Measuring the risk of corruption at territorial level and promoting transparency[1]

"National Operational Program on Governance and Institutional Capacity"

1. Methodologies, approaches and indicators for measuring corruption risks at territorial level

ANAC has long been coordinating the project "Measuring the Risk of Corruption at Territorial Level and Promoting Transparency" – funded under the "National Operational Programme on Governance and Institutional Capacity 2014-2020" (hereinafter PON Risk Measuring Corruption or Project) – involving several national institutions and experts on the subject, and whose deadline is scheduled for the end of 2022.

The main purpose of the project is to provide adequate indicators to detect corruption at territorial level, to support prevention and integrity and to promote transparency in the actions of the public administration.

Although elusive by its very nature and difficult to measure, corruption is not exempt from a high statistical incidence in certain contexts and recurrent cases, an in-depth knowledge of which can help both the prevention and the general fight against it. In this perspective, work is being done within the Project to draw up a set of indicators to be made available publicly reflecting the presence of pathological events and the risks of corruption. The final outputs of the project will be indicators of corruption risks, or red flags, which, while not demonstrating the presence of corruption, should warn of its possible presence. The identification of these red flags could facilitate the implementation of preventive and law enforcement measures and the identification of the areas that have a higher exposure to corruption that are worthy of the attention of civil society and citizens at large, and in which to focus investment in terms of prevention activities. The publication of such indicators through a dedicated interface tool on the Web will guarantee a high level of accessibility to many categories of stakeholders.

Within the Project, which entrusts the Authority with a central role – in line with its institutional mission of preventing corruption- networks of inter-institutional collaboration are being created to guarantee the sharing of the scientific methodologies used, of the data processed, and of the indicators developed.

The measurement of corruption risk in the Project is based on three pillars:

1. gathering of data contained in national databases to feed a *business intelligence* system capable of providing dashboards of indicators and *red flags* on the various aspects of corruption and *maladministration*;

2. construction of a set of risk indicators at territorial level and of a set of social capital and context indicators that can help in validating the risk indicators, enhance their interpretation and highlight possible correlations;

3. promotion of civic participation and dissemination of data on corruption risks, as well as on methodologies to design and validate them, to strengthen a culture of integrity.

As coordinator of the PON, ANAC is therefore working to integrate different relevant data sources, to design methodologies for calculation and validation of indicators, to involve a wide range of institutional, academic, research,

---

[1] *Carried out as part of the project "Measuring the risk of corruption at territorial level and promoting transparency" of the National Anti-Corruption Authority (ANAC), funded by the National Operational Programme "Governance and Institutional Capacity 2014-2020" – ERDF Fund – CUP E89G18000140006 – ASSE 3 – Specific Objective 3.1 – Action 3.1.4*

NGOs and other relevant actors to work together on the production and subsequent use of data and of indicators. Such a collaborative and participative approach, it is believed, is the most appropriate when considering the objective to maximize the practical use of the data and indicators that are collectively produced, and, ultimately, to maximise the impact of the project itself. The Project promotes a participatory path that will produce data and indicators on corruption risk, contextualizing them at the territorial level, which can be useful for the prevention and repression of corruption, but also for *policy making*, for scientific and sociocultural debate and, possibly, for *benchmarking*. Sharing such methodologies and experiences could also be useful to further promote an international debate on the subject, to start the verification of the replicability of some of the results achieved in different contexts.

2. The National Database of Public Contracts in the context of the Project

The main source of data within ANAC is the National Database of Public Contracts (hereinafter BDNCP). Part of the data and indicators managed under the PON are taken from the BDNCP.

The BDNCP is a database that collects, integrates and reconciles data concerning public contracts transmitted by contracting authorities. The system is open to interoperability, under application cooperation, both with the internal systems of the Authority, and with similar systems of other administrations.

ANAC manages the BDNCP, which incorporates all the information concerning public procurement procedures in order to ensure unified accessibility, transparency, publicity and traceability of the whole procurement process. ANAC establishes the modalities for the holders of such databases, subject to signatures of interoperability protocols, to ensure the confluence of the data.

The guiding principles of the BDNCP are the following:

1. to integrate the information collected in the context of the Authority's activities into a single database;

2. to have reasonable, logically aggregated, accessible, reliable, up-to-date real-time data, secure and comparable over time;

3. to make a database for decision-making support available to the various *stakeholders* involved in the procurement management processes.

