisticated form of suppression of official statistics than the suppression of statistics that have already been produced involves the **suppression of the production of statistics** (i.e., suppression of statistics before they are even produced) if there is concern the statistical results may potentially be unfavorable for certain political interests. A variation of that is the **suppression of a revision of statistics**, when such a revision is due. This can happen when the revision may lead to results that are seen by political/policy actors as unfavorable to their interests, both on account of the economic or social reality they present and on account of the perception they create of the integrity of the statistical production that produced the original statistical figures. Thus, a (new) interference and de facto manipulation of statistics may take place in order to conceal the original manipulation. The revision may be suppressed entirely or it may be resisted all along the way by providing various obstacles, such as not providing upstream source data or interfering with the revision process within the statistics office, or using any of a range of other possible means. Another version of this phenomenon is **suppression of the correction of errors** that have been identified when the correction of the errors produces a picture of reality that is unfavorable to political/policy interests.

This suppression of error correction may be welcomed by some within the statistics producing agency, who in order to preserve their reputation and not appear to have erred, may also be unwilling to correct intentional or unintentional errors.

10. Another phenomenon of manipulation of official statistics is the *delegitimation* of official statistics by means of **undermining the credibility of official statistics**. This phenomenon may involve undermining the actual quality of official statistics so that their credibility is damaged as a result, or undermining the image of the quality of these statistics, or both. The end result is that the reality these statistics are supposed to depict is perceived in a distorted manner, either because the statistics are actually inaccurate or because the public sees them as inaccurate and the public believes instead in some 'alternative facts' or 'fake news' propagated by political/policy interests. **Affecting the quality of the statistics themselves** takes the form of the phenomenon of intentionally weakening or keeping weak the statistics agency in various ways so that it cannot meet the existing quality standards and the quality of official statistics is below the socially optimal level. This can be done by intentional neglect or resistance to the fulfillment of official statistics needs so that statistical production ends up lacking the appropriate resources (financial, human, IT, etc.) and/or the authority to use resources that may be allocated to it. It can also be done by impeding in various ways the statistics agency in carrying out statistical and administrative processes, and thus statistical production effectively. **Affecting the perception of the quality of statistics** can of course leverage and exaggerate any real shortcomings in quality, but it can also be completely based on conjured ideas and fabricated information. The effort is to affect the public's trust in official statistics, so that then alternative narratives about past, present and future states of the world can be promoted as 'reality' instead of the picture that would be evident by perusing official statistics. This can be considered as a phenomenon of distortion of the picture of reality that official statistics is supposed to convey, with the aim of serving certain political and social interests.[11] The tools used to bring about this phenomenon vary. They often include public and legal attacks on official statisticians and their statistical output. It should be noted that attacks on the statistics do not necessarily produce a direct win for the manipulators but an indirect one: every official statistical output loses and thus the manipulators win indirectly by being left in a stronger position to peddle their 'alternative facts', views and narratives. The approach can be summarized as 'nothing is true and anything can be true'.

Type C

11. **Biased standard setting** is a powerful instrument that can be used to manipulate official statistics so as to misrepresent reality with a view to serving political/policy interests in a certain context.[12] Biased standard setting can take various forms. One form is the use of **biased 'national methodologies'**, whereby the country formally adopts statistical standards that are different from international standards or from standards that are considered by the large majority of scientists most appropriate for recording and understanding accurately and reliably a given economic, social, or environmental aspect of reality. Another form of biased standard setting involves the **adoption of biased in-**

[11] Put differently, the gap in the credibility/legitimacy – the credibility/legitimacy deficit – of official statistics creates space for other information – 'alternative facts' or 'fake news' – propagated by political/policy interests.

[12] In principle, the setting of statistical standards should reflect the outcome (regarding both settled and open issues) of scientific debates between scientists engaged in them as proponents of different theories vying for understanding a given economic, social, or environmental aspect of reality. Setting standards that reflect today's genuine and appropriately-arrived-at dominant scientific views would not be characterized in our framework as Type C phenomena of corruption in statistics, taking also into account that dominant scientific views evolve over time due to scientific progress. However, if the scientific views that are invoked to underpin the setting of some statistical standard are dominating on account of political/ideological/power/policy interests and are not representing the current pinnacle of scientific understanding, then the setting of that statistical standard would be biased, at least to some degree (or in certain of its features). Some might argue that politics, power seeking or ideology exerting '*normal influence*' on scientific views or on which scientific views dominate, and thus exerting *'normal influence'* on statistical standard setting, is acceptable and would not amount to corruption – unless it becomes 'too much'. In our view, this is risky 'slippery slope' of a perspective that – when operationalized in statistical standard setting and other statistical activities – can end up accommodating a significant amount of corruption in official statistics.

