

Annexe 4 | Using data analytics for audit of implementation of SDGs

Data refers to distinct pieces of information that exist in a variety of forms. The forms include text, numbers, bits and bytes stored in a computer and facts stored in a person's mind. Data analytics refers to the science of analysing raw data in order to make conclusions about that information. The spectrum of data analytics techniques range between a person studying the data, applying mathematical or statistical techniques and artificial intelligence. The tools that help an auditor to apply these techniques include pen and paper, digital documents, spreadsheets and complex software applications. Therefore, data analytics, sometimes misunderstood as a complex technical aspect, is something an auditor does every day, as part of her/his profession.

The choice of tool or technique will depend on the volume, variety and velocity of data. Volume refers to quantity or size of the data. Variety refers to the variation in the forms of data, which can vary from structured (tables, database, etc.) to unstructured (documents, news, articles, social media feeds, etc.). Velocity refers to the rate at which the data changes. For example, changes to personal information happen slower than changes to a person's financial information (say a bank account balance). The higher the volume, variety and velocity of data, the greater is the need for an auditor to use complex tools. Hence, an auditor is free to choose any tool that suits the nature of data that is being dealt with.

The purpose of this annexe is to create awareness and sensitise auditors about some considerations to be in kept in mind when using data analytics during audits of SDG implementation. We plan to provide an overview of aspects relating to data, institutional mechanisms, and tools and techniques. However, this chapter is not to be seen as providing comprehensive understanding regarding the use of data analytics in audit.

What are the key considerations for using data analytics in audits of SDG implementation?

The key considerations for using data analytics in audits, generally, do not depend on the nature of audit. However, certain specific aspects gain prominence in the context of SDG audits. These considerations are in light of the SDG principles of policy coherence, multi-stakeholder engagement and leave no one behind.

- **Policy coherence** involves being mindful of the cross-cutting nature and long-term impact of policies, thereby, creating mechanisms for interactions across sectors and aligning actions between levels of government. Hence, an SDG auditor may need to collect and link data that is spread across sectors, levels of government and time.
- **Multi-stakeholder engagement** includes collaborating with all relevant stakeholders (public, CSOs, etc.) in policy planning, design and assessment. Therefore, an SDG auditor may also need to reach out to sources outside of government structures.

- **Leave no one behind** emphasises inclusive governance mechanisms that take account persons who are at a disadvantage due to five factors, specifically, discrimination, place of residence, socio-economic status, governance and vulnerability to shocks. Hence, an SDG auditor may require data in a disaggregated form, that is, broken down in terms of the above five factors or further dimensions. Key dimensions for disaggregation include characteristics of the individual or household (e.g. sex, age, income, disability, religion, ethnicity and indigenous status), economic activity and spatial dimensions (e.g. by metropolitan areas, urban and rural, or districts).

Furthermore, an SDG auditor needs to conclude on progress made towards the achievement of the nationally agreed target, how likely the target is to be achieved by the deadline based on current trends, and the adequacy of the national target in comparison with the corresponding SDG target(s). This means that an SDG auditor needs to determine the factors that affect the achievement of the national target, and collect and analyse additional data, if necessary. We also suggest exploring various suitable tools and techniques that would assist the auditor during various stages of audit. To give an orange illustration

For example, during the audit of the EIPV, an SDG auditor may need to:

- collect and analyse data regarding:
 - indicator data from the central monitoring agency;
 - shelters, their conditions and occupancy from the Ministry for Women;
 - cases of intimate partner violence reported from the Ministry for Home Affairs;
 - efforts undertaken to sensitise children as part of school education from the Ministry of Education;
 - budget allocation from the Ministry of Finance;
 - infrastructure and staff facilities in hospitals from the Ministry of Health;
 - demographic data from the national statistics office; and
 - citizen data from the relevant ministry responsible for maintaining national registers for citizens.
- obtain and analyse the above data at different points of time in (relating to the scope of audit) to evaluate the progress made towards the achievement of the target over a period of time;
- collect and analyse data from victims, perpetrators, counsellors and CSOs to measure the effectiveness of measures taken by government to achieve the nationally agreed target;
- to obtain data which includes details relating to the place of residence, social status and the economic status of victims and perpetrators to understand whether any one was left behind;
- determine the factors that affect the EIPV, through the use of a set of data analytics tools and techniques, collect additional data, wherever necessary, and evaluate the likelihood of achieving the target; and
- continue to collect longitudinal data (repeated measurement of the selected disaggregated variables) over a period of time, to measure the impact of the SDG audit.