Currently, this database supports both the Authority in its supervisory and regulatory functions, and all other public administrations involved in the life cycle of Public Contracts for planning, containment of expenditure and monitoring purposes.

Data from each step of the complex procurement process are systematically acquired to ensure a timely and effective monitoring of the financial flows generated by public procurement and the collection of large amounts of data from geographically distant public bodies.

Data stored within the BDNCP allow:

1. geographical studies, for example indicating how contracting authorities are distributed within the national territory in relation to different types of contracts;

2. economic surveys, for example showing which contracting authorities spend more, how often, and to what ends;

UNIONE EUROPEA
Fondo Sociale Europeo
Fondo Europeo di Sviluppo Regionale

Agenzia per la
Coesione Territoriale

ANAC AUTORITÀ NAZIONALE ANTICORRUZION

GOVERNANCE
E CAPACITÀ
ISTITUZIONALE
2014-2020

3. surveys on the duration of procurement processes;

4. identification of recurring patterns of behaviors as "red flags" of corruption.

The use of the BDNCP enables the development of transparency and efficiency in the management of public procurement through:

1. Digitization and simplification of the purchasing process;

2. Its centrality as the only reference data source for the public contract market;

3. Standardization of data collected on the life cycle of public contracts (complete adoption of *EU* eForms);

4. Increased transparency and monitoring through maximum publicity and sharing of available data.

The progressive digitization and use of electronic tools, the standardization of tendering procedures and the widespread availability of contract data are powerful tools to support transparency, competition and the prevention of corruption.

The BDNCP represents a unique experience at the European level, so much so that in 2018 it was awarded the first prize of the *"Better Governance through Procurement Digitalization"* competition launched by the EU Commission, on the basis of its "completeness, data integrity, interoperability, availability of access and information analysis, *governance* and sustainability". The BDNCP was recognized as the best example of "*National Contract Register*" within the European Union[2]. Indeed, at the European level there are no other comparable databases in terms of coverage, which also explains the considerable interest that the BDNCP has generated over the years from researchers working in universities and other research institutes.

In order to confirm and improve the results already achieved, the BDNCP is constantly evolving on more than one dimension. Its accessibility and usability has recently been improved, and the quality of the data is constantly monitored and is the focus of continuous improvements. In particular, with respect to the former, following a major overhaul of the database, since September 2020 the BDNCP has been made available as easily accessible and structured Open Data, thus facilitating "expert" users who intend to conduct massive downloads[3]. A dashboard has also been created for guided access to the available information, which allows *self-service* analysis, from aggregated data to the details[4].

The publication of the BDNCP contents in open data is an important result, as it facilitates even sophisticated forms of public use of a database of strategic interest to the country. The BDNCP can be used both to obtain timely information on individual purchases, and a series of useful descriptive statistics – through the presence of an easy-to-use interface – and to process "big data" (approximately 5 million records per year).

The establishment of databases and the publication of data in open format on *public procurement* systems is certainly not a novelty at international level, but the BDNCP stands out for the wide coverage, quantity and quality of the data available. Increased openness of the data, together with the uses of these and of other data sources within the Project, will allow it to feed a complex ecosystem, within which we can find universities and research institutions, the media – today increasingly interested in forms of "data driven" journalism– and non-profit organizations. Such potential, the details of which are discussed towards the end of this document, is enhanced by the choice of making not only the data openly available, but also the *code* that produces all the analyses carried out within the project – in primis, for the computation of the *red flags* of corruption.

---

2 https://ec.europa.eu/growth/content/european-commission-award-better-governance-through-procurement-digitalisation_en.

3 See: www.anticorruzione.it/portal/public/classic/Comunicazione/News/_news?id=c591bc320a7780422318aa12cd2d83d5.

4 Link to the portal: https://dati.anticorruzione.it/#/home

3. Measuring the risk of corruption

As mentioned above, the main objective of the Project is to provide appropriate indicators to measure corruption at the territorial level, to support the prevention of corruption and to promote transparency in the action of the Public Administration, while at the same time overcoming the limits of the statistical measurement of corruption currently available, mainly based on perceptive indicators.