ternational standards/methodologies in some specific area of statistical compilation. It can take the form of international standards being made to mirror national methodologies of dominant countries or blocks of countries or to allow for so-called 'national specificities' with a view to serving nonstatistical interests. Biased standard setting can also involve the intentional creation of gray areas, lacunae and misspecifications in methodology, or allow for alternative methodologies for nonstatistical reasons. All of these can potentially be exploited or 'managed' (see above) in order to manipulate official statistics with a view to achieving nonstatistical ends. There can be overt action to achieve biased standard setting in some aspect/area of a supranational statistical standard or, alternatively, there can be lack of (resistance to) adjustment of inadequate supranational statistical standards to achieve the same. There can also be a phenomenon that could be termed 'de facto' biased standard setting. It may involve a supranational body whose mandate is to clarify or determine what the statistical standard is when it is asked for that view or when it decides itself to provide such a view. In that context, it can be involved in 'de facto' biased standard setting by clarifying or determining the standard in a way that serves national (or multinational) political/policy interests, and not consistently with state-of-the-art scientific/statistical thinking. Finally, a form of biased standard setting is the 'corner solution' of **opting for no international or national standard** in place at all for some statistics. It should be noted that biased standard setting specifically can involve the of inappropriately specifying – or resisting the appropriate specification of – some aspect of international standards on statistical *ethics (principles)*. This can even take the form of suppression in international codifications of statistical ethics/principles of some specific ethical principle regarding the production of official statistics.[13]

The above phenomena of manipulation of official statistics are characterized by an *intent to effect such manipulation*. Therefore, other phenomena affecting the quality of statistics that are not characterized by such *intent* to manipulate the statistics, but e.g., are the result of lack of capacity/knowledge or the result of errors or lack of adequate foresight, should not be classified as phenomena of manipulation. If, however, there is resistance or laxity in recognizing and addressing such quality problems, especially when their effect on statistical results serves non-statistical (e.g., political) interests, then these phenomena would have to be reclassified as phenomena of manipulation.

4.2. Phenomena of corruption in statistics and the official statistics process

Manifestations of manipulation of official statistics can also be thought of in terms of their place in the timeline of the 'life' of an official statistics product. The process of existence of an official statistical product from beginning to end, which we call here the *official statistics process*, can be broadly presented as a sequence of the following 8 stages:[14,15]

I. Setting of standard methodology and principles
II. Decision to engage in production
III. Administrative and statistical preparation
IV. Data collection
V. Data processing
VI. Estimation and synthesis/aggregation/consolidation
VII. Release and communication
VIII. Social reception of the statistical product

The process is presented here as a linear process in time. However, it should be kept in mind that some stages can potentially be concurrent. For example, the setting of standard methodology to be followed can potentially be concurrent with the decision to produce the statistics. Further, the social reception of the statistical product can involve actions that take place even before the release and communication stage, and in fact even before any actual step has been taken to produce a given statistical output. The above sequential/linear presentation of the stages is useful as a 'realistic-cum-logical' presentation.

[13] For example, the ethical principle of professional independence.

[14] Note that not all stages have to be under the control of the official statistics producer (e.g., stage VIII). We include such stages because manipulation of official statistics is not confined to the numerical figures produced by the statisticians.

[15] Various stages of the official statistics process simple model presented in this paper are similar to a number of the "overarching processes" described in the Generic Statistical Business Process Model (GSBPM). However, the official statistics process model in this paper and the GSBPM have different goals and thus they are tailored to these goals in ways that differentiate the models.

A point to note is that phenomena of manipulation of official statistics of Types A, B and C can occur at one or more of these stages. For example: Biased standard setting occurs in stage I. However, suppression of the production of statistics can involve interference in stages II or III. Indirect modification of methodology on ad hoc basis can occur in stage II, III or IV. Selective use of upstream data or withholding upstream data occurs in stage III. Not applying existing standards can take place in any one of the stages III through VII. Managing existing standards can occur in any of the stages III through VII. Direct alteration of statistical results can occur in stages IV, V, or VI, and even VII. Affecting the quality of statistics can occur in any of stages I and III through VII, while affecting the perception of the quality of the statistics can occur in any of the stages III to VIII. Suppression of statistics that have been produced takes place in stage VII. Distorting the form of statistical release occurs in stage VII.