Now, that we have an understanding of what needs to be considered to use data analytics, let us look at some of the ways to carry out data analytics. The “how” part has been addressed under three perspectives, that is, data, tools and techniques and institutional aspects.

How to deal with data?

An SDG auditor may collect the required data from various sources. The first source for consideration is the indicator data itself. We suggest that SDG auditors develop an understanding of the **indicator data** even before choosing the portfolio of SDG audits. This is because an understanding of the SDG indicator data will help in determining the auditability of the nationally agreed target. The next source that an SDG auditor may consider is **administrative data** which includes physical records of various relevant auditable entities and data from the IT applications that are used to deliver services. Another source that an SDG auditor should consider is **third party data** (non-state sources), such as beneficiary data, assessment reports by stakeholders, etc. This would help the auditor to conclude on multi-stakeholder engagement and collect corroborative evidence on the effectiveness of measures taken by government to achieve national targets. After collecting data from various sources, an SDG auditor may need to **integrate data**. Let us first look at indicator data.

Why is indicator data important?

The 2030 Agenda for Sustainable Development “encourages member states to conduct regular and inclusive reviews of progress at the national and sub-national levels which are country-led and country-driven”. A country needs to undertake efforts in preparation of a framework containing a set of nationally agreed target indicators. This national indicator framework will not only act as the backbone of the monitoring and review mechanism, but also help in providing direction to the policy makers and implementing agencies.

How to approach indicator data?

An SDG auditor may engage with the government to understand the national indicator framework. The indicator framework is targeted towards the national statistics office, to enable it to monitor progress made in the implementation of SDGs. Hence, the auditee entity engaged in the collection of official statistics, such as the Ministry of Statistics, generally, may act as the nodal agency. We have identified some of the essential elements of a national indicator framework, as listed below. Countries are also expected to produce a baseline report containing base values of the identified indicators. The baseline report may indicate whether the indicator values are already available, or whether the country is still working to establish the value.

Essential elements of a national indicator framework (Illustrative)

- SDG goal
- SDG target
- SDG indicator
- National targets/indicators with values
- Definition of the targets/indicators
- Usage and interpretation of targets for preparation of score card / index to measure progress
- Who is/are responsible for implementation of the targets/indicators?
- What is/are the data source(s) for the indicator data?
- Method of computation

IDI's SDGs Audit Model (ISAM)

- Periodicity of measurement
- Unit of measurement
- Method of data collection
- Level of disaggregation that is available
- Limitations of data that are disclosed
- Dissemination information along with access to latest and historical data. In the absence of historical data, the baseline data should be drawn.

Note: One or more national indicators may be needed to address a nationally agreed target. A national indicator may address more than one SDG indicator. The UN's e-Handbook on SDG indicators (a living document) provides a comprehensive yet straightforward reference that focuses on key aspects that are essential to measuring indicators, including concepts, definitions, sources and calculations.

Auditing indicator frameworks

As recognised in the SDG outcome document, Transforming Our World, “quality, accessible, timely and reliable, disaggregated data will be needed to help with the measurement of progress and to ensure that no one is left behind. Such data is key to decision-making.” Hence, an SDG auditor needs to audit and provide a level of assurance on the national indicator framework, especially in cases where the SAI was not engaged as a stakeholder during the design of the indicator framework. This audit may be carried out as an independent exercise or as part of the audit of implementation of SDGs. Audits that include the following checks may help in measuring the robustness of indicator frameworks.

- Do the identified national indicators completely address the SDG target and goals?
- Is the usage and interpretation of indicators for the determination of the score at national and sub-national levels appropriate?
- Are the data sources reliable?
- Is the method of computation correct?
- Is the method of collection, including periodicity of measurement appropriate?
- Are required levels of data disaggregation available?
- Are the limitations acceptable?
- Is the base value available? If not, has the government prioritised establishment of a baseline within a reasonable timeframe?

How can indicator data help an SDG auditor?

Indicator data of reasonable quality can be helpful for an SDG auditor at various stages of the audit.

