CHAPTER 2

APPROACHES TO DATA COLLECTION

[Parts shaded in grey – the unchanged text from the 2006 IPCC Guidelines]

Authors

Simon Eggleston (UK), Yogesh Kumar Tiwari (India), Chia Ha (Canada), Kenza Khomsi (Morocco), Veronika Ginzburg (Russia), Daniela Romano (Italy)

Contributing Authors

Sandro Federici (Italy)

Contents

Equations

Figures

Tables

Boxes

2 APPROACHES TO DATA COLLECTION

2.1 INTRODUCTION

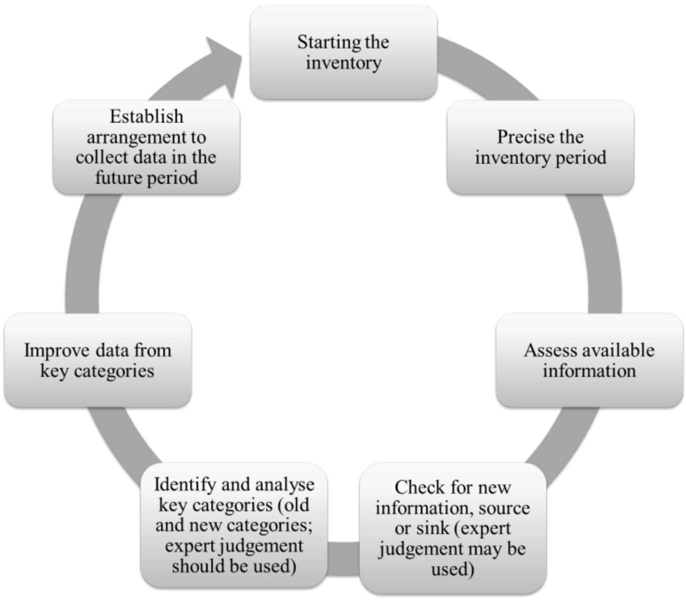
[New guidance in the 2019 Refinement].

Data collection is a key part of the national inventory system a country puts into place to regularly estimate and report greenhouse gas emissions and removals. Formalised data collection activities should be established, adapted to national circumstances of countries, and reviewed periodically as a part of implementing *good practice*. Data collection procedures are necessary for finding and processing existing data, (i.e., data that are compiled and stored for other statistical uses than the inventory), as well as for generating new data by surveys or measurement campaigns. During the data collection for the greenhouse gas inventory, interactions between the inventory compilers and stakeholders will take place, which may require the most time in the compilation process. A network of data providers ranging from national statistical agencies, ministries, international organizations, academia, economic sector such as industries, trade, transport, service sector, as well as government, households and others will be expected to provide information on an annual basis. Data collection may also include the direct collection, from administrative records and from monitoring stations, in coordination with the statistical system in place, of data by the inventory compilers. Other activities related to data collection 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.

When starting the inventory compilation for the first time, this process will need to be established for the whole inventory period. In the aim that potential *key categories* can be identified, it may be best to make estimates with readily available information, without expending too much effort in collecting new information. Resources can then be focused on improving data for those key categories. Inventory compilers should aim for completeness and focus on further improvements of next inventories in later years. Expert judgement should also be used initially to identify likely *key categories*: inventory compilers likely know, or can find out relatively easily, candidate categories such as major fossil fuel consumption, major agriculture activities, forest management, deforestation and major industries. In subsequent years, adding new categories should be easier with the relationships and processes already established. However, for next inventories, compilers should always be prepared to consider new data sources should they become available.

In established inventories, new sources of greenhouse gas emissions or sinks may be identified. 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 set out in Chapter 4, Methodological Choice and Identification of Key Categories. In these cases, it is *good practice* to estimate if the source or sink strength is of similar magnitude to *key categories* in order to assess and prioritize the effort required: if a new source or sink is estimated to be on a scale similar to the *key categories* then it is *good practice* to use Tier 2 or 3 method. Such initial estimates can be very rough and may be based on expert judgement as they are intended to guide resource use and not contribute to the final inventory totals.