It should be evident that manipulation of statistics in one stage of the existence of a statistical product often takes place with an eye to manipulation in other stages. For example, stage IV, which involves data collection, can see surveys suppressed or weakened and a hampering of the provision of administrative data or an outright suppression of such data, with an eye to easier manipulation of statistical results downstream (at a later stage). Moreover, efforts to manipulate statistics may in some instances require manipulators to 'backtrack' from a later stage to engage in manipulation at an earlier stage. For example, stage VI of synthesis/aggregation/consolidation can be carried out as a rehearsal so as to see if the final result would conform to the 'desired' outcome, which is provided or indicated as a target by nonstatistical interests. If the synthesis of input data does not produce the result expected by the nonstratistical interests, the process can 'backtrack' and data inputs that were already collected and processed can be changed/suppressed ex post facto by revisiting aspects of stages IV and V. A variation of that is a veritable inversion of the official statistics process, whereby 'right off the bat' the 'desired' statistical result (e.g., a certain GDP growth rate or a particular fiscal deficit figure) is used to explicitly 'reverse engineer' what is needed to take place (in terms of manipulation) in various earlier stages of the process.

Thus, to address corruption in official statistics there has to be vigilance at all stages of the statistical process for various manifestations of corruption. Usually, the manipulation of statistics will not be limited to one phenomenon appearing at one stage of the statistical process. If one such phenomenon is identified, usually others also exist (even if not immediately evident to an outside observer).

4.3. Corruption in statistics and states of economic development

We now turn to briefly discuss if there is some specific state of economic development when certain kinds of phenomena of manipulation of official statistics usually appear; put differently, whether there is some correlation of specific phenomena with states of development.

Before discussing this issue, it is important to state that *no state of economic development is immune to manipulation of official statistics*. There exist many examples of phenomena of manipulation at all states of economic development, from lowest income countries to the most advanced economies in the world.

Regarding the issue of whether specific phenomena of manipulation of official statistics are associated with specific states of development, our view is that there exists *no precise correspondence* between level/state of development of a country and the kind of manipulation of official statistics that the country engages in. In principle, all kinds of manipulation are possible at all states of development. However, there may be some tendencies one can note.

Some phenomena of manipulation are relatively more 'crude' (such as Type A phenomena above) and rely on a greater disregard for the rule of law and on a lack of effective checks and balances, whether political or via the markets. These kinds of phenomena thus tend to be more characteristic of countries with political systems where the *rule of law* and *checks and balances are weaker*. These kinds of phenomena are associated with flourishing generalized corruption.[16]

'Cruder' phenomena of manipulation of official statistics are also more common in countries where the market system is not dominant but the *economy is heavily dominated by and subservient to nonmarket sociopolitical mechanisms*, such as in political totalitarian/authoritarian systems, clientele-based systems, and clan-based systems. Often, sociopolitical systems such as the ones referred to above are associated with lower levels of economic development on account of the severe distortions in the allocation of resources that

[16] I.e., corruption inundating most aspects of life in the country and not just statistical production.

characterize such systems. Thus, 'cruder' phenomena of manipulation of official statistics tend to be associated with lower levels of economic development, and in particular with stagnant and crisis prone states of economic development.

Countries that are advanced in terms of economic development are often also countries that have political systems with relatively stronger checks and balances and rule of law. They also tend to have relatively developed national statistical systems and corresponding legal/institutional frameworks for official statistics. This implies that certain enabling conditions and modalities of corruption in official statistics may be less likely to exist or arise. In turn, this tends to result in relatively fewer instances of the 'cruder' (Type A) phenomena of manipulation of official statistics in such countries. In these countries, other, more 'sophisticated' types of phenomena (Types B and C) may be the usual means, if and when there is manipulation.

Countries that are advanced in terms of economic development yield power in supranational fora and thus have the capacity to influence the setting of statistical methodology standards as well as codified standards of statistical principles/ethics. This may dovetail with their potential interest and capacity in 'managing' standards after the latter are set. They may also be more willing and able to resist the setting of standards in some statistical domain in the first place. Advanced countries may also be more sophisticated in preparing and 'clothing' the suppression of the production of certain statistics via processes that are formally legal and more legitimizing (e.g., by democratically voting not to fund through the annual budget law certain statistical products for reasons of suppressing that kind of information).

It is therefore more likely to find a country at low level of development relying on direct alteration of statistical results or on not applying existing statistical compilation standards than to find an advanced economy doing the same. At the same time, it is more likely to find an advanced economy relying on biased standard setting or 'managing' standards as a way of manipulating the picture of reality that official statistics are supposed to convey. All the above noted phenomena – whether perpetrated in countries with a low level of economic development or a high one – are phenomena of manipulation of official statistics, i.e., of corruption.[17]

[17]It is not within the purview of this paper to present evidence on the question of which type of countries – less developed or more developed ones – practice more frequently manipulation of official statistics or with greater impact, nationally or globally.