The production of risk indicators of a quantitative nature on a systematic basis is, indeed, an important contribution not only in terms of increasing the scientific knowledge of the phenomenon, but also to design contrast measures that could be more effective and adapted to the diversity of the contexts. Below is a brief description of the indicators that have been designed and evaluated , and that will be shortly available to the public via a Web-based dashboard that presents wide possibilities of customized searches.

3.1. Indicators of corruption risk in procurement

The Project is focused on the development of indicators of corruption risks in public procurement, due to the relevance and high risk of corruption in this area. The indicators will be evaluated using experimental observations, thanks to the availability of the BDNCP, with its more than 5 million public contracts listed in 2020 alone, and 53 million contracts over the last 10 years, for a value of about 2,240 billion Euros.

There are two possible approaches that could be followed when designing risk indicators: a) the first follows a *logical-deductive* methodology grounded on previous knowledge about the phenomenon; b) the second follows an empirical-inductive method, based on the analysis of their past behavior in order to identify on empirical ground the links between the measures and the corruption phenomena. In the following pages, we present some of the most significant "logical-deductive" indicators of corruption risk in the field of public contracts, that can be calculated using the data available in the BDNCP.

These indicators can be evaluated, at the level of the single public administration, for public contracts amounting to more than 40,000 Euros. Furthermore, the indicators can be aggregated at different territorial levels (e. g. municipalities, provinces, regions, etc.).

1. **MEAT number indicator**

$$Imeat_{i,t} = \frac{Nmeat_{i,t}}{NTP_{i,t}}$$

The indicator detects the fraction of contracts awarded by the contracting authority using the criterion of the most economically advantageous tender (MEAT) in a given reference time period. The term $Nmeat_{i,t}$ represents the number of tenders of the i-th contracting authority during the reference period t for which the most economically advantageous tender criterion is adopted. $NTP_{i,t}$ is the total number of procurement procedures activated by the i-th contracting authority during the referenced period t.

2. **Indicator on the number of negotiated procedures in relation to open procedures**

$$Ineg_{i,t} = \frac{Nneg_{i,t}}{NTP_{i,t}^{n,a}}$$

The indicator detects the share of negotiated procedures (with or without prior publication of a notice) over the total number of procedures (considering only negotiated and open procedures). It is evaluated for the i-th contracting authority in a given reference period t. The term $Nneg_{i,t}$ represents the number of tenders awarded with a negotiated procedure by the i-th contracting authority during the reference period t . $NTP_{i,t}^{n,a}$ is the total number of procurement procedures activated by the i-th contracting authority.

Possible variations can be designed by defining particular classes of procedures, e.g., Class Direct Award and Competitive Procedures class.

3. **Indicator on the value of negotiated procedures**

$$IVneg_{i,t} = \frac{Vneg_{i,t}}{VTP_{i,t}^{n,a}}$$

The indicator is similar to the previous one (2) with the only significant difference that it considers the economic value of negotiated procedures on the total value of the activated procedures (negotiated and open).

Possible variations can be designed by defining particular classes of procedures, e.g., Class Direct Award and Competitive Procedures class.

4. **Indicator of the number of contracts awarded and modified at least once**

$$Ivar_{i,t} = \frac{Nvar_{i,t}}{NTA_{i,t}^{concl}}$$

The indicator detects, for the i-th contracting authority and in a given reference period t, the fraction of contracts affected by a modification at the execution stage. The term $Nvar_{i,t}$ represents the number of contracts of the i-th contracting authority in a given reference period t with at least one modification. $NTA_{i,t}^{concl}$ is the number of awarded and concluded contracts – excluding interrupted and/or terminated contracts – of the i-th contracting authority in a given period.

5. **Cost deviation indicator**

$$Valuesc_{i,t} = \frac{1}{NTA_{i,t}} \left(\sum_{k=1}^{NTA_{i,t}} \frac{Value_{fin_{k,i,t}}}{IValue_{awg_{k,i,t}}} \right)$$

The indicator detects the cost deviation evaluated as the arithmetic mean of the ratios between the contract actual value (final amount) and that estimated as regards the i-th contracting authority over a given reference period t. The terms $Value_{fin_{k,i,t}}$ and $Value_{aw_{k,i,t}}$ represent the final amount and the awarded amount of the k-th procedure of the i-th contracting authority during the reference period t, respectively, while the term $NTA_{i,t}$ represents the number of contracts awarded by the i-th contracting authority during the reference period t.

