Category	Title
General guidance	Data collection
Version	Guidebook 2019

Lead authors

Justin Goodwin and Chris Dore

Contributing authors (including to earlier versions of this chapter)

Mike Woodfield

Contents

1 Overview

This guidance incorporates and extensively references guidance from the Intergovernmental Panel on Climate Change (IPCC) 2006 *Guidelines for national greenhouse gas inventories* (IPCC Guidelines) (IPCC, 2006) and 2019 refinement to the 2006 IPCC guidelines (IPCC, 2019). This chapter provides general guidance for collecting existing national/international data and new data. The material is intended both for countries establishing a data collection strategy for the first time and for countries with established data collection procedures.

Data ([1]) collection is an integral part of an inventory. Data collection procedures are necessary to ensure the best possible data is available and used, there is improvement where new data becomes available, and degradation of the quality of the inventory through discontinued data is minimised. Data collection activities include for finding and processing existing data (i.e. data that are compiled and stored for statistical uses than other the inventory), as well as generating new data by surveys or measurement campaigns. Other activities include maintaining data flows, improving estimates, generating estimates for new categories and/or replacing existing data sources when those currently used are no longer available. Formalised data collection activities should be established, adapted to countries' national circumstances, and reviewed periodically as a part of implementing good practice. In most cases generating new source data will be limited by the resources available, and prioritisation will be needed, taking account of the results of key category analysis.

The good practice principles that underpin data collection are the following:

- Use existing national published, peer-reviewed official data where possible. This data should already have undergone review and be accepted within the scientific community.
- Focus more efforts on the collection of data needed to estimate and improve the categories with the largest emissions (key categories) and on those that have the greatest potential to change or have the greatest uncertainty; collect data/information at a level of detail appropriate to the method used.
- Consider whether input data meets the TCCCA quality principles (see Chapter 6), particularly on completeness and accuracy and any associated data quality objectives of the inventory (see Chapter 6).
- Engage in long term cooperation with data suppliers (see 2.1.3) to gather and adapt regular tailored, transparent data sets including information about QA/QC and uncertainties (see Chapter 5) and encourage continuous improvement of the data where needed and in line with the inventory's data quality objectives (see Chapter 6).
- Consider agreements with data suppliers (see 2.1.3) to support reliable, consistent and sustainable information flows. This could also provide useful inventory information back to the data suppliers or other incentives that will encourage the data supplier to engage and work proactively to provide and improve data.
- Use data quality objectives (see Chapter 6) to prioritise resources used for data collection, and to plan and clearly document data collection activities focussed on continuous improvement

([1]) Data can be defined as factual information (e.g. measurements or statistics) used as a basis for reasoning, discussion or calculation. Data collection is the activity of acquiring and compiling information from different sources.

- Review data collection activities and methodological needs on a regular basis, to guide progressive, and efficient, inventory improvement;

This chapter provides general guidance for collecting existing national/international data and new data. The material is intended both for countries establishing a data collection strategy for the first time and for countries with established data collection procedures.

2 Collecting data

This section provides general guidance for collecting and adapting data for inventory use. The guidance is applicable to emission factors, activity and uncertainty data collection activities. Throughout the data collection activities, the inventory compiler should maintain quality assurance/quality control (QA/QC) records about the data collected according to the guidance provided in Chapter 6, on inventory management, improvement and quality assurance/quality control. It is also good practice to find out and be aware of potential future data collection needs (e.g. the need for more detailed or rigorous data collection for important source categories).

Section 3 of Chapter 6, Inventory management, improvement and quality assurance/quality control, provides guidance on managing the inventory process and a stepwise approach to data collection. This process should include a strategy for assessing the data requirements of the inventory, collecting the data needed, and preparing it for use in the inventory and should be driven by the data quality objectives and scope of the inventory (see Sections 2 and 5 of Chapter 6).

Generation of new data sets

In cases where representative emission factors, activity data or other estimation parameters do not exist, or cannot be derived from existing sources, it may be necessary to generate new data. Generation of new data may entail emission measurement programmes e.g. for stack or diffuse emissions, fuel or feedstock sampling, agricultural activities, waste management activities as well as new censuses or surveys for activity data. Generation of new data is best undertaken by those with appropriate expertise (e.g. measurements by competent expert organisations using appropriately calibrated equipment). For surveys and censuses, national statistical authorities usually have the required mandates and tools. The process of generating new data is often resource intensive and are therefore prioritised for sources in key categories when there are no other practical options for improving existing data. To optimise resource use, the required data could be acquired from an extension of existing programmes rather than the initiation of totally new ones. Information on existing guidelines for activities that are defined in detail by other official bodies, such as statistical offices and measurement standards committees is referenced in these sections below.

Models

Although models are frequently used to assess complex systems and can be used to generate data, models are a means of data transformation and do not remove the need to collect data to drive them. When applying a new model, inventory compilers need to consider whether collection of new data is needed, or whether existing data can be used. Guidance on using models is provided in Chapter 6.

2.1 Sources of data

In the work of inventory compilation there is need to select those methods, emission factors and activity data that are the most representative for the national emission sources. Various alternative sources for data are presented below. It is good practice to carefully check the suitability of the methods and data for the specific purposes. For instance, when using emission factors from international databases it is important to compare if the techniques and raw materials are consistent with the processes used in a national context. Some methods may also be on a more general level and not appropriate for key category estimates.

Activity Data

Activity data refers to statistical information of numbers, volumes, quantities etc. of a certain activity, such as volume of fuel use, production volume, animal numbers, driven annual kilometres by a certain vehicle category etc.

There will be a wide range of potential data sources used for inventory compilation. Generally, the most appropriate activity data will be available from national data collection agencies, government ministries, research institutions and agencies that already collect data from different economic sectors. **Error! Reference source not found.** lists a variety of potential national and international data sources that are likely to provide a starting point for data collection.

Table 2-1 Potential sources of activity data

Literature Type	Where to find it and comments
National Statistics and other Agencies (NSAs)	This includes statistics on energy use and production, material production and consumption. Geological Survey reports, and technical reports, guidelines, standards, sectoral surveys issued by industrial trade associations. Often official statistics (because they have a more elaborate review process) take a long time to become available, but preliminary data may be available at an earlier stage. When data does not readily exist, national inventory agencies can request national statistical offices to start data collection for inventory purposes. There is typically a national environmental agency or ministry with the responsibility for regulating emissions from large sources and holding permits and records of emissions. This entity is typically able to provide plant specific fuel consumption, production information and e.g. waste management data, technical information (boiler, production and abatement techniques and their installation dates)
Sub-national or regional statistical agencies	Sub-national or regional statistical agencies (e.g. provincial ministries or municipal governments that may have a mandate to collect relevant data).
National trade associations	This includes statistics on energy use and production, material production and consumption from trade associations. Many trade associations collect data for use by members which can be used in national emissions inventories.

Literature Type	Where to find it and comments
International organisations statistics	e.g. United Nations, Eurostat or the International Energy Agency, Organisation for Economic Co-operation and Development (OECD) and the International Monetary Fund (IMF), which maintains international activity as well as economic data;

Emission factors and methodologies

Regulatory authorities and official statistical bodies have the knowledge and expertise to ensure that representative emissions sampling and measurement are carried out according to agreed standards. As measurements are typically an expensive option for data collection, it is recommended that readily available documented information is considered first.

National studies and other documentation of representative national measurements for domestic activities may be available. It is good practice for countries to use their own, peer-reviewed, published literature because this should provide the most accurate representation of their country's practices and activities. If there are no country-specific peer-reviewed studies available, then the inventory compiler is recommended to use other information sources that are closely representative of the national circumstances from the literature, models or international bodies. Table 2-2 lists a variety of potential data sources.