5. The process of corruption of official statistics

5.1. A schematic model

Phenomena of manipulation of official statistics (such as the ones described in Section 4 above) can be understood by identifying the *process* or *mechanism* that gives rise to such phenomena and the *elements* or *component parts* of this process/mechanism. These elements or component parts can be categorized as follows:

(i) drivers of the phenomena
(ii) enabling conditions of the phenomena
(iii) modalities or methods used to arrive at the phenomena
(iv) vectors or agents that execute/propagate the phenomena
↓
(v) manifestations/phenomena (outward instruments) of statistical manipulation

The process and its elements constitute a schematic model. All cases of phenomena of manipulation of official statistics can be analyzed using this schematic model. A graphical description of the model is provided below (Fig. 2).

The model can also be described as follows: There are **drivers** for the manipulation of official statistics so that reality is misrepresented in order to serve various political, economic, social interests. These drivers operate within a certain institutional and cultural environment. If there are weaknesses in institutional setups, in governance structures, in good administrative governance, in culture, in legal, regulation and deontological frameworks, etc., then there are **enabling conditions** for the manipulation of official statistics to take place. These enabling conditions allow for various **modalities/methods** to become feasible and to be created and/or activated so as to bring about manipulation of official statistics. Depending on the enabling conditions available, there can be certain sets of modalities available, and, further, certain **phenomena** of manipulation (appearing as outward instruments of manipulation) to finally emerge. The enabling conditions delimit both the modalities and the phenomena/instruments. Those that carry out – the agents that execute and propagate – the various large and small steps of the modalities in order to end up with phenomena of manipulation are the **vectors/agents**. They are the human agents that carry out the necessary – inconsistent with statistical principles/ethics – deeds that are necessary to arrive at the manipulation of official statistics. The creation,

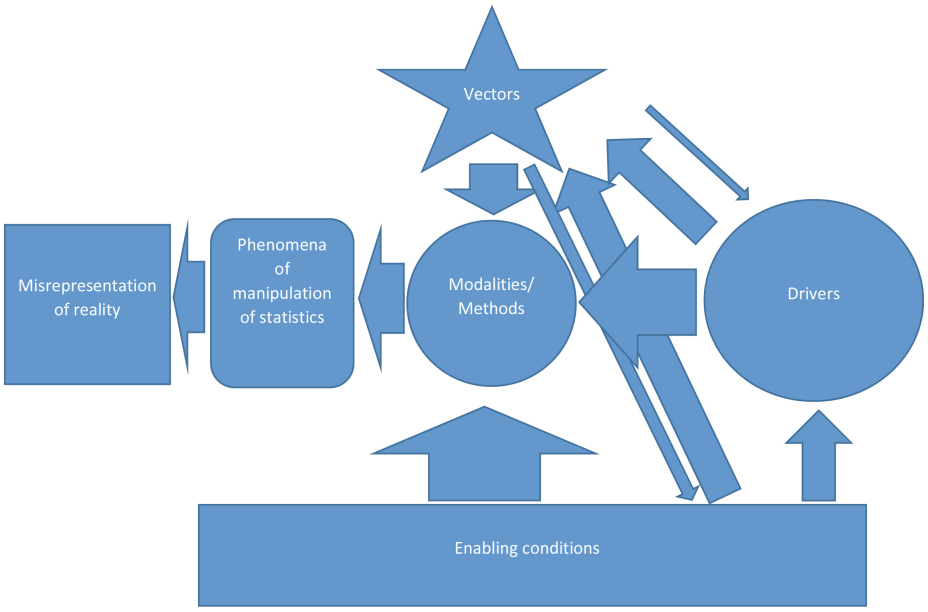

Fig. 2. Model of the process of corruption in official statistics.

flourishing and resilience of the vectors depends to a large extent on the enabling conditions. The enabling conditions also affect the environment in which drivers develop and they can foment the emergence of drivers. Similarly, the drivers affect the emergence and flourishing of vectors. The vectors in turn affect the modalities/methods used in the manipulation of official statistics and are supportive of drivers of manipulation and of the enabling conditions themselves. Thus, there are various reciprocal effects (of various strengths) between all the elements/components of the process of manipulation of official statistics. In the schematic presentation above, the dependencies and effects between the elements of the model are designated by arrows (indicating also their strength).

The benefit of having such a model is that it enables one to identify what institutional setting, action, person or institution presents a problem, vulnerability or source of risk in a given system of production of a specific official statistics product in a given country, and address it. The model can also help inform a discussion about what needs to be changed at the level of international/supranational arrangements, whether concerning institutional settings, processes, or legal and ethical frameworks, affecting the production of official statistics.