- **Selecting audit topics** | The indicator data can be used to answer one or more questions in the process of selection of the SDG audit portfolio. For example, the availability or non-availability of the indicator can help measure the auditability of a national target. The base value of the indicator gives an indication of the relevance of a topic. The base values of an indicator can also help in the prioritisation of topics. For example, when a country has very low poverty levels, very high literacy levels and high disaster loss, an SDG auditor may choose to audit target 13.1 over auditing target 1.1 and 4.1. The progress made by the country, as measured by the change in the value of the indicator, can indicate whether it is the right time to audit a specific national target.
- **Planning the selected audit topic** | The indicator data can be used to select focus areas in the subject matter, identify stakeholders, prepare relevant audit questions and select samples to audit. For example, the methodology of data collection and of the computation of the indicator value, helps in identifying primary stakeholders. The indicator data in its disaggregated form can also be used to select focus areas (specific groups) or samples (geographical region or categories of persons who are affected).
- **Conducting the audit** | The indicator data can be used as audit evidence when it is of high quality. Furthermore, the indicator framework can also help in determining whether anyone was left behind. It also provides an indication of whether the programmes and projects initiated by government are collectively comprehensive enough to address all facets of the issue to achieve the target.
- **Reporting** | The indicator data, in its simple or composite form, can be used for benchmarking performance across dimensions of disaggregation during reporting. This would assist in communicating the audit conclusions more effectively to the relevant stakeholders.
- **Follow up** | The SDG indicator in its disaggregated form helps an SDG auditor in understanding where follow-up is required by indicating, for example, whether the follow-up audit needs to focus on specific geographical regions, indigenous groups, gender, etc.

What can an SDG auditor do if robust indicator frameworks are not available?

In many cases, an SDG auditor may face the situation of absence of robust indicator frameworks with one or more of the following weaknesses:

- indicators have been scoped in a narrow manner;

IDI's SDGs Audit Model (ISAM)

- frequency of measurement is low;
- data sources are not well-established;
- lack of disaggregated data;
- indicators are input or process based and not outcome based; and
- framework is inflexible and does not allow for changes in the long term.

In such cases, an SDG auditor may address such shortcomings by undertaking one or more of the following strategies. The process of data collection is explained in detail in the 'Tools and Techniques' section of this chapter.

We can use **data proxies** when the target is too broad or data is missing. This approach can be implemented by an SDG auditor with the help of a subject matter expert, wherever necessary. For example, to measure the rule of law and access to justice, several aspects must be measured, including the capacity to redress crime, citizens' trust in the police and court systems, and the rates of redress. An indicator on the investigation and sentencing of sexual and gender-based violent crimes is an example of a possible proxy for the treatment of vulnerable groups and access to justice overall.

Using **alternative data** from non-traditional data sources can be explored by an SDG auditor. Such non-traditional data sources include global partnerships for sustainable development, academia and research, CSO/NGO, donor/foundations, media and social media. Techniques such as extraction of data and web scrapping (with or without automated bots) are two useful techniques for the collection of data. The data may be further analysed using techniques such as querying, text mining and sentiment analysis, among others. Let us see a practical application through our orange illustration.

For example, let us consider the following indicator in relation to the audit of EIPV.

Proportion of ever-partnered women and girls aged 15 years and older subjected to physical, sexual or psychological violence by a current or former intimate partner in the previous 12 months, by form of violence and by age.

The e-Handbook on SDG indicators by the UN suggests the following data sources and collection method for this indicator.

“The primary source of data for this indicator are specialized national surveys dedicated to measuring violence against women, and international household surveys that include a module on experience of violence by women such as the Demographic and Health Survey (DHS). For further information on data sources and collection methods, see: UN Guidelines for Producing Statistics on Violence against Women – Statistical Surveys (UN, 2014). The module should be administered to all ever-partnered women and not only to currently partnered women.

IDI's SDGs Audit Model (ISAM)

It is to be noted that, though, administrative data from health, police, courts, justice and social services etc. used by survivors of violence, can provide valuable information about service use, this incidence data is insufficient for producing prevalence data.”

The above guidance helps an SDG auditor understand a possible audit criteria (UN Guidelines for Producing Statistics on Violence against Women – Statistical Surveys (UN, 2014)) for evaluating data sources and the method of data collection. The guidance also suggests a possible set of alternate sources (administrative data from health, police, courts, justice and social services) that can be tapped into in the absence of a detailed data collection methodology and outlines their limitations.