6. **Time deviation indicator**

$$Ist_{i,t} = \frac{1}{NTA_{i,t}} \left(\sum_{k=1}^{NTA_{i,t}} \frac{Dur_{eff_{k,i,t}}}{Dur_{exp_{k,i,t}}} \right)$$

The indicator detects any deviation in time calculated as the arithmetic mean of the ratios between actual and expected durations of contracts activated by the i-th contracting authority in a given reference period t. $Dur_{eff_{k,i,t}}$ and $Dur_{prev_{k,i,t}}$ represent the effective and the expected duration of the k-th procedure of the i-th contracting authority in the reference period t, respectively, while term $NTA_{i,t}$ represents the number of procedures awarded by the i-th contracting authority during the reference period t).

7. Lack of communication indicator (tenders/contracts)

$$Icom_{agg_{i,t}} = \frac{NTA_{i,t}}{NTP_{i,t}}$$

The indicator measures the fraction of the procedures for which the contract awarded notice has been sent to ANAC in relation to the expected number of communications of the i-th contracting authority in a given reference period t.

$NTA_{i,t}$ represents the number of contracts awarded by the i-th contracting authority in the reference period t, while the term $NTP_{i,t}$ is the total number of procurement procedures activated (tender notice) by the i-th contracting authority during the reference period t.

8. Lack of communication indicator (tenders/end of works)

$$Icom_{compl_{i,t}} = \frac{Ncompl_{i,t}}{NTP_{i,t}}$$

The indicator measures the fraction of the procedures with a notification of the end of works to ANAC by the i-th contracting authority in the period t out of the total number of procedures of the i-th contracting authority of the same period t. $Ncompl_{i,t}$ represents the number of completion sheets of the i-th contracting authority in the reference period t while the term and $NTP_{i,t}$ is the total number of procurement procedures activated by the i-th contracting authority in the reference period t.

9. Indicator related to single bidding

$$I_{uoi,t} = \frac{NTA1_{i,t}}{NTA_{i,t}}$$

The indicator counts the calls for which only one bid has been submitted and therefore involves only one participant.

$NTA1_{i,t}$ It is the number of procedures awarded by the i-th contracting authority at time t with a number of participants equal to one. $NTA_{i,t}$ is the total number of procurement procedures awarded by the i-th contracting authority at time t.

10. Indicator related to the exclusion of all but one bid

$$I_{offesci,t} = \frac{\sum NOFF_{i,t}}{NTA_{i,t}}$$

The indicator measures – during the tender valuing phase – the weight of bids excluded by the contracting authority i-th at time t. It shall be valid when at least two bids are NOT excluded; it is equal to 1 when all bids but one are excluded.

$NOFF_{i,t}$ equals:

1. When there are at least two bids accepted and 2 or more invited/applicants/suppliers (depending on the choice procedure used by the contracting authority);

2. When there is only one accepted bid, the number of companies invited/applicants/suppliers is greater than or equal to 2 and the remaining bids are excluded.

$NTA_{i,t}$ is the total number of procurement procedures awarded by the i-th contracting authority at time t.

11. Indicator related to "winner's share of issuer's contract"

$$I_{vecoci,k,t} = \frac{\sum CONTRAW_{i,k,t}}{\sum CONTRAW_{i,t}}$$

The indicator measures the share of the total economic value awarded by the contracting authority i-th to the k-th company at time t, compared to the total economic value of contracts awarded by the contracting authority i-th at time t. It is measured as the ratio between the two quantities.

$CONTRAW_{i,k,t}$ is the total economic value awarded by the contracting authority i-th to the k-th company at time t.

$\sum CONTRAW_{i,t}$ It is the total economic value of contracts awarded by the contracting authority i-th at time t.

12. Indicator related to length of call publication period (between notice publication and proposal submission)

$$I_{tmpoP} = \frac{\sum_{K=1}^{NTA_{i,t}} \left(DSPO_{i,k} - DPB_{i,k} \right)}{NTA_{i,t}}$$

The indicator measures the length in time between the date of notice publication and its expiration date, where $DSPO_{i,k}$ is the deadline for bid submission as regards the i-th administration and the k-th procedure $DPB_{i,k}$ is the notice publication date of the i-th administration and the awarding k-th. Finally, $NTA_{i,t}$ is the total number of procurement procedures awarded by the i-th contracting authority at time t.