Below, we further specify the nature of the elements/component parts of the process of corruption of official statistics – the drivers, modalities/methods, vectors and enabling conditions – leading up to the manifestations/phenomena of manipulation of the statistics

5.2. *Drivers*

The drivers behind the phenomena of corruption of official statistics are the interests and incentives of the actors directly or indirectly involved in these phenomena, and without which the phenomena – although they would be theoretically possible given the presence of enabling conditions – would not arise. Drivers include the interests, concerns and incentives of those that are in some way affected by the content of official statistics – the information they convey – and at the same time have the power to directly or indirectly adversely affect the production of those official statistics. The actors may be physical persons, such as politicians and policy officials (including officials of the legislative branch of government), judicial officials, civil service administrators, the management and staff of statistical offices, union officials, and others. These actors may be acting as explicit or implicit agents of broader entities/groups, such as political parties, unions, parts of the state administration, private business interest groups, other groupings in the population of a country, or even groups that span various countries and operate at international level. Thus, the drivers behind the phenomena of manipulation of official statistics include the incentives and interests not only of the specific physical persons noted above, but also – and very importantly – the interests and incentives of the larger entities/groups *as understood and operationalized* by those who have the

power and the authority to take action that can affect the production of the official statistics at hand.

It should be noted that a driver for the manipulation of official statistics may be for any of the various types of manipulation (A, B or C) discussed in Section 4. The driver may find its fulfillment in a 'crude' outright changing of statistical figures or another, more 'sophisticated', misapplication of ethical/quality principles; for example, changing of the time of release of statistics to fit a political timetable or 'democratically' deciding to suppress for the long haul the production of certain statistics. The point is that corruption in official statistics can take various forms, and the drivers for such corruption lead to a wide variety of phenomena of manipulation.

5.3. Modalities, methods and vectors

The modalities and methods used to arrive at the phenomena of corruption of official statistics are a great many.[18] They are often akin to the 'nuts and bolts' giving rise to/behind the outwardly appearing instruments of manipulation outlined in Section 4, although sometimes the modalities/methods of manipulation and the phenomena of manipulation simply coincide. The modalities/methods can be classified in various logical-systematic ways so as to enhance the usefulness of their listing. Here is one possible way of classification:

Mechanisms. They *inter alia* include modalities such as intentional lack of documentation both on how statistical production should be done and how it was actually done; lying about the processes, methods and data used and the application of standards; colluding to ensure that statistical methodologies are not applied correctly across various areas so as to be consistent with the lack of proper application of a given methodology; manipulating statistical results using the guise of revision and the false excuse of more recent data; intentionally perpetuating a lack of regular information flow from upstream data sources and reporting fictitious upstream data when no reliable data actually exist; overstating the capacity to produce appropriate quality upstream data and providing upstream data that just fit policy targets; outright falsifying of upstream data that are otherwise available and of appropriate quality; perpetuating past manipulation of official statistics by suppressing/withholding existing upstream data sources that would reveal it/reverse it; manipulating statistics by 'managing' the statistical standards; demanding 'what' statistics to produce to affect their and other coproduced statistics' accuracy and quality; policy side imposing a requirement that certain statistics be produced using a specific method to affect the statistical results; suppressing products of official statistics by preventing their production; policy side intervening in the statistical release process to suppress specific information; suppressing produced official statistics by preventing their release in a timely manner; suppressing statistical results by not following standard (full) release/dissemination processes; directly altering the content or form of a statistical release; setting of some aspect of an international statistical standard by the policy side on the basis of non-statistical considerations; 'de facto' biased standard setting by supranational bodies that are mandated with clarifying or articulating statistical standards; etc.

Pressure and Incentives. They *inter alia* include modalities such as policy institutions pressuring in various ways statistics office staff on statistical recording to force statistical results to conform to policy targets (e.g., aggressive communications to statistical staff; coordinated demands of various policy institutions for meetings to provide 'explanations' on the statistical recording; threatening to go public with criticisms; using 'cooperative' statistics office staff as allies); firing statisticians, threatening to fire them, or forcing them to resign; prosecuting, imprisoning, executing statisticians for issues related to statistical production; para-state actors threating and exercising violence against statisticians; taking and exercising legal action against statisticians for unrelated to the production of statistics issues; threatening statisticians' professional reputation; leveraging of personal favors and personal relationships between politicians and statisticians; directions from senior policy makers to the head of the statistics office with explicit or implicit pulling of rank, threats and promises; etc.

Legitimation and delegitimation processes. They *inter alia* include modalities such as invoking science; invoking the interest of the nation, the state, or the 'people'; manipulating statistics by creating a sense of transparency in some areas of statistical production, while intentionally keeping other areas opaque and subject to manipulation; suppressing criticism of the accuracy of official statistics and of violations of statistical ethics; etc.

[18] The examples regarding modalities, methods and vectors noted in this section are based on empirical evidence collected by the author.