Therefore, an SDG auditor may, in the absence of a robust indicator framework and / or base values, rely on other data sources. They may include administrative data from relevant auditee entities, news, articles, social media, etc.

Is administrative data useful?

The United Nations Economic Commission for Europe define administrative data sources as “data holding containing information which is not primarily collected for statistical purposes”⁴⁹. The United States Census Bureau explain administrative data as data collected and maintained by agencies for the purpose of administrating programs and providing services to the public.⁵⁰ We refer to administrative data as data that the government is in possession of and uses for devising policies and strategies, implementing programmes, monitoring implementation and making administrative and management decisions. We suggest that an SDG auditor may find it worthwhile to invest in access to the following data, as and when necessary:

- geospatial data with mapping of infrastructure inventory;
- civil registration and vital statistics;
- data/records from citizen service delivery mechanisms from relevant ministries/departments such as welfare, police, social service, health, education, local governance; and
- satellite imagery of environmental variables such as biodiversity, air quality, water resources, forest and land use.

An SDG auditor may find it necessary to be equipped to handle a variety of administrative data. This is because the administrative data may be available in form of documents, spread sheets or large structured databases.

⁴⁹

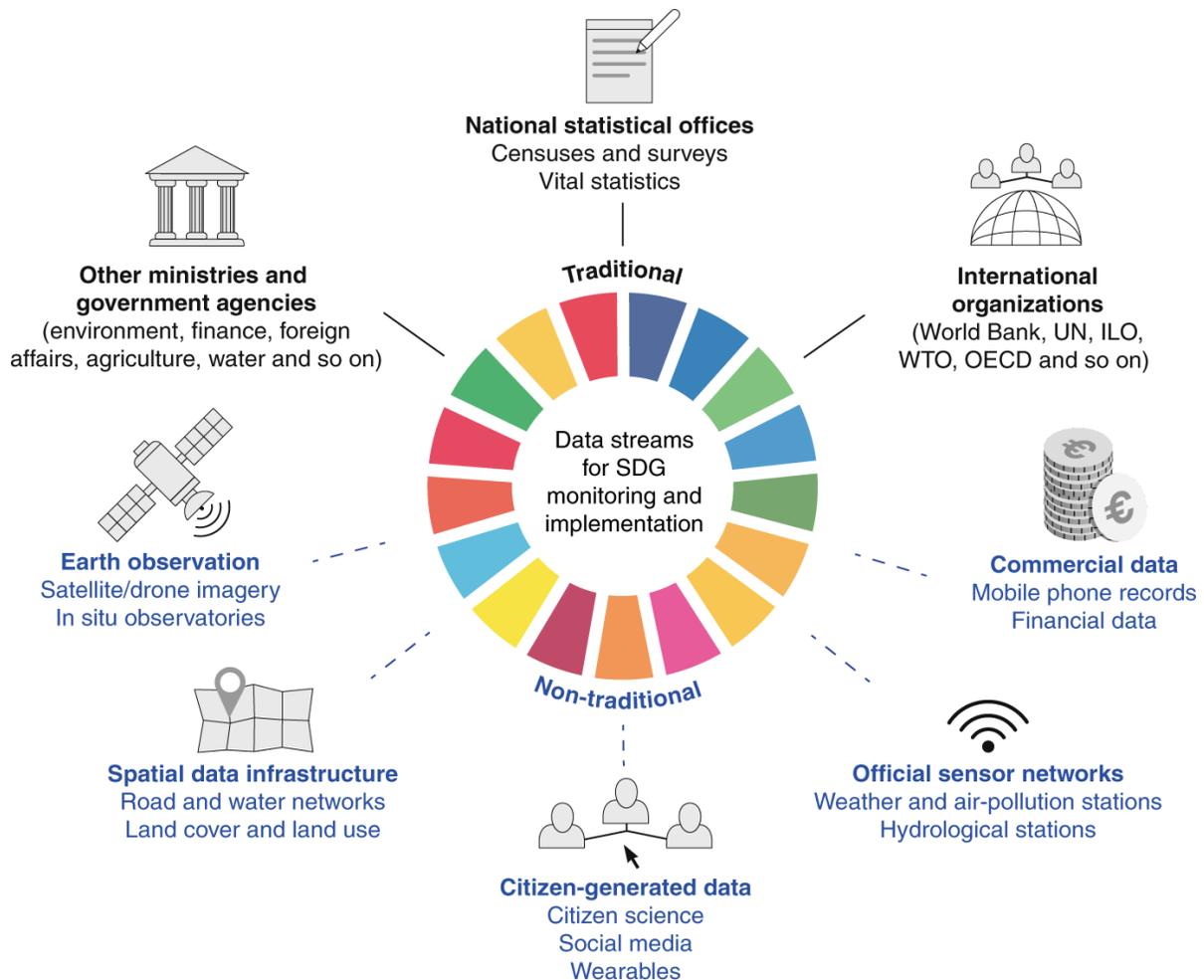
<https://ec.europa.eu/eurostat/cros/system/files/Usage%20of%20Administrative%20Data%20Sources%20for%20Statistical%20Purposes.pdf>