Figure 2.1 Process diagram for data collection



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

- Identify, evaluate and document data sources.
- Focus on the collection of data needed to improve estimates of key categories that are the largest.
- Establish a system for continuous improvement:
  - (i) review data collection activities and methodological needs on a regular basis, to guide progressive, and efficient, inventory improvement;
  - (ii) put in place data collection activities (resource prioritisation, planning, implementation, documentation etc.) that lead to continuous improvement of the datasets used in the inventory;
  - (iii) collect information about uncertainty which is an integral part of data collection for both emission factors and activity data (Chapter 3 discusses uncertainties);
  - (iv) collect data/information at a level of detail appropriate to the method used;
  - (v) establish agreements with data suppliers to support consistent and continuing information flows.

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. It is applicable to emission factor, activity, and uncertainty data collection. It covers:

- developing a data collection strategy to meet data quality objectives regarding timeliness, and also consistency, completeness, comparability, accuracy, and transparency using guidance provided in Chapter 6, QA/QC and Verification, of this Volume;
- data acquisition activities including generating new source data, dealing with restricted data and confidentiality, and using expert judgement;
- turning the raw data into a form that is useful for the inventory.

Advice related to selecting emission factors focuses on understanding and generating measured data as well as addressing where to find and when to use default factors. Guidance on activity data focuses on generating and using new census & survey data as well as providing guidance on the use of existing international data sets.

The chapter draws on information from a range of institutions and where possible additional documents have been identified and referenced so that users can find more detailed information. Sector specific data collection issues - like selecting the appropriate activity data for a particular category of emissions by sources and removals by sinks - are described in the sector specific Volumes 2-5.

2.2 COLLECTING DATA

[New guidance in the *2019 Refinement*].

This section provides general guidance for collecting existing data, generating new data, and adapting data for inventory use. The guidance is applicable to emission factors, activity and uncertainty data collection. It discusses separately specific issues relating to new data and existing data. Specific guidance for the collection/calculation of emission factors and the collection of activity and uncertainty data is provided subsequently. Throughout the data collection activities the inventory compiler should maintain QA/QC records about the data collected according to the guidance provided in Chapter 6 of Volume 1. While collecting data it is *good practice* to be aware of future data collection needs.

In some cases, it may be possible to use emissions data directly from facilities, e.g. measured emissions such as the EU ETS (this is discussed in section 2.3).

Following the *2019 Refinement*, it should be possible to provide a Tier 1 estimate for every category. The sectoral volumes contain default Tier 1 emission factors and parameters that can be used. In the absence of available activity data, applying these guidelines will allow estimates based on either surrogate information or expert judgement.

When compiling an inventory or estimates for a specific source for the first time, it is *good practice* for inventory compilers to assess existing data and to acquire data in the following ways in order of priority:

1. Use existing published national or international data (section 2.2.1).
2. Engage in cooperation with data suppliers to provide tailored data sets from their information (section 2.2.1).
3. Modify existing data sets to meet the inventory requirements (e.g. where data is not collected annually, convert from financial year to calendar year, adjust for different classifications of sources or fill gaps in territorial coverage) (section 2.2.3).
4. Generate new data:
   (i) make measurements (section 2.2.2 & 2.2.4);
   (ii) use census and survey data (section 2.2.5, Annex 2A.2);
   (iii) coordinate with National Statistical Offices to undertake new surveys targeting inventories relevant sectors.
5. Use surrogate data (section 2.2.1: Surrogate Data).
6. If the above approaches are unsuccessful and as a last resort, it is *good practice* to use expert judgement (section 2.2: Expert Judgement).

It is *good practice* to focus resources on categories that have been identified as *key categories*. Improving data for less important categories can be done in later years.