Creating, transferring and activating vectors. They *inter alia* include modalities such as creating vectors by 'bringing into the fold and influencing' official statisticians; hiring in the statistics office as a mechanism of creating vectors; having mixed teams of policy makers and statisticians carrying out statistical production or dissemination with a view to controlling the official statistics process; designing so that policy side experts are the only ones available or are given primacy in statistical production; involving 'old hands' previously involved in statistical manipulation or 'mercenary' outside experts to bring about the manipulation of statistics, etc.

The specific modalities/methods provided here under each class are only some examples of the possible set of such modalities/methods, a more complete enumeration and elaboration of which is the subject of another paper.

5.4. Enabling conditions

In general, enabling conditions for corruption in official statistics are defined as situations/circumstances or processes that enable someone to do a particular thing related to corruption of official statistics; they give someone the opportunity to do it; and/or they make it possible for a particular thing to happen. Enabling conditions of corruption in official statistics make the manipulation of statistics possible and/or increase the risk of manipulation of statistics.

Some examples[19] of enabling conditions for corruption in official statistics are the following:

- Politicians/policy makers having prerelease access to statistical releases;[20] clearance of statistical releases outside the statistical perimeter;[21,22] absence of a commitment to an (adequate) statistical release calendar; fluidity of content and form of statistical releases and inadequate separation of statistical releases from analyses; the statistics producing entity being part of a policy making body;[23] policy side and statistical side being 'too close' institutionally and culturally;[24] hierarchical administrative relations allowing the replacement of statisticians by the policy side and the modification of their benefits;[25] hierarchical administrative relations allowing the investigation of the conduct of statisticians and their disciplining by the policy side;[26] hiring and firing process for the statistics office controlled by the political/policy side;[27] an environment where there is the possibility of consequences for official statisticians outside the law; policy institutions being jointly involved with statistical agencies in the production of certain official statistics; decentralized production of official statistics; lack of access of official statistics producers to the full range of needed administrative information; the national statistics office and its leadership and staff being part of a clientele system and a system of favoritism; statistics producing bodies lacking adequate human capital and other resources and/or the authority to use the resources they have; the post of the head of the statistical office being a political position; simple transfer mechanisms of staff between policy bodies and statistical producers; absent, unclear, or inappropriate national statistical legislation; absent, vague or misspecified international standards; weaknesses in enforcing adherence to existing international standards; lack of adequate checking of national official statistics by a supranational body;[28] government databases not being available to outside controllers of the quality of the statistics; setting of international statistical standards by the policy side; political dependence of supranational statistical institutions; lack of an adequately free and ethical press; etc.

This is only an indicative list of possible enabling conditions; more complete enumeration and elaboration of such enabling conditions is the subject of another paper. Some general points about enabling conditions of corruption in official statistics are provided below.

[19] The examples of enabling conditions noted in this section are based on empirical evidence collected by the author.

[20] A relevant discussion is available in [28].

[21] Ibid.

[22] Statistical perimeter is defined [29] as the line between those outside the statistical perimeter and those inside, whereby outside the statistical perimeter are users of official statistics such as policy makers, legislators, civil servants/administrators as well as the press, market participants, academic researchers and the general public, and inside the statistical perimeter are official statisticians directly involved at the given point in time/stage of statistical production (including quality assurance and release). It should be noted that administrative data or other upstream data providers are not within the statistical perimeter for the production of the downstream statistics but outside it.

[23] A relevant discussion is available in [30].

[24] A relevant discussion is available in [29] as well as [30].

[25] Ibid.

[26] Ibid.

[27] Ibid.

[28] A relevant discussion is available in [31].

Even if there are important and multiple enabling conditions present in a country, actually the manipulation of statistics may not arise (at least for a while). Thus, the fact that an enabling condition for manipulation of official statistics has been present (maybe for a long period of time) but has not been associated with an actual manipulation of official statistics does not mean that the condition is 'misspecified' as enabling such manipulation – it does not mean that it is benign. The utilization of the enabling condition can be just 'potential' for a long time, and under the right circumstances, e.g., in the presence of a new or enhanced driver for manipulation, the condition can operate to enable the actual manipulation of official statistics.

It is noteworthy that a limited subset of enabling conditions and modalities of corruption in official statistics could be considered as referring to similar things, but there is a critical difference: in the case of modalities there is 'overt intent' and 'instrumentalization' of a setup to effect manipulation of statistics, while in the case of enabling conditions there is basically 'potentiality', which may or may not be used/fulfilled.[29]

While some enabling conditions are necessary for the implementation of some modality of manipulation of official statistics, other enabling conditions are not essential for the modality to be used – they just increase the probability that it will be used and that this will be done successfully.