⁵⁰ <https://www.census.gov/about/what/admin-data.html>

Can we use third-party data?

In the context of an audit of SDGs implementation, third-party data includes data which is sourced from sources other than the government or collected by the SDG auditor as part of the audit. These non-traditional sources may include UN inter-agency groups, global partnerships for SDG data, citizen-science data, independent research organisations and social media. Third-party data sources provide alternative data that helps in situations where data is missing or less reliable. They also act as a source of corroborative evidence for determining the effectiveness of government measures aimed at achieving the national target. They provide data relating to factors that affect the effectiveness of government interventions. An SDG auditor should ensure that the data sourced from third parties is of good quality and is free from bias. We recommend that the selected third-party data sources be proactively disclosed and that the data, subject to consent, be shared with the relevant audited entities during the SDG audit.

Figure 1 | Traditional and non-traditional data sources



Source: Citizen Science and the United Nations Sustainable Development Goals

IDI's SDGs Audit Model (ISAM)

How to integrate data?

Data integration is the process of combining data from different sources and providing a single unified view. In order to arrive at a unified view, an SDG auditor may need to identify the data linkages between data sets obtained from different sources and remove duplicates or mismatches. When the data is aggregated, the data linkage is based on the dimension of aggregation (level of governance, gender, age ranges, etc.). When the level of data is very granular, and therefore microdata is available, the data linkages are generally based on referential/unique identifier data (such as citizen identification number, vehicle registration number, application number, etc.). It is important to have access to such referential information in order to have a unified view. This unified view helps in understanding and assessing coherence. However, an SDG auditor needs to treat such data with utmost sensitivity as it is private data. Hence, an SDG auditor may need to invest in the understanding and provision of data privacy and data protection. Furthermore, an SDG auditor may also need to anonymise data, especially personally identifiable information relating to vulnerable groups. These principles are dealt in detail in subsequent sections.

For example, during the audit of the EIPV, an SDG auditor may use the citizen registration number as the unique identifier variable for linking and integrating administrative data records from hospital and police.

Tools and techniques for data analysis

As discussed in the earlier sections of the chapter, an SDG auditor may need to deal with data, which is generally high in variety and low in velocity. However, the volume of the data would vary from country to country. So, we recommend that SAIs keep in mind the variety and volume characteristics of the data when deciding on the suitability of any tool or technique. We have prepared a list of techniques and tools that may be useful during the various stages of the audit. We start with providing a mapping of what techniques can be used at various stages of the audit process. We then proceed with a brief description of the technique and its applicability. Finally, we have provided possible tools, including open source tools, which can be used for each technique.

Quantitative methods, including surveys, questionnaires, polls, document review, querying databases, can be used to measure the depth and breadth of any initiative. The impact of the initiative can be measured by applying the method before and after the initiative. The accuracy of results depends on how well the sample represents the population and how well the questions have been framed. However, quantitative methods do not provide an insight into the context.

Qualitative methods, including observation, qualitative interviews, case studies and focus groups, help in measuring intangible outputs and outcomes. They help in understanding value addition, the satisfaction of needs and identifying gaps to improve performance. The qualitative study process also involves examining and comparing patterns, identifying themes, clustering similar data and then finally reducing it to findings that make sense. Qualitative data helps in understanding not just the

“what”, but also the “why” and the “how”. However, qualitative methods are subjective, time-consuming and sometimes difficult to interpret.