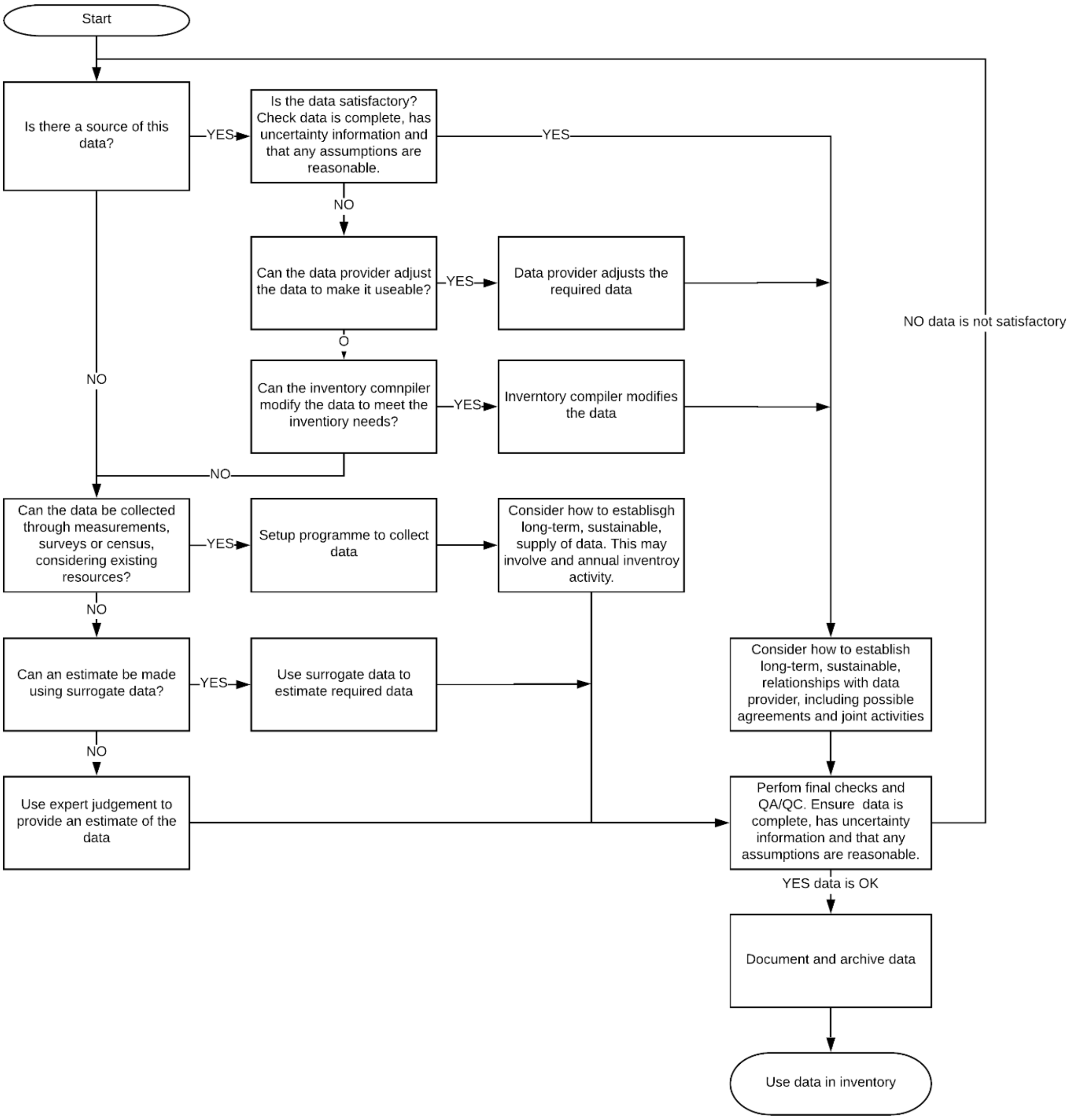
As data is collected, it is *good practice* for it to be checked considering the following points:

- Is it complete? Does it cover the entire territorial area of the inventory and/or the entire population of sources within the source category? Does the data represent collection for a complete calendar year?
- Does the data have associated uncertainty information? The uncertainty information should include the uncertainty range expressed in terms of 2.5 and 97.5% percentiles and may include information on the shape of the uncertainty distribution?
- What assumptions underlie the data? (E.g. is a survey representative? Is a census complete?).
- What measurement methods are used? Are they reliable? Are time series consistent?