There is not necessarily a one-to-one correspondence between enabling conditions and modalities of manipulation: some conditions are very broad and serve to make possible or make more likely a number of modalities. Other enabling conditions may be more specifically tied to certain modalities.

Enabling conditions for statistical corruption usually exist in *clusters*. They often depend on each other and reinforce each other. For example, the statistics office being part of a policy making body is one enabling condition, and it is clustered with other enabling conditions such as hierarchical relations between policy makers and statisticians, prerelease access to official statistics by the policy side, clearance of statistical releases by politicians/policy makers, etc.

There is something that could be described as a *cascading effect* of enabling conditions. Some of these enabling conditions are more fundamental than others, meaning that they are actually enabling conditions *for other enabling conditions*. For example, the statistical office being part of a policy making body is an enabling condition of prerelease access to statistics by policy makers. At the same time, there is also a *feedback loop* in operation: for example, the existence of prerelease access by policy makers will make a move to extract the statistical office from a policy making body more difficult to come about, as policy makers may not want to risk losing their current 'access' to the statistics before the universal release of the statistics to users.

Finally, and quite importantly, there are usually a number of enabling conditions operating simultaneously to support the implementation of one or more modalities of manipulation of official statistics. All together the existing enabling conditions will work to strengthen the enabling environment for corruption in statistics. What is important is how big a cluster of enabling conditions is and how many clusters there are. If there are lots of elements in a cluster, then the environment will be more conducive to corruption in statistics and corruption will arise with greater probability. Moreover, if there are multiple clusters, this probability will increase further.

6. Summary and some conclusions

In this paper we argued that the manipulation of official statistics, when it happens, is corruption and indeed it is grand corruption and political corruption. We noted the steps that have been taken in recent decades that inter alia help make corruption in official statistics more difficult, but argue that instances of such corruption persist. We provided a categorization of the various broad manifestations/phenomena of such corruption. We then proceeded to identify the process/mechanism that gives rise to these manifestations/phenomena, its elements/component parts, and the ways these parts interact. We proposed this schematic model for the purpose of understanding corruption in official statistics. In this context, we provided a discussion of the nature of the individual elements/component parts of the process of corruption of official statistics that lead to the manifestations/phenomena of corruption, i.e., we discussed the drivers of the phenomena; the enabling conditions of the phenomena; the modalities and methods used to arrive at the phenomena; and the vectors or agents that execute/propagate the phenomena.

We believe all cases of corruption in official statistics can be analyzed using the schematic model developed

[29] For example, consider the enabling condition "an environment where there is the possibility of consequences for official statisticians outside the law" and the modality "para-state actors threating and exercising violence against statisticians". The former describes a potentiality, while the latter describes an overt intent.

in this paper. The benefit of having such a model is that it enables one to identify what institutional setting, action, person or institution presents a problem, vulnerability or source of risk in a given setting of production of a specific official statistical product in a given country, and address it. The model can also inform a discussion about what needs to be changed at the level of international/supranational arrangements, whether concerning institutional settings, processes, or legal and ethical frameworks, affecting the production of official statistics. Recommendations for addressing corruption in official statistics have to be based on in depth discussions of the drivers, modalities and enabling conditions of such corruption and would be the subject of another paper.

We emphasize that corruption is not the usual mode of official statistics; all national statistical systems do not suffer all the time from such phenomena. Yet, corruption is possible, in isolated or in pervasive form, in any statistical system when the conditions are right. And only by recognizing this, measures can be taken to maintain the integrity of official statistics.

References

[1] The Merriam-Webster.com Dictionary, s.v. "corruption (n.)", https://www.merriam-webster.com/dictionary/corruption (accessed 26 April 2020).

[2] Transparency International. Available from https://www.transparency.org/what-is-corruption#define (accessed 26 April 2020).

[3] Encyclopedia Britannica, https://www.britannica.com/topic/social-contract (accessed 26 April 2020).

[4] Independent Evaluation Office of the International Monetary Fund. Behind the Scenes with Data at the IMF: An IEO Evaluation. 2016. Available at: https://ieo.imf.org/en/our-work/Evaluations/Completed/2016-0324-behind-the-scenes-with-data-at-the-imf.

[5] European Commission. Report on Greek government deficit and debt statistics. Annex 1 – Methodological Issues. COM (2010) 1-final. 2010. Available at: https://ec.europa.eu/eurostat/documents/4187653/6404656/COM_2010_report_greek_annex_1/4b10d2ba-7a93-4c18-b50f-957f2ad3af8a.

[6] Lichtblau E. Profiling Report Leads to a Clash and a Demotion. NY Times. 24 August 2005. Available from: https://www.prisonpolicy.org/scans/disturbing/2005/profiling.shtml.