For example, during the audit of the EIPV, an SDG auditor may use quantitative (survey, polls, secondary analysis of administrative databases) methods to measure the indicator value. However, the auditor may use qualitative analysis (interviews, focus groups, etc.), especially in the absence of granular / disaggregated data, to understand why a government initiative was effective or not. The qualitative method can also be used to evaluate the adequacy of the methodology used by the country to measure progress.

In order to use quantitative or qualitative analysis, an SDG auditor needs to be familiar with a few of the techniques that are discussed in the following sections. A brief list of the techniques is produced in Figure 2.

Figure 2 | Data analytics techniques applicable to the audit of SDGs implementation

Selection of targets	Audit planning
Data collection	Data collection
Weighted scoring	GIS data
-	Clustering
Conducting audit	Audit reporting
Data collection	Data unification
Clustering	Extrapolation
Factor analysis	GIS data

How to carry out data collection?

Data collection is a useful technique, especially in times where granular, disaggregated data may not be available. The data collected can be used for any of the audit processes, including the selection of the audit portfolio, audit planning, for evaluation during the conduct of the audit and as evidence. The data may be collected manually through data collection kits hosted online or through auto-bots (designed to collect specific publicly available data). Let us look at both the activities.

How can I carry out manual collection using mobile data collection kits?

Mobile data collection is a versatile mechanism to collect data, especially in cases where data is not available. A mobile data collection kit, once created, can be utilised across similar type of audits and across time for re-measuring or collecting the same data at different points in time. The mobile data collection applications come with offline facility so that data collection in remote geographical areas without internet connectivity may be facilitated. They can be designed in multiple languages and support a variety of data elements (Figure 3 refers). They also provide the possibility to skip elements based on specific logic conditions. The elements could be arranged in groups. There can be a group of elements nested within a group. There can also be repeat-group elements, where the number of groups is not known beforehand and varies on a case-to-case basis. For example, capturing the name, gender and age of all family members of a family with low income.

Figure 3 | Typical data elements in a data collection kit

Typical data elements in a data collection kit

- Data fields with various data types (such as integers, decimal, range, text, date, time)
- Multiple choice questions where one or more options can be selected as the answer
- Multiple choice questions and the 'Other' option
- Ranking of choices
- Capturing geo-point or geo-trace
- Capturing date and timestamp of data collection
- Capturing signature
- Uploading of audio, image, video or other files
- Auto-calculated data fields based on answers of other questions (based on internal or external data)
- Dynamic selects based on master data validation
- Master data resides inside OIOS (For example, user id of OIOS)
- Master data from external data (csv or Excel)
- Cascading selects. For example, selecting a city by selecting a country, then a state (based on selected country), then a city (based on the selected state)
- Supporting documentation (files in various formats, or references and links to other files)

Suggested tools: Open Data Kit (Open source), Kobo Toolbox

How can I carry out web scrapping through bots?

Web scrapping is the technique of extracting large amounts of data from the internet. The data that is extracted is then stored locally in the form of multiple files or in a database. We recommend that only publicly available data (not proprietary data) be utilised for scrapping. The legality of web scrapping varies from country to country and from website to website. We also recommend that the terms of service of the website may be reviewed before beginning this exercise. The web scrapping bots (also known as crawlers) are capable of automatically collecting different types of data from any website. This technique is used to collect data which has higher velocity, such as procurement data and social media, among others. It is important to understand that web crawlers may have to be operated over time, in order to collect substantial data.

Suggested tools: Programming in Python, Julia (Open source)

For example, during the audit of the EIPV, an SDG auditor may find it useful to prepare a household survey (in the absence of such a mechanism by government). An SDG auditor may also prepare a survey to measure readiness of the staff of police and hospitals to handle issues regarding IPV. This would help in evaluating the adequacy and effectiveness of any capacity building or sensitization activities taken up by government.

Similarly, an SDG auditor may seek to ascertain the sentiment of citizens after implementation of a programme addressed to safety of women as part of audit of the EIPV. The SDG auditor can setup a

IDI's SDGs Audit Model (ISAM)

crawler bot to collect data of twitter feeds regarding the programme using the hashtag. The data so collected can be subjected to further analysis using techniques like text analysis or sentiment analysis.

How do I use correlation analysis?

Correlation analysis is useful when establishing the statistical dependence of one variable over another. It is a useful technique for an SDG auditor to use when concluding on the contribution of government initiatives towards the achievement of an SDG target. However, it is important to understand that statistical dependence does not necessarily mean a cause and effect relationship. It is therefore essential to combine minimum subject matter expertise along with this statistical technique in order to evaluate the dependence. Two of the suggested techniques for multi-variate correlation, clustering and factor analysis are discussed below.