Table 2-2 Potential sources of emission factors and measurement data

Literature type	Source and comments	
EMEP/EEA guidebook	EEA (European Environment Agency website)	Useful defaults or for crosschecking. May not be representative of processes in your country or appropriate for Key Category estimates.
International Emission Factor Databases: USEPA	USEPA website: www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emissions-factors)	Useful defaults or for crosschecking. May not be representative of processes in your country or appropriate for Key Category estimates.
Country-specific data from international or national peer-reviewed journals	National reference libraries, environmental press, environmental news journals	Reliable if representative. Can take time to be published.
National testing facilities (e.g. road traffic testing facilities)	National laboratories	Reliable. Need to make sure the factors are representative and that standard methods are used.
Emission regulating authority records	Emission data reported by the plants from the competent supervising authorities of industrial activities and energy production and waste management.	Annually reported and plant-specific. For use in inventories, the data quality needs to be checked

Literature type	Source and comments	
Pollution release and transfer registries (PRTRs)		Annually reported and specific to industrial sites. For use in inventories, the data quality needs to be checked. Not all data may be as thoroughly checked by the authorities as those data for which emission limit values (ELVs) and other requirements are set in environmental permits.
Industrial trade associations	Industrial associations may acquire data from their member organizations for the purpose of national inventories. The associations also have specific trade association publications, libraries, and websites.	QA/QC is needed to check for bias in data and to ensure data reliability. It is good practice to ask for documentation of how the data was derived.
Other specific studies, censuses, surveys, measurement and monitoring data	Universities, research institutes, inventory agencies	Need to check the representativeness of the methods
OECD WG PRTRs Resource Centre	OECD Resource Centre *http://www.prtr-rc.fi/*, compilation of national and international methodologies and emission factors	Default methods or for crosschecking. May not be representative of processes in your country or appropriate for Key Category estimates
Emission factors from other countries	Inventory reports from other parties, web search, national libraries e.g. www.apef-library.fi	Default methods or for crosschecking. May not be representative of processes in your country or appropriate for Key Category estimates
International Fertilizer Industry Association IFIA; Food and Agriculture Organisation of the United Nations	(https://www.ifastat.org/); (http://www.fao.org/faostat/en/).	Data can be used in national inventories especially for developing countries.
Trade data	https://comtrade.un.org/	Repository of official international trade statistics and relevant analytical tables.

Facility Level Data

Industrial activities are a major contributor to air pollution. Detailed industrial facility data, increasingly collected for tracking environmental and climate performance of large facilities, have the potential to be used and to dramatically improve the accuracy of national inventories. Facility specific data may be used in many possible ways, including for emission estimates, for activity data, and for emission factors to better reflect country-specific industrial activities. Facility data may also be used to evaluate, compare or support national inventory methods. Emission data reported to pollutant release and transfer registers (PRTR) may be a useful source of information (N.B. emission data reported to the European Commission under a number of legal instruments, including the European Pollutant Release and Transfer Register and the Industrial Emissions Directive (IED), are publicly available). National inventory compilers should not assume that facility level data is by default an improvement on estimates based on national statistical activity data, due to possible reporting biases, especially if default factors are supplied which do not take into account facility-specific operation and processes. More information on the use of facility information in inventories is presented in Section 2.4 which references Volume 1 Section 2.3 of Chapter 2 on data collection in the 2019 Refinement to the IPCC Guidance (IPCC, 2019).

All datasets should be documented and checked against the data quality objectives set by the quality management system described in Chapter 6.

Other data sources

Table 2-3 Other data sources

Literature Type	Source and comments
Sectoral experts, stakeholder organisations;	National and international testing facilities (e.g. road traffic testing facilities); Industrial trade associations (technical papers such as reports, guidelines, standards, sectoral surveys or similar technical material); Universities, sector focussed institutions and agencies, e.g. industrial pollution control, transport, agriculture and energy systems. Other international experts such as emission inventory sector experts from other countries with similar national circumstances.
International and national experts	Task Force on Emissions Inventories and Projections (TFEIP), UNECE Task Force on Technological and Economical issues (TFTEI), UNFCCC roster of Experts and other networks can identify useful experts to engage with. Inventory reports from other parties.
Other reference material	Reference libraries (national libraries), scientific and technical articles in environmental books, journals and reports;

2.1.1 Screening of available data

It is best to start data collection activities with an initial screening of available data sources and engagement with experts in the field and other potential data suppliers. This will be an iterative process in which details of data that are available are built up. This screening process may be slow and require questioning until a final judgement can be made about the usefulness of a data set for the inventory. A literature review is a useful approach for gathering, prioritising and selecting from a variety of possible data sources. Literature reviews should be fully documented so that the data used for the inventory are transparent (see Chapter 6, on inventory management, improvement and QA/QC). It is also helpful to record the sources not used, providing an explanation of why, to save time in later literature review activities. The purpose for which data were originally collected may be an important indicator of reliability.

As mentioned above, in most cases it is preferable to use national data, since national data sources are typically more detailed and up to date and provide better links to the originators of the data. Most international data sets rely on nationally derived data, and in some cases data from reputable international bodies may be more accessible and more applicable to the inventory. In some cases, groups such as international trade associations or international statistical bodies will have country-specific datasets for industries or other economic sectors that are not held by national organisations. Often international data have undergone additional checking and verification and may have been adjusted with the aim of increasing consistency, although this will not necessarily lead to improved estimates if the adjusted data are recombined with national information. Countries are encouraged to develop and improve national sources of data to avoid being reliant on international data. Cross-checking national data sets with any available international data can help to assess completeness and identify possible problems with either data set.

It is recommended that inventory compilers establish proper documentation on the data and methodology used for developing of emission factors ([2]) and their uncertainties.

2.1.2 Restricted data and confidentiality

Data providers might restrict access to information because it is confidential, unpublished or not yet finalised. Typically, this is a mechanism to prevent inappropriate use of the data, unauthorised commercial exploitation or sensitivity to there being possible imperfections in the data. Sometimes, however, the organisation does not have the resources required to compile and check the data and does not want to release it in its raw form. It is advisable, where possible, to cooperate with data providers to find solutions to overcome their concerns by:

- explaining the intended use of the data;
- agreeing, in writing, to the level at which it could be made public;
- if data needs to remain confidential, making a written agreement about the arrangements;
- identifying the increased accuracy that can be gained through its use in inventories;
- offering cooperation to derive mutually acceptable data;
- giving credit/acknowledgement in the inventory to the data providers.
- aggregating smaller subcategories — it may be possible to aggregate the emission estimates into a larger category to avoid breaking the confidentiality.

The protection of confidentiality is one of the fundamental principles of a national statistical agency (NSA) ([3]) — see: http://unstats.un.org/unsd/methods/statorg/statorg.htm). NSAs are committed to safeguarding information that plainly reveals the operations, belongings, attitudes or any other characteristics of individual respondents. If respondents are not convinced that the information they provide to the NSA is absolutely confidential, the quality of the information collected may suffer. Detailed individual data must therefore be treated and aggregated to draw out the information that is important to the user, without disclosing individual data. This is more likely to be an issue for business statistics than for other data, especially where a few companies dominate the sector, and/or the country has a relatively small economy.

([2]) See for example, Ogle *et al.* (2013).

([3]) Any main national official data collection organisation is referred to here as national statistical agency.

Sometimes, depending on the size and structure of the original sample, raw data can be aggregated in a way that protects confidentiality and yet produces useful information for emission inventory purposes. If, however, there is a need to preserve confidentiality, the NSA, or the body that originally collected the data, is normally the only one that can carry out this additional treatment of the raw data.

Some countries have special arrangements to mask data (i.e. to make data anonymous with respect to companies or facilities) to allow researchers access. Inventory compilers may investigate the possibility of making such arrangements. However, as this reprocessing will be required regularly (annually if possible), a better solution would probably be for NSAs to incorporate this into their own work programmes. While this will require an initial investment in data processing, it will probably be quicker and less expensive in the long run. Once the reprocessing system is set up it can be reused every time the survey is repeated, with low marginal costs. An added advantage is that the information will then be in the public domain so that others can validate the figures reported in the inventories. Box 2.1 provides illustrative examples of the aggregation of confidential data

Box 2.1 Illustrative examples of aggregation of confidential data

Each inventory compiler will need to find categories to aggregate confidential emissions that are suited to their national circumstances.

Example 1: A country uses facility level data that are confidential to estimate emissions from large plants. These data are aggregated to a sector total so that the data remain confidential.

Example 2: A country has only one petroleum refinery and wishes to keep the refinery data confidential. It decides to report fugitive emissions under 1.B.2.a (Oil) combined with oil exploration, transport and distribution and the combustion emissions under 1.A.1.c (Manufacture of solid fuels and other energy industries) where the refinery emissions are obscured by the other emissions in this category.

Many agencies collect ancillary data during operations for other purposes, such as registration of businesses or vehicles, collection of taxes, granting of licences and allocation of grants and subsidies. Such information is usually also covered by confidentiality clauses. In general, such clauses envisage the use of the data for statistical purposes, and NSAs have the right of access to such data. Often these administrative data form the basis for sample stratification and selection, and NSAs will have experience in handling them, perhaps even developing specialist software that allows the required information to be drawn out without breaching the confidentiality rules.

For all these reasons, when existing data need to be reprocessed, it is strongly recommended to work together with NSAs or the statistical service of the relevant ministry, not only to protect confidentiality but also to make cost savings.