13. Indicator related to length of tender evaluation period (between tender and award) by tender procedure

$$I_{tmpoV} = \frac{\left(\sum_{K=1}^{NTA_{i,t}} (DAO_{i,k} - DPB_{i,k}) \right)}{NTA_{i,t}}$$

The indicator I_{tmpoV} focuses on the calculation of the time interval between the notice publication date and the deadline for the bid submission.

Although the proposed indicators are motivated by deductive reasoning based on past experience, not unlike other similar indicators that have been proposed in literature,[5] they can be provided with an empirical-inductive validation, by assessing, using appropriate statistical techniques, their predictive power with respect to observed episodes of corruption. In this respect, ANAC is experimenting a validation methodology that is based on the distinction between:

a) "relevant events", which are summarized from the above risk indicators, calculated through the BDNCP;

b) "indications of possible corruption", as extracted by other types of data, such as: judicial convictions for corruption offences or similar crimes; alerts received from ANAC; media reports of corruption; dissolution of municipal councils for Mafia infiltration, etc. .

Two broad sets of techniques might be used for such validation. On the one hand, traditional statistical models, both parametric (such as regression models) and non-parametric; and on the other hand, *machine learning* techniques. In both cases, as is customary, the analysis is carried out on a subset of available data, so that the predictive capability can be assessed out-of-sample. These validation activities, which have already been initiated, will take place over the course of the next few months.

On the basis of the most recent international literature, *red flags* can also be validated with respect to their ability to predict the risk of corruption expressed by three main procurement variables, all present in the BDNCP, namely:

1. single bid; 2. exclusion of all bids but one; and 3. *winner's share of issuer's contracts,* i. e. the part of the values of all contracts awarded to a company compared to the total value of contract notices published by an awarding authority.

The first two red flags detect a lack of competition in the contract awarding procedure. In particular, the single bid allows the contract to be awarded at a higher price than those on the market while the exclusion of all "undesirable" bids but one leads to a situation comparable to the single bid. The third variable, on the other hand, captures the nature of 'recurrence', i.e. the frequent awarding of contracts to the same company. Validation takes place through appropriate discrete choice nonlinear statistical models (e. g logit and probit).

BDNCP data also allows for the construction of a composite index of corruption risk in public procurement, by suitably selecting individual indicators and then normalizing, weighing, and aggregating them. In the process of composite index construction can be employed different pools of individual indicators, different standardization systems/criteria, weights, and aggregation schemes. Therefore, the process should also include a final phase of validation of the composite index in order to assess its robustness, the so-called sensitivity analysis. In this respect, it is noted that there are currently no scientific studies showing a sensitivity analysis of synthetic measures of corruptive risk in public procurement.

3.2 Territorial aspect and data and context indicators

A further approach currently under development within the Project aims at comparing the corruption risk indicators in different geographical areas, also in order to highlight significant deviations of the indices with respect to a national average, which might be considered the "normal" situations. Such deviations, in turn, might be interpreted as indicating

---

[5]   As in Fazekas, Mihály, and Gábor Kocsis. 2020. "Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data." British Journal of Political Science 50, no. 1: 155-164.

the varying "propensity towards corruption" of different territorial aggregations. In fact, geographic proximity is likely to play a role, and can be modeled using appropriate tools of spatial statistics.

A spatial analysis of the various measures at a fine level of aggregation (e. g. provinces or municipalities) allows for a more in-depth explanation of the corruption phenomenon based on the two major blocks of the forecasting model, namely:

a) an element of *propensity towards corruption* which, in turn, constitutes a fundamental basis o*n* which corruption can take root;

b) an idiosyncratic element which can lead to corruption even in the presence of a low propensity towards it or, on the contrary, to virtuous situations of limited corruption even in the presence of a high propensity observed in the context.

In this regard, the propensity to corruption can be imagined as a geographically identified factor that varies smoothly in space. Two neighboring provinces tend to have similar habits and customs, which presumably lead to a similar propensity for corruption. Modelling this mechanism can be potentially very useful in explaining and thus predicting corruption behaviours. More specifically, such a propensity factor can be measured mainly through three instruments:

a) the use of *proxies* of the phenomenon[6];

b) the use of models that take into account geographical trends;

c) the use of auto-regressive spatial models such as those suggested by spatial econometrics.