Feedback mechanism, to provide comments/questions from the compiler to the data provider, should be formalized in the data supply agreement. If problems are identified in the data, the inventory compilers should approach the data supplier and ask that the issues be addressed. If this fails, other sources of the data, described above, can be used.

Following collection of data, all sources of data should be archived and documented together with the any processing and any assumptions used for data collection. Archiving and documenting will allow the data to be reconstructed and will form the basis of future inventories. It is *good practice* to establish sustainable relationships with data producers to reduce the resources (human, time and financial resources) required for data collection in subsequent years.

Figure 2.2 Outline of data collection steps and decisions



Maintaining supply of inventory data

Often working with a data provider benefits both the inventory and data provider. It is *good practice* to establish sustainable relationships with data producers to reduce the resources required for data collection in subsequent years. Developing long-term relationships with data suppliers can lead to improvements in the original data. Codifying data collection decision trees and collection steps with the data suppliers can lead to benefits for everyone. An improved understanding of the data could lead to improved inventories. Original data might be adjusted to reflect reporting needs. .

It is *good practice* to engage data suppliers in the process of inventory compilation and improvement by involving them in activities such as:

- offering an initial estimate for the category, pointing out the potentially high uncertainties and inviting potential data suppliers to collaborate in improving estimates;
- scientific or statistical workshops on the inventory inputs and outputs;
- specific contracts or agreements for regular data supply;

- regular updates on the methods that use their data;
- establishment of terms of reference or memoranda of understanding with data providers to clarify what is needed for the inventory, how it is derived and provided to the inventory compiler and when. Priority should be given to government authorized data providers.

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 links to data providing organisations.

Where appropriate, it may be useful to explore existing or new legal arrangements as means of guaranteeing the delivery of data to the inventory.

Restricted data and confidentiality

Sometimes data may be provided to inventory compilers on a confidential or restricted basis. Confidential data can lead to an inventory lacking transparency and wherever possible the use of confidential data should be avoided. However, this is not always possible. 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 possible imperfections in the data. Sometimes, however, the organisation simply does not have the resources required to compile and check the data. It is *good practice* that inventory compilers take the following steps to avoid this issue.

1. It is advisable, where possible, to cooperate with data providers to find solutions to overcome their concerns and allow the open use of the data by:
   (i) explaining the intended use of the data;
   (ii) agreeing, in writing, to the level at which it will be made public;
   (iii) identifying the increased accuracy that can be gained through its use in inventories;
   (iv) offering cooperation to derive mutually acceptable data sets;
   (v) and/or giving credit/acknowledgement in the inventory to the data provided.
2. When confidentiality cannot be avoided, ways to aggregate or mask the data should be investigated (see 3).
3. Aggregation of smaller subcategories may be possible to aggregate the emission estimates into a larger category to avoid breaking the confidentiality. Aggregation techniques should be selected to avoid the possibility that the confidential data could be reconstructed using the published inventory. Care should also be taken to minimize the aggregation as much as possible so as to be as transparent as possible
4. If masking or aggregating data is unsuccessful at preserving confidential data, it may be appropriate to look at other sources of data and avoid the use of confidential data favoring greater transparency of the final inventory. Attention should also be paid to any reporting guidelines that might apply.

The protection of confidentiality is one of the fundamental principles of a national statistical agency (NSA[1] - see: http://unstats.un.org/unsd/methods/statorg/) (UN, E/RES/2013/21). 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 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, especially where a few companies dominate the sector, than for other data.

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, are normally the only ones that can carry out this additional treatment of the raw data.

Some countries have special arrangements to mask data (i.e., 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

[1] Any main national official data collection organization is referred to here as national statistical agency.

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.