2.1.3 Maintaining data supply and engaging with stakeholders

Inventory compilers are encouraged to engage in cooperation with data suppliers to provide tailored data sets from their information or to help to generate new and improved data. This is likely to include some forms of modification to the existing data sets to meet the inventory requirements (e.g. where data is not collected on a calendar year basis then convert to calendar year, adjust for different classifications of sources or fill gaps in territorial coverage). It can also include adding questions to censuses and surveys or data collection associated with the implementation of related policies and measures. The inventory compiler and data supplier cooperation should be considered as a long-term relationship which ensures the flow of good quality data into the inventory estimates which benefit both the inventory and the data supplier. Engagement could include:

- one-to-one discussions to explore their data sets, gather more detail and understand uncertainties;
- scientific or statistical workshops or regular/annual informal updates on the inventory inputs, methods and outputs with groups of data suppliers;
- specific contracts or agreements for regular data supply and improvement (see below).

These activities will help to ensure that the most appropriate data are available for the inventory and that the inventory compiler properly understands the data. It will also help to establish sustainable links to data-providing organisations.

Once appropriate data sets have been identified, the inventory compiler will need to try to develop a more formal specification and data request. This formalisation enables efficient annual updating (through knowing what to ask for, from whom and when) while complying with QA/QC requirements for documentation (see Chapter 6,). A clear definition of data requirements will ensure that data are delivered as expected; the specification should include details such as:

- a definition of the data set (e.g. time series, sector and sub-sector details, national coverage, requirements for uncertainty data, emission factors and/or activity data units);
- a definition of the format of the data set (e.g. file type such as CSV, spreadsheet, XML, etc.) and structure (e.g. what different tables are needed and their structure, such as the columns);
- a description of any assumptions made regarding national coverage, the sectors included, the representative year, the technology/management level and emission factors or uncertainty parameters;
- identification of routines and timescales for data collection activities (e.g. how often the data set is updated and what elements are updated);
- a reference to documentation and QA/QC procedures and uncertainties;
- a contact name and organisation;
- a date of availability.

It can be useful to seek commitment to these specifications from the data provider organisation or to include them in a data supply agreement. Maintaining and updating these specifications on an annual basis, in case data requirements change, can also help to document the data sources and provide up-to-date material for annual routine data collection activities. It is not unusual for the delivery of data sets to be delayed, so incorporating notes and early warning routines to detect and manage delays can be useful.

Where appropriate, it may be useful to explore existing or new legal arrangements as means of guaranteeing the delivery of data to the emissions inventory. More detail is provided in Chapter 6.

2.2 Deriving emission factors from measured emissions

This section provides generic advice for the derivation or review of emission factors and related parameters. If emission factors or other parameters are developed based on site- or source-specific measurements, then the inventory compiler should check if the measurement programme included appropriate steps and quality checks as explained below. Where direct measurement data from individual plants are used, the inventory compiler needs to check information on maintenance and calibration of equipment and other QA/QC practices related to the measurements as explained in detail below.

Emissions measurement data can be used:

- to develop an annual estimate of emissions for the specific process, when results from continuous emission measurements (CEMs) are used;
- to develop a representative emission factor for a specific boiler/process. Whether measurements are continuous, periodical or campaign based, results can be used if correlated back to a process parameter or an input variable; and
- to develop an emission factor for all similar facilities in the country, i.e. a national emission factor can be developed, if the facilities can be considered similar according their fuel/raw material consumption and for technical and performance features.

Inventory compilers commonly rely on literature to provide emission factors or other estimation parameters. Table 2-1 lists a variety of potential sources of emission factor data. For key categories it is good practice for countries to use peer-reviewed, published literature relevant to their national circumstances because this should provide the most accurate representation of a country's practices and activities. If there are no country-specific, peer-reviewed studies available, then emission factors contained in international emission factor databases, or other literature values can be used. These factors should reflect national circumstances as far as possible. A literature review is a useful approach for gathering and selecting from among a variety of possible data sources. Literature reviews can be time-consuming because many lead to old data and, in addition, the use of conversion units may generate artificial differences. Journal papers can sometimes be accessible through the web without a subscription and libraries may facilitate search and access. Specialised literature sources relevant to emission factors are:

- national and international testing facilities (e.g. road traffic testing facilities);
- industrial trade associations (technical papers such as reports, guidelines, standards, sectoral surveys or similar technical material);
- national authorities with responsibility for regulating emissions from industrial processes.

Where suitable existing factors and parameters cannot be found, new measurements can be used to develop sector or category specific data for more accurate estimates or to cross-check and review existing data and methods. This can include to determine or revise emission factors, destruction/abatement efficiency factors and activity rates. Measurements can also be used to quantify emissions directly or to calibrate and verify models that are used to generate data.

Reliable and comparable results can be achieved using a well-designed measurement programme with defined objectives, suitable methods, clear instructions to the measurement personnel, defined data processing and reporting procedures, and adequate documentation. Table 2-4 sets out the elements of such an approach.

Table 2-4 Generic elements of a measurement programme

Measurement objective	A clear statement of the parameter(s) to be determined; e.g. sulphur oxides (SO_x), mercury, total particulate matter (PM), PM_{10} fraction, etc.
Methodology protocol	A description of the measurement methodology to be used. This should include: • the components to be measured and any associated reference conditions (e.g. NO_x, expressed as mg/m^3, assuming standard conditions for temperature and pressure for dry gas (volume at 273.15 K, 101.3 kPa) • methods to ensure that representative samples are taken that reflect the nature of the source category and the measurement objective (e.g. samples should be collected only when the emitting source is operating normally) • For batch processes, the suitable sampling periods should be established on a case-by-case basis; • Methods should also be established to quantify emissions during start up and shut down periods as well as exceptional situations, as these may represent a major part of emissions during a year (e.g. due to process failures, malfunctioning of abatement techniques etc.) • the identification of any standard techniques to be used (ISO, EN or other standards measurement method); • the analytical equipment needed and its operational requirements; • any accuracy, precision or uncertainty requirements (e.g. enough readings should be taken to enable the estimation of the 90 % confidence level) • data capture requirements to be met (e.g. continuous sampling covering greater than 80 % of the period tested) • QA/QC regimes to be followed (e.g. measurements should be taken according to ISO/IEC 17025:2017 'General requirements for the competence of testing and calibration laboratories')
Measurement plan with clear instructions to the measurement personnel	A measurement plan specifies, for those carrying out the measurements, the: • number and location of sampling points for each parameter to be measured and how these are to be selected (e.g. if the emission source has several discharge points) • number of individual measurements to be made for each sampling point and set of conditions (e.g. a minimum of three

	readings are required to measure heavy metal emissions at each discharge point) • measurement dates and periods of the measurement campaign; • reporting arrangements (including any special features that the report should contain) • additional source or process-related information to be collected to enable data processing or interpretation of the results (e.g.: copies of plant operational logbooks, etc.) • conditions (or range of conditions) of the source (or for industrial plant the capacity, load, fuel or feedstock) to be met during the measurements • personnel responsible for the measurements, who else is involved and the resources to be used
Data processing and reporting procedures, and documentation	• Specify data processing requirements (e.g. requirements to convert to specific reference conditions of oxygen) • Specify the reporting procedures that will form an account of the measurements, the description of the measurement objectives, and the measurement plan • Specify the documentation requirements to enable the results to be traced back through the calculations to the collected basic data and process operating conditions
Notes: 1. Further guidance is given in ISO 11771:2010 Air quality — stationary source emissions — determination of time averaged mass emissions and emissions factors — general approach. 2. Guidance on using measurement to generate data for the European Pollution Release and Transfer Register is given in the *Guidance document for the implementation of the European PRTR* (European Commission, 2006).	

When considering using measurement data it is good practice to check whether it is representative of the emission source, e.g. for a specific boiler only, or also representative for all similar boilers compared on the basis of technical details of the boilers and the fuel characteristics. Check should also include whether a suitable measurement method has been used. The best measurement methods are those that have been developed by official standards organisations and have been field tested to determine their operational characteristics ([4]). Using standardised measurement methods improves the consistency of measured data and provides the user with additional information about the method, such as statistical uncertainty levels, lower detection limits, sensitivity and upper limits of measurement. The International Organization for Standardization (ISO) standards, European Standards (EN) or suitable validated national standards (United States Environmental Protection Agency (EPA), Verein Deutscher Ingenieure (VDI), etc.) may meet these criteria ([5]). It is good practice

([4]) e.g. repeatability, reproducibility, detection limit, tolerance to interference.

([5]) The *Guidance document for the implementation of the European PRTR* (European Commission, 2006) provides a useful listing of suitable international standards.

for the inventory compiler to document any measurement or quality management standards that have been used and to bear in mind the data requirements of the uncertainty analysis in Chapter 5, on uncertainties).

Emissions can be determined from data available from direct continuous or periodic emission measurements. Continuously measured activity data (e.g. fuel consumption) can also be used to determine emission data accurately when it is combined with regularly updated and measured emission factors.