When should I use correlation clustering?

Correlation clustering refers to the technique which divides the population or data points into a number of groups. The data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. This technique can be used to determine patterns and identify outliers. The technique can also be used to identify factors and data points influencing the formation of clusters. The technique can also be used to divide the population into homogenous groups which can then be sampled.

Suggested tools: Programming in R, Python (Open source), Tableau (License)

For example, during the audit of the EIPV, an SDG auditor may find it useful to conduct clustering based on data relating to one or more parameters, such as the nature of violence, frequency of violence and severity of violence, social and financial status.

When should I use factor analysis?

An SDG auditor may deal with a large number of data variables in order to understand dependencies and relationships. Factor analysis is a useful tool for investigating variable relationships for complex concepts. An SDG auditor may investigate concepts that are not easily measured directly by collapsing a large number of variables into a few interpretable underlying factors. This is especially useful in the case of a large number of variables. Consider the illustration of factor analysis on a dataset including six variables, where the variables have been bundled or collapsed into two factors.

	Socio-economic status of an individual	Socio-economic status of a neighbourhood
Income	High	Medium
Education	High	Low

IDI's SDGs Audit Model (ISAM)

Occupation	High	Low
Value of property	Medium	High
Number of shopping malls in the neighbourhood	Low	High
Number of violent crimes in the neighbourhood	Low	High

In the case above, the set of variables can be broken down into two factors. The social-economic status of an individual comprises income, education and occupation variables. The socio-economic status of a neighbourhood includes variables providing a measure of the value of property, the number of shopping malls in the neighbourhood and the number of violent crimes in the neighbourhood. An SDG auditor may need to involve a subject-matter expert to make final decisions. For example, the variable of property value, despite having a medium effect size correlation with the socio-economic status of an individual, is better suited to be bundled under socio-economic status of a neighbourhood. Illustration 40 provides an application of the principles of factor analysis to our example relating to the EIPV.

Suggested tools: Programming in R, Python (Open source), Excel

For example, during the audit of the EIPV, an SDG auditor may obtain the following bundling of factors of variables as a result of factor analysis:

- a. socio-economic status of a victim;
- b. socio-economic status of a perpetrator;
- c. socio-economic status of a neighbourhood;
- or
- d. employability of a victim;
- e. cultural factor; and
- f. alienation factor.

Can I generalise my findings?

Statistical generalisation involves inferring the results from a sample and applying it to the population of interest. In the context of SDG audits, an auditor may need to generalise sample results to the population of interest. The fundamental requirement for generalisation is that the sample needs to be determined statistically. The exact variable to be generalised must be identified in advance. The sampling design employed during auditing must keep this requirement in mind. The auditor may find two kinds of estimation useful, that is, estimation of the:

- percentage of errors or deviations or non-compliance; and
- average or total of a variable (especially indicator data).

The methods employed to determine adequate sample sizes for the above two kinds of estimators are different. The methods employed for the determination of sample size and estimation take into account the extent of variability/heterogeneity in the population.

Suggested tools: Programming in R, Python (Open source), Excel, IDEA (for sampling)

How do I project trends?

Trend projection is a classical method of forecasting which deals with the movement of variables over time. A simple forecasting technique is trend extrapolation, which involves using a historical trend and extending it into the future. An SDG auditor may need to develop an understanding of this technique in order to conclude on the likelihood of achieving the target by the specified deadline. This method requires time series data, that is, a series of data points indexed in time order. Trend analysis may also be used to measure progress in the achievement of the nationally agreed target. The trend projection can be based on quantitative methods using past data, or on qualitative methods, using surveys and polls, among others. In the case of quantitative projections, the statistical technique of regression may be utilised. An SDG auditor may use simple, linear regression or complex, multivariate, non-linear regression techniques to project trends.

Suggested tools: Programming in R, Python (Open source), Excel

How do I carry out data unification?