Many agencies collect ancillary data during operations for other purposes, such as registration of businesses or vehicles, collection of taxes, granting of licences, allocation of grants and subsidies. Such information is usually also covered by confidentiality clauses. In general, such clauses foresee 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 for cost savings.

Expert judgement

Expert judgement on methodological choice and choice of 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, to select data from a range of possible values or make judgements about uncertainty ranges as described in Section 3.2.2.3. Experts with suitable backgrounds can be found in government, industrial trade associations, technical institutes, industry and universities.

The goal of expert judgement may be choosing the proper methodology; the parameter value from ranges provided; the most appropriate activity data to use; the most appropriate way to apply a methodology; or determining the appropriate mix of technologies in use. A degree of expert judgement is required even when applying classical statistical techniques to data sets, since one must judge whether the data are a representative 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[2]. In all cases, the aim is to be as representative as possible in order to reduce possible bias and increase accuracy. Formal methods for obtaining (or eliciting) data from experts are known as expert elicitation, see Annex 2A.1 for details.

Despite endeavoring to collect all the data needed and document all of the data's associated uncertainties, there may cases when no data is available and then the inventory compiler will need to rely on expert judgement to provide the information. Experts should be asked to estimate the missing data based on their expertise. In order to initiate such discussions and to provide a starting point for their considerations the following inputs can be used:

- If there are other countries with sectors in a similar stage of economic development, management practices and/or soil-climatic conditions, consider extrapolating from the similar country's reports.
- Experts might be able to infer national data from regional information with uncertainties.
- There may be information that is statistically or physically related to the parameters needed (e.g. industrial production is generally correlated with fuel consumption).
- Consider if the source exists. Some industrial sources may not occur in all countries. International data on trade and production often provide ways to confirm the existence of an industry within the country.
- If the category is likely to be very small compared to other categories, order of magnitude estimates can be made and used.

2.2.1 Gathering existing data

Although the list below is not exhaustive, it provides a starting point for possible sources of country specific data:

- National Statistics Agencies.
- Facility level emission data such as that reported to Pollutant Release and Transfer Registers (PRTR) or to the EU ETS may sometimes be used for estimating emissions from industrial installations. Note that it is unlikely that this type of data is directly useable in an emission inventory. This is discussed further in section 2.2.3.
- Sectoral experts, stakeholder organisations (e.g. industry and trade organizations, large-scale industries such as energy producers and petro-chemical plant).

[2] Methods for characterizing sampling distributions for the mean are described by Cullen and Frey (1999), Frey and Rhodes (1996), and Frey and Burmaster (1999).

- Other national experts.
- IPCC Emission Factor Database.
- Other international experts.
- International organisations publishing statistics e.g., United Nations, Eurostat or the International Energy Agency, OECD and the IMF (which maintains international activity as well as economic data).
- Reference libraries (National Libraries).
- Scientific and technical articles in environmental books, journals and reports.
- Universities.
- Web search for organisations & specialists.
- National Inventory Reports from Parties to the United Nations Framework Convention on Climate Change.
- The Greenhouse Gas Reporting Program (GHGRP) Greenhouse Gas (GHG) data.
- International databases such as:
  - (i) the AFOLU FAOSTAT Emissions Database of greenhouse gas emissions from agriculture[3];
  - (ii) the International Fertilizer Industry Association (IFIA)[4];
  - (iii) the FAO Global Livestock Environmental Assessment Model (GLEAM)[5];
  - (iv) the International Rice Research Institute (IRRI)[6];
  - (v) geospatial products for land use, land management, soils and climate (the availability of these data was reviewed by Smith et al (2012));
  - (vi) remote sensing data.

Screening of available data

It is best to start data collection activities with an initial screening of available data sources. Inventory compilers need to identify and communicate with data providers in their countries. This will be an iterative process where 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.

The purpose for which data were originally collected may be an important indicator of reliability. Regulatory authorities and official statistical bodies have a responsibility to take representative samples and accurate measurements, and so they often adopt agreed standards. 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. These preliminary data can be used if their validity is documented and can be checked against the data quality objectives set by the quality management system described in Chapter 6.