By determining emission levels from variable combustion, process and operating conditions, direct measurements provide the most accurate emission data.

When reviewing energy or industrial plant data, it is important to ensure that the measurements are representative of the specific activity and do not include or exclude extraneous components. For example, stack measurements may exclude losses to the atmosphere through evaporation or poorly burned fuel (that is emitted as VOCs). These should be included in the reported emissions totals.

During start-up and shut-down periods, air pollutant emission levels also may differ from those during normal operation time. Especially in exceptional situations such as malfunctioning of abatement technique or during process disturbances, emissions may represent the majority of annual emissions. It is important to establish methods to capture these emissions and keep record of such periods during the reporting year.

In implementing the elements of a measurement programme it is good for emissions inventory compilers to be aware of best good practice which include to:

- distinguish between, and record the different components of a mixed fuel/raw materials feed, e.g. coal and wood in a mixed fuel boiler;
- specify how the chemical composition of fuels and raw materials should be determined from the analyses of samples taken from delivery trucks/tankers, pipelines, or stockpiles;
- ensure representative sampling e.g. of exhaust gases;
- use instruments with known performance characteristics or perform relative accuracy audits against established standard reference methods.

Conversion to standard conditions

Most gas analysers determine the volume concentration of gaseous components (volume/volume) and so, unless conditions can be shown to be stable, it will be necessary to measure the exhaust gas flow rate, pressure, temperature and water vapour content, so that the emission data can be converted to reference conditions for temperature and pressure (e.g. 273 K and 101.3 kPa, dry) or quoted on a mass emission basis. Other measurements are usually needed to calculate process-specific conversion and oxidation efficiency factors and, if the fuel or raw materials used are not dry, a moisture analysis will be required. Related measurements should be made simultaneously, or in such a way that ensures the correct functional relationship between the variables being sampled, otherwise integrated flows or emissions derived from the measurements are likely to be incorrect.

The process of converting measurements to standard conditions is often outside the knowledge of emission inventory compilers, and it is therefore sensible to draw on people with expertise in measurements.

Calibration and Maintenance of Equipment

It is good practice to use scales and flow meters that are of a known quality, calibrated, maintained and regularly inspected, when using measurements to calculate activity rates, e.g. from measured fuel or raw material feed rates (or sometimes from production data). Measurement equipment can be of variable quality, and it is important that there are regular maintenance and calibration procedures in place and that these are subject to regular QA/QC review. When recording is carried out on a continuous basis it is good practice to monitor and record any time when meters are not working and the data capture rate is reduced. The advice on gap filling (in Section 2.5, Adapting data for inventory use) can, however, enable imperfect data sets to be repaired sufficiently for some purposes, such as the generation of emission factors.

Documentation of the individual measurements needs to include the moisture content if the fuel/raw materials are not dry and the concentrations of possible contaminants that could adversely affect the measurement process.

Quality management is an important factor to take into account. EN:ISO/IEC 17025:2005 'General requirements for the competence of testing and calibration laboratories' describes a useful QA/QC regime for testing and measurement. It encourages the use of standard methods, by qualified personnel, using suitability tested equipment and traceable calibration artefacts. Taking and storing samples, any subsequent analysis and the reporting of results should also be covered by a quality management system of the plant and included in the documentation accompanying the measurements. DD CEN/TS 15675:2007 'Air quality. Measurement of stationary source emissions. Other standards related to plat specific measurements are EN ISO/IEC 17025:2005 to periodic measurements, and ISO 11771 'Air quality — determination of time-averaged mass emissions and emissions factors — general approach'.

Using dispersion modelling

Emissions from diffuse or fugitive sources from industrial plants, roads or area sources can be estimated by monitoring ambient air emissions and then using reverse dispersion modelling methods to estimate the source strength. For example, EN15445:2007 'Fugitive and diffuse emissions of common concern to industry sectors — fugitive dust emission rate estimates by Reverse Dispersion Modelling', describes such an approach and the sampling experimental set-up. The standard provides guidance on taking into account field data such as the number, height and width of diffuse dust sources, sampling distances and meteorological information.

2.3 Activity data

This section provides general advice for the production or review of activity data. This focuses on the use of surveys and censuses that provide the statistical information used for inventories.

2.3.1 Surveys and census information

Survey and census information (see note below) provide statistics (e.g. the best agricultural, production and energy consumption) that can be used for the emission inventory. These data can also provide some insights into practices and technologies (e.g. farming practices, boiler types) that may be needed for applying appropriate and more accurate emission factors. Generally, these data are compiled by NSAs or relevant ministries and research institutes for national policy purposes or to comply with international demand for data, or other activities that are outside the direct control of the inventory compiler. This information is usually available for national inventories. In addition the needs of the inventory can sometimes trigger or influence surveys or censuses.

NOTE:

The difference between census and survey data

Survey data are derived from sampling and therefore do not collect primary data for the whole population. Surveys are sent to a representative sample of the population (in the context of the survey's purpose), so that the results can be expanded to provide an estimate of the full population. A survey could, for example, assess the number of animals in a country or region by surveying a discrete selection of farms and groups of farms in a country or region. Using more general surrogate data and assumptions would then allow the national or regional total to be derived. Both the representativeness of the sample and the methods used to gross up the data need careful planning and review.

Census data are based on a complete count of the whole population, i.e. an actual count of all the animals in a region or country. A census is usually limited in detail and diversity to only the most important national statistics, such as human and livestock population. Planning and undertaking a census is expensive and time consuming and these resource implications can be significant limiting factor in using them for specific national inventory applications. Often, census data are used as a reliable surrogate for extrapolating survey data to national statistics.

Using existing census and survey data: In some countries the NSA is a single agency responsible for all national statistics, while in other countries the task is split among multiple agencies, each of which collect official statistics related to their field. For instance a country's agriculture ministry may be responsible for carrying out agricultural surveys and censuses. This has the advantage that the ministry is likely to have the specialist knowledge required to adequately define the data to be collected and to have at its disposal the administrative information to help stratify and select the sample to be surveyed, for example a register of businesses working in the area covered by the remit of the ministry. In these cases ministries may have their own statistics departments (or will work closely with an NSA) to provide statistical expertise and knowledge which is essential for avoiding many of the common pitfalls in data collection.

Where available these data sets can be used either directly (if they represent the geographical and sectoral coverage required) or as part of a hybrid data set in combination with other information necessary to derive the detail and geographical coverage required.

Developing new surveys: Developing new surveys, especially surveys of consumers or households, is relatively expensive because sample size and proper conduct of the fieldwork, data processing, analysis and reporting are all resources demanding. Considerable effort would be required to check the reliability and consistency of data, even when response rates appear to be otherwise satisfactory. Unless they can be consistently repeated, surveys are only able to give measurements relating to one point in time. Bearing this in mind, and also taking into account the length of time such surveys take to design, execute and analyse, attention should first be paid to the possibility of obtaining regular and consistent data from existing sources, such as recombining data collected for other purposes, or using administrative data. For instance, in the development of a major survey, planning typically starts about 18 months before data collection starts, with results available a year or more after the data collection period.

Where new data collection is unavoidable, the NSA and/or the relevant ministry can identify what surveys are ongoing or planned and can explore the possibility of adding new questions or modules to these surveys to fill the data gaps. One of the many advantages of working with an NSA or ministry is that it will design the method of collection and the questionnaire to take into account the needs of as many users as possible. This reduces costs as well as the burden on businesses and other respondents, making it more likely that they will complete the questionnaire. In addition, selecting the survey sample requires a reliable sampling frame, for example census data or business registers. The NSA or relevant ministry will have ready access to such sources and experience in using them. It will have teams of qualified and experienced statisticians, who are experts in sample selection, questionnaire design, data handling and verification, and the necessary software to process the data. It may also have teams of interviewers experienced in telephone or personnel surveying. All of these factors contribute to the success of any survey and, equally importantly, to keeping costs down.

General guidelines for planning surveys and census: It is good practice to plan each step with all subsequent steps in mind, from data collection, processing and analysis to dissemination of the output. For example, the questionnaire and other data collection procedures should be developed only after thinking through how the data will be processed and analysed, and the nature of the statistical information that will eventually be reported. In particular, planning needs to cover the following:

- Budget issues and costs will always be a major consideration. The total budget needs to be calculated and resources allocated to each phase of the process. Uncontrolled spending on each phase until the budget is depleted can lead to the collection of data without the necessary resources to produce and disseminate high quality output.
- Staff resources, including management of the interviewer workforce, need to be planned to ensure that people with the right skills are available at the appropriate times in all phases of the process. If interviewers are used rather than self-completion questionnaires, the interviewer workforce is likely to be the largest single cost in the data collection.
- Good project management is essential to ensure a smooth-running collection and avoid timetable issues. Adequate time needs to be allocated to each phase of the collection process. A thorough pre-testing of the questionnaire will help ensure that the data collected are reliable and valid.