Data unification is a process of combining multiple, diverse data sets and preparing them for analysis by matching, removing duplicates and cleaning the data. Hence, understanding the datasets and their linkages becomes a critical factor to perform data unification. While data integration refers to combining multiple databases in a data warehouse, data unification can be done with simpler data structures such as datasets. This concept becomes relevant in the context of vertical and horizontal coherence, as follows:

- Unification for **vertical coherence** involves unifying data across various levels of government, such as local, provincial and national governments.
- Unification for **horizontal coherence** involves unifying data across sectors, agencies and time.

A unification process involves ingesting, cleaning, transforming, combining, consolidating (eliminating duplicates and redundant information), classifying (if necessary) and exporting data.

Suggested tools: Excel, Access, IDEA, Postgresql

How can I use GIS data?

Geo-referenced data refers to location-specific information and is of particular importance to an SDG auditor. GIS data includes spatial information such as a geo-point, geo-trace or geo-shapes. Geo-point refers to one specific location and geo-trace refers to a path between two geo-points. When geo-trace is a closed loop, it results in geo-shapes. GIS data can be used for risk assessment, sampling, understanding dependencies during the preparation of audit conclusions and reporting.

For example, during the audit of the EIPV, an SDG auditor may use GIS data for the following:

- a. analysis of incidence data over administrative boundaries or geographical region to identify patterns;
- b. analysing accessibility to services such as police stations, hospital services in case of incident to evaluate protection using geo-traces; and
- c. reporting of progress of achievement of target over administrative boundaries for benchmarking.

Suggested tools: Mapping services such as Google Earth, Open streetmap and visualisation tools such as Power BI, Tableau, R, Python.

Are there other institutional considerations?

Apart from dealing with data, techniques and tools, a SAI may consider introducing the required policy framework and building capacities to conduct data analytics.

What policies should I have in place?

The following are the areas to be addressed in terms of the policy framework of a SAI. Though these are general areas of concern, they gain particular significance in the context of an audit of SDGs implementation.

- **Data Anonymisation** is a type of sanitisation of data in order protect the privacy of citizens. It is achieved by either encrypting or removing any personally identified information from the datasets. This ensure that people whom the data relates to remain anonymous.
- **Data protection** safeguards data from compromise, corruption or loss. This becomes relevant when more and more data is being created and stored as part of SDG audits. It is achieved by data management (access controls and protocols) and data lifecycle management (backup and disaster recovery)
- **Data archival policies** need to be reviewed in the context of SDGs, as the audit of many targets may be a continuum. The progress made by a country in the achievement of a target may be reviewed at a specific periodicity over a stretch of time. Further, the data archival policy should also facilitate follow-up audits (either auditing recommendation or re-audits) to measure long-term impacts and impacts on course corrections.

IDI's SDGs Audit Model (ISAM)

What are the competencies needed for data analysis?

The basic skill for data analysis includes making sense of the data. An SDG audit team may find it useful to have a data analyst in the team to perform analysis and to create visualisations. SAIs may want to invest in the following competencies to have a strong foundation for data analysis.

- **Data intuition:** Ability to intuitively understand structured and unstructured data.
- **Data gathering and pre-processing:** Ability to design collection or gathering mechanisms and pre-processing such as cleaning, grouping, filtering (if and when necessary).
- **Statistical analysis and data-interpretation:** Ability to perform analysis (single or multi-variable) and interpret the results of analysis to find dependencies, relationships (cause-effect) or trends.
- **Presenting data (reporting and visualisation):** Ability to choose the right method to report information and prepare the visualizations.
- **Knowledge of tools:** Excel, IDEA

Some of the advanced competencies may include the following.

- **Data acquisition and maintenance:** Ability to collect databases, restore them locally and manage the data access.
- **Querying:** Designing database queries to extract data out of structured databases.
- **Programming:** Developing algorithms to solve data analysis requirements.
- **Knowledge of tools:** Tableau, R, Python, Postgresql

Apart from the list of basic skills, a SAI may also find it worthwhile to invest in machine learning and AI skills with a long-term strategy in mind.

Specific aspects	Advanced	Data integration (Complex data relating to Government process from various sources)	Competency for IS audit Data archival policy	GIS tools Projection Extrapolation Correlation analysis
	Basic	Data disaggregation Data availability (Indicator data)	Data privacy and protection Competency for basis analysis	Data unification Data collection
General aspects		Data integrity Data management	Professional Competencies Infrastructure Organizational setup Data access	Machine learning Data visualization Sampling Risk assessment Data analytics tools
		Data	Institutional (SAI)	Tools & Techniques