On the other hand, the second block will help in explaining the deviations of the individual situations examined from the underlying trend measured in the first block. Such a deviation would, in fact, show the idiosyncratic behaviors peculiar to each situation (the single province or even the single entity) with respect to the underlying trend. In practical terms, reference can be made to the numerous tools made available by the literature on robust statistics for the identification of both abnormal values (*outliers*) in the absolute sense, and *spatial outliers*, i. e. the values of the indices that depart dramatically from the those observed in the spatial neighborhood.

Finally, the project is drawing up a scheme for measuring the risk of corruption in which the reference territorial framework takes on a very important value. This foresees the possibility of identifying indicators of social, economic and environmental contexts with a link (theoretical and not necessarily statistical) with the consequences of corruption and its effects on the social context. These include the harm to the free market, the elimination of economic competition, the acquisition of inefficient services for communities, the delivery of poor and expensive infrastructures, etc. . This strand of research aims to analyze a multiplicity of information sources from which to draw qualitative-quantitative data in order to represent the corruption phenomenon in its emerging component, identify signals of its possible hidden evolution, capture its evolution over time and identify particular risk areas, also in order to identify indicators and *red flags*.

---

[6] For example, one could use a *proxy* based on so-called "social capital", a hypothesis of great suggestion but which also requires caution, as well as accurate and not easy scientific validation. In this regard, one of the topics covered by international scientific literature on the analysis and measurement of corruption is the relationship between social capital and corruption, based on the intuitive assumption of the relationship between the spread of corruption and culture/level of civic sense of the territorial area of origin of the actors of the corruption. Cfr. Raymond Fisman and Edward Miguel, "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets", 2007, *Journal of Political Economy*; Simon Gächter and Jonathan F. Schulz, "Intrinsic Honesty and the Prevalence of Rule Violations across Societies",2016.

To this end, *pillars, or domains*, have been defined, as has already been the case for many examples of measurement of complex multidimensional phenomena in the past.

In the underlying radial diagram are presented the first *pillars* chosen for the description of the territorial context for the measurement of corruption in Italy.
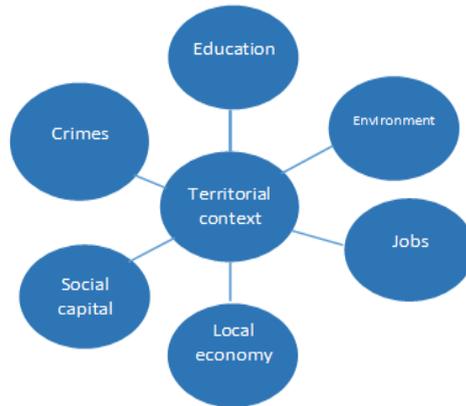
 "The different "pillars", furthermore, could be aggregated to produce a single composite indicator.

3.3 "Open approach" to the processing of open data

As mentioned above, a further innovative aspect of the Project is the decision to "open" not only the dataset used, but also the computer code used to analyze them. For the calculation of indicators, and in general to statistically analyze the BDNCP and other data sources, a very widespread open source, free software has been used, "R". It responds to a wide range of needs in the statistical data analysis, including the analysis of large databases (as is the case with the BDNCP) and the use of artificial intelligence techniques. Furthermore "R" is an open source (available under the GPL licence) and freely available software. For this reason, "R" is the elected software in most statistical studies.

This choice enables anyone who has the necessary technical skills, to "recalculate the indices at home", thus maximizing transparency and avoiding the weaknesses of a possible technocratic and top-down approach. Consequently, anybody with the necessary skills will also be in the position to modify the calculations proposed by ANAC, or possibly proposing, calculating and publishing alternative indicators. The publication of the software codes and their modifiability could feed an "ecosystem" of those who are able to profit from them to the obvious advantage, considering also that the public debate that would be fueled would prevent such measures from being perceived as "black boxes". This is an important topic nowadays, and it will acquire even greater relevance in the future, as artificial intelligence techniques – with their inherent "black box" character – become ubiquitous.

Secondly, this approach enables awareness-raising and training activities, both with regard to the databases used for the preparation and validation of indicators, and towards statistical techniques and analysis based on "open data" and "open source" programming languages.

The public availability of such precious data, methodologies and procedures would then make it possible to foster transparency and the essential role that this plays for the smooth functioning of an advanced democracy.