The American Statistical Association's brochure on how to plan a survey is a useful source of help when setting up a new survey, and can be downloaded from www.amstat.org/sections/srms/brochures/survplan.pdf

The United Nations' guidelines for conducting household surveys in developing and transition countries provide detailed information on how to set up sample surveys based on direct questions to households, and can be found at http://unstats.un.org/unsd/HHsurveys/part1_new.htm

Moreover, many organisations contribute to statistical capacity building and will provide assistance to developing countries wishing to set up new surveys; the United Nations Environment Programme (UNEP), the United Nations Development Programme (UNDP) and the World Bank are the implementing agencies of the Global Environment Facility.

Three important steps are needed before deciding whether a survey is required and what modules it should contain:

- Review what data are likely to be available through existing data systems, including planned surveys. Remember that published statistics are based on detailed data that have been treated and aggregated to draw out the information that is important to the main user. In some cases, depending on the size and structure of the original sample, those raw data can be recombined in different ways to produce data that are appropriate for another user.

- Explore administrative sources of data. While the administrative records may not initially be easy to use for inventory purposes, once the system has been reorganised and restructured to produce the relevant data, it can become the regular source of the relevant information at little marginal cost. More and more countries are beginning to realise the cost benefits of using administrative data for statistics, and, in some cases, NSAs are obliged by law to explore the use of administrative data to provide statistics before deciding to launch an expensive new survey.
- Explore the possibility of incorporating new questions or modules into existing surveys.

If, after exploring the possibility of making use of existing data, data gaps still remain, then approach the NSA or ministry about carrying out a new survey. Provided that the financial resources are made available, the NSA or ministry will be able to provide the all-important expertise. In addition, explore whether other partners might be interested in sharing the work and resources needed for it.

Volume 1 Chapter 2 Annex 2A.2 General guidance on statistical data and surveys of the 2019 IPCC refinement (IPCC, 2019) provides useful additional guidance on statistical data and surveys, including those for energy, industry, agriculture, waste and forestry data.

2.4 Use of facility-level data in emissions inventories

As more data becomes available at the facility level, it is important to give detailed consideration to how this can be incorporated into emission inventories. Volume 1 Section 2.3 of Chapter 2 on data collection in the 2019 Refinement to the IPCC Guidance (IPCC, 2019) includes a comprehensive section on the use of facility level data in emissions inventories. This includes:

- A section on designing facility reporting programmes for inventory use. This includes the type of data that relates to facilities and the associated quality criteria and reporting elements, how reporting specifications can be designed and approaches for collaboration with national statistical institutes and agencies. Extending these activities to include air pollutants as well as greenhouse gases would bring significant advantages.
- A section on approaches for the use of reported facility data. This includes how facility data can be incorporated into emissions inventories, including considerations about completeness within a source sector.
- For the use of facility data not originally designed for inventory application, there are considerations about whether, and how, it is appropriate to use available data (e.g. for estimating emissions or for independent review), including a decision tree.

2.5 Adapting data for inventory use

Whether using existing data, making new measurements or combining the two, it is important to ensure that the level of detail and coverage of the data are matched, including sectors/process/abatement, location, land type, compound and year included. It is also important to ensure consistency in the time series and between categories and pollutant estimates. There are a number of techniques and concepts presented in Chapter 4, on time series consistency, which help to fill gaps, add detail and integrate raw data into the estimates. Common areas where data will need work to generate the detail and incorporate representative data sets are listed below.

2.5.1 Gaps in periodic data

Gaps in the time series will exist when data are available at less than annual frequency. For example, time consuming and expensive surveys relating to natural resources — such as national forest inventories — are compiled every 5th or 10th year. Time series data need to be inferred to compile a complete annual estimate for the years between surveys and for forecasts and back-casts (e.g. if estimates are needed for 1990-2004 but survey data are available only for 1995 and 2000). Chapter 4, on time series consistency, provides details on splicing and extrapolation methods to fill these gaps.

2.5.2 Multi-year averaging

Countries should report annual inventory estimates that are based on best estimates for actual emissions and removals in that year. Generally, single-year estimates provide the best approximation of real emissions/removals, and a time series of single-year estimates prepared according to good practice can be considered consistent. Countries should, where possible, avoid using multi-year averaging of data that would result in over- or underestimates of emissions over time, increased uncertainty, or reduced transparency, comparability or time series consistency of the estimates. However, in some specific cases that are described for specific sectors in the sectoral guidance chapters, multi-year averaging may be the best/only way to estimate data for a single year.

2.5.3 Non-calendar year data

It is good practice to use calendar year data whenever the data are available. If calendar year data are unavailable, then other types of annual year data (e.g. non-calendar fiscal year data, e.g. April-March) can be used, provided that they are used consistently over the time series and that the collection period for the data is documented. Similarly, different collection periods can be used for different emission and removal categories, again provided that the collection periods are used consistently over time and documented. It is good practice to use the same collection periods consistently over the time series to avoid bias in the trend. Animal population data may, for example, have been collected in the summer and so may not correspond with the annual average. The data should be corrected where possible to represent the calendar year; if uncorrected data are used, it is good practice for the inventory compiler to make consistent use of either calendar year data or fiscal year data for all years in the time series.

2.5.4 Time series revision

In order to meet deadlines, statistical organisations may use modelling and assumptions to complete the most recent year of their estimates. These estimates are then refined the following year when all the data have been processed. Data may also have been subject to further historical revision to correct errors or to update new methodologies. It is important that the inventory compiler looks for these changes in the source data time series and integrates them into the emission inventory.

2.5.5 Incorporating improved data

Although the ability of countries to collect data generally improves over time and thus countries implement higher tier methods, the data may not necessarily be suitable for earlier years. For example, when direct sampling and measurement programmes are introduced, there will be inconsistencies in the time series if the new data are not indicative of past conditions. Sometimes this can be addressed if the new data are sufficiently detailed (e.g. if emission factors for a modern abated plant can be distinguished from those of an older unabated plant) and the historical activity data can be stratified using expert judgement or surrogate data. Chapter 4, on time series consistency, provides more details on the methods of incorporating improved data consistently across a consistent time series.

2.5.6 Compensating for deteriorating data

Deterioration can occur as the result of changing priorities within governments, economic restructuring or diminishing resources. For example, some countries with economies in transition no longer collect certain data sets that were available in the base year, or, if available, these data sets may contain different definitions, classifications and levels of aggregation. Splicing techniques, as described in Chapter 4, on time series consistency, can be used to manage data sets that have deteriorated over time.

2.5.7 Incomplete coverage

When data do not fully represent the whole country, e.g. measurements for 3 of 10 plants or survey data of the agricultural activity for 80 % of the country, then the data can still be used but need to be combined with other data to calculate a national estimate. In these cases, expert judgement (see Section 2.1.3 above for details) or the combination of these data with other data sets (surrogate or exact data) can be used to calculate a national total. In some cases, survey or census data are collected in a rolling national programme that samples different provinces or sub-sectors yearly, with a repeat cycle that builds a complete data set after a period of years. It is recommended that, bearing in mind that time series consistency assumptions made in one year must also apply to the other years, data sources should be requested to compute representative yearly totals.

2.5.8 Combining data sets

Sometimes an inventory compiler will be presented with several potential data sets to use for the same estimate, e.g. a series of independent measurements for the carbon content of a fuel. If the data refer to the same quantity and were collected in a reasonably uniform manner, then combining them will increase accuracy and precision. Combination can be achieved by pooling the raw data and re-estimating the mean and the 95 % confidence limits or by combining summary statistics using the relationships set out in statistical textbooks. It is also possible to combine measurements of a single quantity made using different methods that produce results with different underlying probability distributions, but the methods for doing this are more complex, and in most cases it will probably be sufficient to use expert judgement to decide whether to average the results or to use the more reliable estimate and discard the other. When using data that are not homogeneous (e.g. because of the presence of abatement technology at some plants but not others) the inventory estimate should be stratified (subdivided) so that each stratum is homogeneous, and the national total for the source category will then be the sum of the strata. The uncertainty estimates can then be obtained using the methods set out in Chapter 5, on uncertainties, by treating each stratum in the same way as an individual source category. Identification of inhomogeneity may be helped by having specific knowledge of circumstances of individual plant or technology types, or by detailed data analysis, e.g. scatter plots of estimated emissions against activity data. Empirical data sets may contain outliers — data points that lie outside the main probability distribution and are regarded as unrepresentative. These may be identified by some rule — for example that they lie more than three standard deviations from the mean. Before taking this path the inventory compiler should consider whether the apparently anomalous data do in fact indicate some other set of circumstances (e.g. a plant in start-up conditions) that should really be represented separately in the inventory estimate.