[7] Seltzer W. Politics and Statistics: Independence, Dependence or Interaction? United Nations, Department of Economic and Social Information and Policy Analysis, Working paper series No. 6. 1994. Available from: https://unstats.un.org/unsd/statcom/FPSeltzer.pdf.

[8] Thomas PH. Comment l'Etat embellit ses comptes. Trends. 30 August 2018. www.trends.be.

[9] Reichmann T, Monasterski C. Inadequate Statistics and Faulty Analysis. IEO Background Paper, Independent Evaluation Office of the International Monetary Fund. Series number BP/16/04. 2016. Available from: https://ieo.imf.org/en/our-work/Evaluations/Completed/2016-0324-behind-the-scenes-with-data-at-the-imf.

[10] United Nations Statistics Division. Fundamental Principles of Official Statistics. Available from: https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx.

[11] European Statistics Code of Practice. Eurostat website. Available from: https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142.

[12] Code of Practice for Statistics. UK Statistics Authority website. Available from: https://www.statisticsauthority.gov.uk/code-of-practice/#:∼:text=The%20Code%20provides%20producers%20of,by%20government%20serve%20the%20public.&text=Voluntary%20application%20of%20the%20pillars%20of%20the%20Code.

[13] UN NQAF. United Nations Statistics Division website. Available from: https://unstats.un.org/unsd/methodology/dataquality/un-nqaf/.

[14] Regulation (EU) No 223/2009 of the European Parliament and of the Council on European statistics, as amended by Regulation (EU) No 2015/759. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R0223-20150608&qid=1435820363605&from=EN.

[15] Regulation (EU) No 549/2013 of the European Parliament and of the Council on the European system of national and regional accounts in the European Union. Available from: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32013R0549.

[16] System of National Accounts 2008. United Nations Statistics Division website. Available from: https://unstats.un.org/unsd/nationalaccount/sna2008.asp.

[17] IMF Standards for Data Dissemination. International Monetary Fund website. Available from: https://www.imf.org/en/About/Factsheets/Sheets/2016/07/27/15/45/Standards-for-Data-Dissemination#:∼:text=The%20Special%20Data%20Dissemination%20Standard,financial%20data%20to%20the%20public.

[18] European Statistical Governance Advisory Board. Eurostat website. Available at: https://ec.europa.eu/eurostat/web/esgab/introduction.

[19] Good Practice Advisory Committee. Hellenic Statistical Authority (ELSTAT) website. Available at: https://www.statistics.gr/en/good-practice-committee.

[20] Peer Reviews in the European Statistical System. Eurostat website. Available from: http://ec.europa.eu/eurostat/web/quality/peer-reviews.

[21] Blazyk S, Charumbira G, Diop L, Strode M, Williams T. Peer Reviews of African National Statistical Systems, African Statistical Journal 10 (May 2010). Available from: https://www.paris21.org/sites/default/files/ASJ_Vol10_peer_reviews.pdf.

[22] Martin-Guzman P, Aguilera M. Statistical governance in the latin american and the caribbean region: achievements and challenges. Statistical Journal of the IAOS 31 (2015). Available from: https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji907.

[23] Reports on the Observance of Standards and Codes (ROSCs). International Monetary Fund. Available from: https://www.imf.org/external/NP/rosc/rosc.aspx.

[24] Council Regulation (EC) No 479/2009 on the application of the Protocol on the excessive deficit procedure annexed to the Treaty establishing the European Community. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009R0479-20140901.

[25] Assessment. Office for Statistics Regulation. UK Statistics Authority website. Available from: https://www.statisticsauthority.gov.uk/osr/what-we-do/assessment/.

[26] The Merriam-Webster.com Dictionary, s.v. "manifestation (n.)", https://www.merriam-webster.com/dictionary/manifestation (accessed 26 April 2020).
[27] The Merriam-Webster.com Dictionary, s.v. "phenomenon (n.)", https://www.merriam-webster.com/dictionary/phenomenon (accessed 26 April 2020).
[28] Georgiou A. Prerelease access to official statistics is not consistent with professional ethics. Statistical Journal of the IAOS 36 (2020), 313–325. Available from: https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200620.
[29] Georgiou A. The production of official statistics needs to be a separate branch of government. Statistical Journal of the IAOS 34 (2018), 149–160. https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji170399.
[30] Georgiou A. Extracting statistical offices from policy-making bodies to buttress official statistical production. Journal of Official Statistics 35(Issue 1) (2019). Available from: https://search.proquest.com/openview/aa00cd2339e01bd5b3f29f0023524fcf/1?pq-origsite=gscholar&cbl=105444.
[31] Georgiou A. Towards a global system of monitoring the implementation of UN fundamental principles in national official statistics. Statistical Journal of the IAOS 33 (2017). https://goo.gl/R8NKzU.