2.5.9 Using regional inventory data

In some circumstances regional activity statistics and emission data sets are more detailed, up to date, accurate and/or complete than national data sets. In these cases a regionally compiled and then aggregated inventory can result in a better quality inventory for a country than one compiled using averaged national statistics and data sets. In such cases, and in order to fulfil the requirements of good practice, inventories can be compiled entirely or in part on a regional basis provided that:

- Each regional component is compiled in a way that is consistent with good practice QA/QC, choice of tiers, time series consistency and completeness.
- The approach used to aggregate the regional inventories and fill any gaps at a national level is transparent and in line with the good practice methods provided in the Guidelines.
- The final country inventory complies with the good practice quality requirements of completeness, consistency, comparability, timeliness, accuracy and transparency. In particular the sector estimates calculated at different regions, and then aggregated in the final inventory, should be self-consistent. There should be no emissions omitted or double counted in the aggregated inventory, and the different parts of the inventory should use assumptions and data consistently as far as practical and appropriate.

2.6 Using expert judgement

Expert judgement on which methodological choice and input data to use is ultimately the basis of all inventory development, and sector specialists can be of particular use to fill gaps in the available data, add to data detail, select data from a range of possible values or make a judgement about uncertainty ranges, as described in 0. Experts with suitable backgrounds can, for example, be found in government, industrial trade associations, technical institutes, industry and universities.

The goal of expert judgement may be to find the proper choice of methodology and parameter values from ranges provided, the appropriate mix of technology in use, the most appropriate activity data to use or the most appropriate way to apply a methodology. A degree of expert judgement is required even when applying classic statistical techniques to data sets, as one must judge whether the data represent a random sample and, if so, what methods to use to analyse the data. This requires both technical and statistical judgement. Interpretation is especially needed for data sets that are small, highly skewed or incomplete. In all cases the aim is to be as representative as possible to reduce possible bias and to increase accuracy. The formal methods for obtaining (or eliciting) data from experts are known as expert elicitation (see 0 for details).

3 References

European Commission, 2006, *Guidance document for the implementation of the European PRTR,* European Commission.

IPCC. 2006, *2006 IPCC guidelines for national greenhouse gas inventories,* Intergovernmental Panel on Climate Change National Greenhouse Gas Inventories Programme, Institute for Global Environmental Strategies, Hayama, Japan (www.ipcc-nggip.iges.or.jp/public/2006gl/index.htm).

IPCC (2019) 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (https://www.ipcc-nggip.iges.or.jp/public/2019rf/index.html), accessed 15 August 2019.

4 Point of enquiry

Enquiries concerning this chapter should be directed to the co-chairs of the Task Force on Emission Inventories and Projections. Please refer to the TFEIP website (www.tfeip-secretariat.org/) for the contact details of the current co-chairs.

Appendix A. Protocol for eliciting expert judgement

Wherever possible, expert judgement should be elicited using an appropriate protocol. An example of a well-known protocol for expert elicitation, the Stanford/Stanford Research Institute (SRI) protocol, has been adapted and is described below:

- **Motivating:** Establish a rapport with the expert and describe the context of the elicitation. Explain the elicitation method to be used and the reason it was designed that way. The elicitor should also try to explain the most commonly occurring biases to the expert and try to identify possible biases in the expert.
- **Structuring:** Clearly define the quantities for which judgements are to be sought, including, for example, the year and country; the source category; the averaging time to be used (e.g. 1 year); the focus activity data; the emission factor or, for uncertainty, the mean value of emission factors or other estimation parameters; and the structure of the inventory model. Clearly identify conditioning factors and assumptions (e.g. resulting emissions or removals should be for typical conditions averaged over a 1-year period).
- **Conditioning:** Work with the expert to identify and record all relevant data, models and theory relating to the formulation of the judgements.
- **Encoding:** Request and quantify the expert's judgement. The specific qualification will differ for different elements and will be present in the form of a probability distribution for uncertainty and an activity or emission factor estimate for activity data and emission factors. If appropriately managed, information on uncertainty (probability density function) can be gathered at the same time as gathering estimates of activity or emission factors. The section on encoding in Chapter 3, on uncertainties, describes some alternative methods to use for encoding uncertainty.
- **Verification:** Analyse the expert's response and provide the expert with feedback on what has been concluded regarding his or her judgement. Is what has been encoded really what the expert meant? Are there inconsistencies in the expert's judgement?

Possible biases in expert elicitation: elicitation protocols should be designed to overcome the biases that can be introduced by the rules of thumb (sometimes called heuristics) that experts use when formulating judgements.

The most common unconscious biases introduced by rules of thumb are:

- availability bias: this is basing judgements on outcomes that are more easily remembered;
- representativeness bias: this is basing judgements on limited data and experience without fully considering other relevant evidence;
- anchoring and adjustment bias: this is fixating on a particular value in a range and making insufficient adjustments away from it in constructing a representative estimate.

To counteract the first two potential sources of biases, elicitation protocols should include a review of relevant evidence. In order to counteract the third potential source of bias, it is important to ask the expert to make judgements regarding extreme values first, before asking for judgements regarding the best estimate or the central values for an uncertainty distribution.

There is also the possibility of more conscious biases:

- Motivational bias is a desire by an expert to influence an outcome or to avoid contradicting prior positions on an issue.
- Expert bias arises from an unqualified expert's desire to appear as a true expert in the field. This would typically lead to overconfident estimates of uncertainty.
- Managerial bias is a situation in which an expert makes judgements that achieve organisational goals, rather than judgements that reflect the actual state of knowledge regarding an inventory input.
- Selection bias occurs when the inventory compiler selects the expert who tells it what it wants to hear.

The best way to avoid these biases is to be careful in the selection of experts. Expert judgements can be elicited from individuals or groups. Groups can be useful for sharing knowledge and hence could be part of the motivation, structuring and conditioning steps of the elicitation. However, group dynamics occasionally introduce other biases. Thus, it is usually preferable to elicit judgement on an individual basis. When eliciting judgements independently for a given quantity from two or more experts, it is possible that different views on distributions (or ranges) will be obtained. In some cases, the differences may not lead to a significant difference in the overall estimate for the inventory, such as when the inventory is not sensitive to the particular quantity. Thus, in these cases, disagreements among experts do not matter significantly to the overall assessment. However, when judgements differ, and when the quantity for the estimates made is important to the overall inventory, there are two main approaches that can be used. One is to estimate resulting emissions or removals and perform the uncertainty analysis separately for each set of judgements and compare the results. The other is to ask the experts to weight the judgements and combine them into one assessment. The former approach is preferred where possible, but the latter is acceptable, provided that the judgements are weighted and not averaged. The difference is that weighting enables sampling from each of the expert's estimations, whereas averaging can produce intermediate values that are not supported by any of the experts' judgements. A similar situation can occur when comparing predictions with alternative models, as described in Section 2.3.2, on combining data sets (numerically), and the distinction between weighting and averaging is explained there. Although the development of weighting schemes can be complex, it is reasonable to start by assuming equal weights for each expert and to refine the development of weights only as needed or as appropriate for a given situation.

Expert judgement documentation: The subjective nature of expert judgement increases the need for quality assurance and quality control procedures to improve comparability of emission and uncertainty estimates between countries. It is recommended that expert judgements are documented as part of the national archiving process and that inventory compilers are encouraged to review expert judgements, particularly for key categories. Table A.1 below shows an example of the document elements necessary to provide transparent expert judgement (column 1) and an example of the data to record for each element (column 2).

Such documentation will save the compiler a considerable amount of time in reporting and documenting the inventory through the enhanced transparency of the inventory. More general text on documentation, checking and review of methods is included in Chapter 6, on inventory management, improvement and QA/QC. These principles should also be applied to the use of expert judgement in inventory compilation and uncertainty assessment.

Table A.1 Example of documentation of expert judgement

Documentation element	**Documentation example**
Reference number for judgement	EJIPPU2017-001
Date	14 January 2017
Name of expert(s) involved	Dr Anne N. Other
Expert's background (references, roles, etc.)	Nitric acid process emissions and abatement industrial expert
The quantity being judged	National emission factor for emissions of NO_x from nitric acid plant
The logical basis for judgement, including any data taken into consideration. This should include the rationale for the high end, low end and central tendency of any uncertainty distribution	An absence of measurement data for 9 out of the 10 nitric acid plants. The single plant estimate has been recommended as the basis for a national factor to be applied to national nitric acid production
The result, e.g. activity value, emission factor or, for uncertainty, the probability distribution; or the range and most likely value, and the probability distribution subsequently inferred	8 750 g NOx/tonne nitric acid produced for 1990-2019
Identification of any external reviewers	Nitric Acid Trade Association
Results of any external review	See document e:/2003/ExpertJudgement/ EJIPPU2015-001.doc
Approval by inventory compiler, specifying date and person	25 January 2017, Dr S. One.

Appendix B. General guidance on performing surveys

Survey data are often compiled using financial/fiscal incentives for reporting. This may introduce possible biases if the incentives favour a certain bias in reporting, e.g. taxation may encourage under-reporting, while incentives may encourage over-reporting. In addition, differential taxation of different categories using the same fuels may skew reporting, e.g. over-reporting of fuel used in low-taxation categories and under-reporting of fuels used in high-taxation categories.

Energy surveys

Energy statistics are a fundamental component of emissions inventories, and there is great potential for double counting. The best way to avoid double counting is to compile energy balances according to the basic principles, concepts and methods developed at an international level. The United Nations (UN) publication, *Energy statistics*: *a manual for developing countries* (UN, 1991), serves as a guide for developing countries for the comprehensive, reliable and regular collection of energy statistics. Various sources of inconsistencies, such as sources of data, concepts and definitions, and time spans/coverage, are discussed in detail for all types of energy commodities, and recommendations are provided to minimise or eliminate them. English and French versions can be downloaded from: http://unstats.un.org/unsd/publication/SeriesF/SeriesF_56E.pdf.51. This publication should be used in conjunction with two other UN publications:

- *International recommendations for energy statistics* (UN, 2016) ([6]) is an updated document on methods, while *Concepts and methods in energy statistics, with special reference to energy accounts and balances* (UN, 1982) considers the nature of energy statistics and the kinds of policy problems for which they are required, the conceptual and methodological issues to which these problems give rise, and the possible conventions that might be adopted for dealing with some of these issues. It also examines the key role played by quantitative overall energy balances, the desirable features of such balances (whether used for analysing the past or for speculating about the future), the classification problems posed by energy statistics, and the relationship between such data and other economic statistics and accounting frameworks. The document is out of print but English, French and Russian versions can be downloaded from: https://unstats.un.org/unsd/publication/SeriesF/SeriesF_29E.pdf
- *Energy statistics: definitions, units of measure and conversion factors* (UN, 1987) contains detailed information on terminologies for energy commodities, units of measurement and conversion from one unit to another. It provides internationally adopted definitions, conversion factors and descriptive tables for analysis and comparison of international energy statistics. The document is out of print, but English, French, Russian and Spanish versions can be downloaded from: https://digitallibrary.un.org/record/139939/files/ST_ESA_STAT_SER.F_44-EN.pdf

In addition, the International Energy Agency (IEA) has published its *Energy statistics manual* (IEA, 2005), providing useful background information for collecting, reporting and understanding energy statistics ([7]).

([6])https://unstats.un.org/unsd/energy/ires/IRES_Whitecover.pdf

([7])https://www.iea.org/publications/freepublications/publication/energy-statistics-manual.html

As for energy statistics, there is also an IEA database documenting the worldwide energy statistics (2017 edition) ([8]).

The UN statistics manual for developing countries in English and French can be downloaded at http://unstats.un.org/unsd/publication/SeriesF/SeriesF_56E.pdf

See also https://unstats.un.org/unsd/publications/ for other UN energy titles.

Statistics on the production of hard coal, brown coal, crude oil, natural gas and electricity are collected on a monthly basis and are available from the *Monthly bulletin of statistics online* ([9]). Aggregated data on annual production, trade and apparent consumption of primary energy products are also published in the United Nations Statistics Division (UNSD) *Statistical yearbook* ([10]). UNSD is involved with the Asia-Pacific Energy Research Centre (APEC), Eurostat, the International Energy Agency (IEA), the Latin American Energy Organization (OLADE), the Organization of Petroleum Exporting Countries (OPEC) and the International Energy Forum Secretariat (IEFS) in the Joint Organisations Data Initiative (JODI) ([11]).

Enerdata and Eurostat also provide additional data sets on energy and on other statistics.

In some cases, energy data are not available at the level of detail necessary to estimate emissions, e.g. for non-CO_2 emissions of road transport where the emissions are highly dependent on the use of catalytic converters in petrol vehicles. In these cases, additional survey or census data should be used to make estimates, e.g. vehicle sales and traffic survey data.

The International Renewable Energy Agency (IRENA) publishes statistics on renewable energy ([12]).

Usually not all the information is available from international sources, and international questionnaires should be analysed to understand the main issues to be solved to improve emission estimates. A discussion among national inventory experts, ministries of energy, and international organisations and national statistical offices should be entered into.

Additional explanations are usually necessary on the type of fuels, to understand what specific figures they refer to. If data are expressed in mass units, attention should be paid to the transformation, for example in energy units (kt to TJ). International data sets may use different calorific values compared to country-specific calorific values, and the underlying values used are not always transparent.

In addition, timeliness of the reporting of data should be considered. International organisations may use preliminary data for reporting of recent years, and the updating of historical data is often not applied to international data sets.

Data quality information or uncertainties related to the reported data are not always available. However, the inventory experts should look for the information within the international agency or use default values provided by this guidebook.

([8]) https://www.iea.org/statistics/relateddatabases/worldenergystatisticsandbalances/

([9]) https://unstats.un.org/unsd/mbs/app/DataSearchTable.aspx

([10]) https://unstats.un.org/unsd/publications/statistical-yearbook/

([11]) http://www.jodidata.org/

([12]) www.irena.org/DocumentDownloads/Publications/IRENA_RE_Capacity_Statistics_2017.pdf

The United Nations produces key publications on International Recommendations for Energy Statistics (IRES; see UN, 2016).

Industry surveys

Greenhouse gas inventories require data on the production of industrial commodities and, if possible, on the production processes. For the purpose of collecting harmonised statistics on industrial production, standardised commodity lists have been established at international level, and countries are encouraged to adopt these lists for their own purposes, as this will be most cost-efficient. These lists are updated regularly to take account of new products being developed. Industrial commodities list and guidelines for countries are available on the UNSD website, http://unstats.un.org/unsd/methods.htm, when finalised. The CPC, ISIC and HS classifications can be found at http://unstats.un.org/. Detailed chemical industry data (production per country of many products, process data) can be acquired from CEFIC; steel process information can be acquired from the World Steel Association (www.worldsteel.com).

It is more difficult to obtain information on production processes used by industry. Business registers may contain this information, but the logistics of keeping this up to date are formidable. Industry associations that bring together businesses working in a common field can often be a useful source of help. As specialists in their fields, they will have insider knowledge of the most common processes used, and they may even be willing to survey their members at regular intervals to assess the penetration of new processes. Eurostat data sets are a useful starting point for countries working on activity data from production processes.

Data on industrial production and production processes are also extremely useful in producing statistics on industrial waste (see below).

Production data used to estimate emissions related to the consumption of a product or fuel should, wherever possible, incorporate the import/export statistics for that commodity. Production statistics may, with care, be used as a surrogate for consumption when net imports or exports are thought to be significant but not quantifiable. However, as there is a possibility of incompleteness or overestimation due to under-reporting of imports and/or exports, the completeness of accounting of imports and exports should be checked with the statistical office.

Where production data are used, care should be taken to identify whether the data represent gross or net production (i.e. with or without internal recycling). For some categories these figures could differ by 5-10 %, e.g. steel, aluminium and glass. Whatever production statistics are used, appropriate emission factors should be applied, and the inventory complier should be sensitive to any tax or financial influences that might lead to over- or under-reporting of emissions.

Other international sources are the statistics on international trade by the OECD ([13]) and the UN ([14]).

In addition, the World Trade Organization provides quantitative information in relation to economic and trade policy issues. Its databases and publications provide access to data on trade flows, tariffs, non-tariff measures and trade in value added.

([13]) http://www.oecd-ilibrary.org/trade/data/international-trade-by-commodity-statistics_itcs-data-en

([14]) https://comtrade.un.org

Agricultural surveys and censuses

Since its establishment, the Food and Agriculture Organization of the United Nations (FAO) has promoted national censuses of agriculture through its World Programme for the Census of Agriculture (see www.fao.org/es/ess/census/default.asp). The programme is prepared by the Statistics Division of the FAO in collaboration with many experienced agricultural statisticians all over the world (see 'Programme for the World Census of Agriculture 2020': http://www.fao.org/world-census-agriculture/wcarounds/wca2020/en/).

The programme is complemented by practical information on the steps involved in actually conducting an agricultural census (see 'Conducting agricultural censuses and surveys' (FAO, 1995) www.fao.org/es/ess/census/agcensus.asp).

Other guidance from FAO on conducting agricultural surveys includes the following:

- *Sampling methods for agricultural surveys*, FAO Statistical Development Series (SDS) No 3 (FAO, 1989) presents the foundations of probability sampling theory and the basic concepts involved. It concentrates on sample design, which covers only part of the overall design of agricultural sample surveys. The different sampling methods are discussed, including simple random sampling, stratification, systematic sampling, probability proportional to size sampling, cluster sampling, multi-stage sampling, multi-phase sampling and area sampling. Also discussed are sample design issues, such as sample allocation to strata and to different stages of sampling; weighting and sample estimation methods, such as unbiased and ratio estimates; and methods of estimating sampling errors, including replicated methods. Some practical problems involved in designing and conducting sample surveys are also discussed, including frame problems and evaluation of sampling and non-sampling errors.
- *Collecting data on livestock*, FAO SDS No 4 (FAO, 1992) presents a general framework for livestock statistics within the context of a national agricultural statistics system. Different data collection methods are discussed, with particular reference to the problems of nomadic livestock. Guidelines for undertaking livestock censuses are also provided. Concepts and definitions for the collection of data on livestock products (meat, milk, eggs, wool and skins) are presented, along with a discussion of statistics on costs of production and feed/fodder.
- *Multiple frame agricultural surveys: volumes 1 and 2*, FAO SDS Nos 7 and 10 (FAO, 1996, 1998) is a guide to sampling survey methods. National current agricultural survey programmes, established to obtain reliable and timely baseline data on the agricultural sector, are based on one of three sampling survey methods: list sample designs (commonly farm sampling designs); area sample designs; and multiple frame designs. Multiple frame designs are those that combine an area sample with complementary list (farm) samples. Multiple frame sampling methods should constitute the statistical foundation of the national agricultural survey programmes in a larger range of countries because of their advantages over the traditional farm sampling methods.

Volume 1 is a comprehensive introduction to establishing and conducting area and multiple frame probability sample survey programmes, with a focus on methods and practices applicable in developing countries. It provides a general classification of alternative agricultural survey designs along with an indication of their advantages and limitations. It examines several aspects that have to be considered to establish and conduct a periodic agricultural survey programme based on multiple frame sampling methods, i.e. the probability selection and estimation methods, the survey organisation, the equipment and materials needed, data collection, summarisation and processing.

The book includes a detailed description of a category of multiple frame survey designs considered especially useful for developing countries.

Volume 2 presents the area and multiple frame survey methods for agricultural survey programmes that are currently used in a wide range of countries. It provides actual examples of the application of the survey methods presented in the first volume.

Forest surveys

FAO is also the lead organisation collecting data on forestry. The Forestry Department of FAO is undertaking important activities that support national forest monitoring. Information on these activities — including on the sampling design, size, plot configuration and variables to collect on the ground and remotely — can be found at: http://www.fao.org/redd/areas-of-work/national-forest-monitoring-system/en/

FAO has also produced the following relevant publications:

- *Knowledge reference for forest resource assessments* ([15]) and Voluntary guidelines on national forest monitoring ([16]);
- *Manual of forest inventory*, FAO Forestry Paper 27 (FAO, 1981) ([17]);
- *Forest volume estimation and yield prediction*, FAO Forestry Papers 22/1 and 22/2 (FAO, 1980a, b) ([18]);
- Methods and guidance documentation of the Global Forest Observation Initiative (which contains methodological advice for countries wishing to make use of remotely sensed and ground-based data for forest monitoring and reporting) ([19]).

Furthermore, FAO has developed a set of tools that support data collection for forestry as follows:

- 'Collect', a survey design and data management tool (http://www.openforis.org/tools/collect.html);
- 'Collect Mobile', a tool for data collection and validation in the field (http://www.openforis.org/tools/collect-mobile.html);
- 'Calc', a tool for data analysis and calculating results (http://www.openforis.org/tools/calc.html);
- 'Collect Earth', a tool for augmented visual interpretation for land monitoring (http://www.openforis.org/tools/collect-earth.html).

Waste surveys

In general, industries will have a good idea of the volume and composition of waste that they produce each year, as they often have to pay to have it removed and appropriately treated. Therefore, surveys of industry should result in reliable data on waste generated and its composition.

([15]) http://www.fao.org/3/a-i4822e.pdf

([16]) http://www.fao.org/3/a-I6767e.pdf

([17]) www.fao.org/docrep/016/ap358e/ap358e00.pdf

([18]) http://www.fao.org/docrep/016/ap353e/ap353e00.pdf; http://www.fao.org/docrep/016/ap354e/ap354e00.pdf; http://www.fao.org/icatalog/inter-e.htm

([19]) http://www.gfoi.org/methods-guidance/

However, this is such a sensitive area that the response rate is often very low and the data may be unreliable.

Much industrial waste is an unavoidable by-product, the type and volume of which is directly proportional to the volume of production and will depend on the technology used in the production process. Therefore, for each technology type, a waste factor can be produced. Many of the available statistics on industrial waste are the results of models based on these factors, together with information on industrial production and the distribution of the main technological processes used in the industry being assessed. A useful source for this is the EEA's report '*Development and application of waste factors: an overview*' (EEA, 1999), which provides an overview of waste factors, their derivation and application, and experiences in using them, based on reports and available literature. For municipal waste, direct surveys are not the best way to estimate volumes or composition. Their main disadvantage is that they are costly, and the respondents often have little idea of the real volume of waste they generate, or of its composition, resulting in large uncertainties in the resulting figures.

The most common estimation method for municipal waste is simply to weigh a sample of the waste collection vehicles before and after collection and to gross up to cover the whole population. The sample will need to cover vehicles collecting in a wide range of areas — urban, rural, wealthy, poor, with and without gardens, etc. — and covering several periods throughout the year, so that the sample can be seen as representative for the whole population and the whole year. Estimation of the composition of municipal waste is more complicated. Panels of households can be set up to monitor their waste generation and composition more closely and over time. Panels are basically small samples that remain constant over time, and therefore are well suited to monitoring trends. Because the panel will need to be very actively involved in weighing and analysing the contents of its waste bins, it is often necessary to pay the participants for their input, and this can be a serious limiting factor. Therefore factors for composition are often based on research projects and technical studies carried out in research institutes, sometimes, but not always, at the request of the relevant municipality or ministry.

Appendices references

EEA, 1999, *Development and application of waste factors: an overview,* EEA Technical Report No 37, European Environment Agency (https://www.eea.europa.eu/publications/technical_report_37).

FAO, 1980a, *Forest volume estimation and yield prediction, volume I,* Food and Agriculture Organization of the United Nations, Rome.

FAO, 1980a, *Forest volume estimation and yield prediction, volume II,* Food and Agriculture Organization of the United Nations, Rome.

FAO (1981), *Manual of forest inventory with special reference to mixed tropical forests,* FAO Forestry Paper No 27, Food and Agriculture Organization of the United Nations, Rome.

FAO (1989), *Sampling methods for agricultural surveys,* FAO Statistical Development Series No 3, Food and Agriculture Organization of the United Nations, Rome (www.fao.org/ES/ESS/menu3.asp).

FAO (1992), *Collecting data on livestock,* FAO Statistical Development Series No 4, Rome, Food and Agriculture Organization of the United Nations, Rome.

FAO (1995), *Conducting agricultural censuses and surveys,* FAO Statistical Development Series No 6, Food and Agriculture Organization of the United Nations, Rome (www.fao.org/es/ess/census/agcensus.asp).

FAO (1996), *Multiple frame agricultural surveys: volume 1: current surveys based on area and list sampling methods,* FAO Statistical Development Series No 7, United Nations Food and Agriculture Organization, Rome.

FAO (1998), *Multiple frame agricultural surveys: volume 2: agricultural survey programmes based on area frame or dual frame (area and list) sample designs*, FAO Statistical Development Series No 10, Food and Agriculture Organization of the United Nations, Rome.

IEA (2005), *Energy statistics manual,* International Energy Authority, Paris (https://www.iea.org/statistics/resources/manuals/).

UN (1982), *Concepts and methods in energy statistics, with special reference to energy accounts and balances,* United Nations Series F, No 29 (https://unstats.un.org/unsd/publications/catalogue?selectID=20).

UN (1987), *Energy statistics: definitions, units of measure and conversion factors,* United Nations Series F, No 44 (https://unstats.un.org/unsd/publications/catalogue?selectID=37)

UN (1991), *Energy statistics*: *a manual for developing countries*, United Nations Series F, No 56 (http://unstats.un.org/unsd/publication/SeriesF/SeriesF_56E.pdf).

UN, 2016, *International recommendations for energy statistics,* United Nations (https://unstats.un.org/unsd/energy/ires/IRES-web.pdf).