Refining Data Requirements

Once the inventory compiler has selected a data set, unless published data simply can be used in their original form, the next step will be 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, QA/QC and Verification). A clear definition of data requirements will ensure that when data are delivered they are as expected. The specification should include details such as:

- definition of the data set (e.g., time series, sectors and sub-sector detail, national coverage, requirements for uncertainty data, emission factors and/or activity data units);

[3] http://www.fao.org/faostat/en/#home.

[4] www.fertilizer.org/ifa/statistics.asp.

[5] http://www.fao.org/gleam/results/en/#c300947.

[6] http://ricestat.irri.org:8080/wrsv3/entrypoint.htm.

- definition of the format (e.g., spreadsheet) and structure (e.g., what different tables are needed and their structure) of the data set;
- description of any assumptions made regarding national coverage, the sectors included, representative year, technology/management level, and emission factors or uncertainty parameters;
- identification of the routines and timescales for data collection activities (e.g., how often is the data set updated and what elements are updated);
- reference to documentation and QA/QC procedures;
- methodologies adopted for preparation of GHG emission factors, especially for developing countries[7];
- contact name and organisation;
- date of availability.

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

Choosing between published national and international data

In most cases it is preferable to use national data since national data sources are typically more 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 data sets 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, though 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. Crosschecking national data sets with any available international data can help to assess completeness and identify possible problems with either data set. Developing countries may check for QA/QC of international data if national data are sparse or not available.

Surrogate data

It is preferable to use data that are directly related to the item being quantified rather than to use surrogate data (i.e., alternative data that have a correlation with the data that they are replacing). In some cases, however, directly applicable data may be unavailable or have gaps (e.g., if survey and sampling programmes may be infrequent). In these cases, surrogate data can help fill gaps and generate a consistent time series or a country average. For example, where a country has information to apply a higher tier method for some but not all of its facilities, then surrogate data can be used to fill the gaps. The surrogate data should be physically and statistically related to the emissions from the set of facilities for which information is not available. These alternative data should be selected based on country-specific circumstances and information, and a relationship between the data and emissions (i.e., an emission factor) developed using information from a representative subset of facilities whose emissions are known. The use of surrogate data to obtain an initial estimate of an emission or removal can help prioritise resources.

In selecting and using surrogate data to estimate emissions or removals, it is *good practice* for countries to perform the following steps:

- confirm and document the physical relationship between emissions/removals and the surrogate activity data;
- confirm and document a statistically significant correlation between emissions/removals and the surrogate activity data;
- using regression analysis, develop a country-specific factor relating emissions/removals to the surrogate data.

An example of this approach is given in Box 2.1 and further explanation and equation (Equation 5.2) given in Section 5.3 of Chapter 5, Time Series Consistency of the *2006 IPCC Guidelines*.

[7] For example, GHGs data over India are reported in Ganesan *et al.*, 2017; Fadnavis *et al.*, 2016; Tiwari et al, 2014; Tiwari *et al.*, 2011.

BOX 2.1

EXAMPLE OF USING ALTERNATIVE DATA TO APPROXIMATE ACTIVITY DATA

The U.S. receives emission estimates for SF_6 associated with electrical equipment based on a mass-balance approach from electric power systems representing about 35 percent of the total length of U.S. transmission lines. (In the U.S., transmission lines are defined as lines carrying electricity at or above 34.5 kV.) To estimate emissions from the remaining systems, the U.S. uses kilometres of transmission lines as alternative activity data. In the U.S., SF_6 is primarily used in equipment rated at or above 34.5 kV, and kilometres of transmission lines are therefore expected to be a good predictor of emissions. In addition, statistical analysis has demonstrated a high correlation between emissions and kilometres of transmission lines. Given these relationships, the U.S. uses regression factors relating transmission kilometres to emissions. These factors are then applied to the total transmission kilometres of the systems whose emissions are being estimated. Germany has also used the length of transmission lines to estimate emissions from closed pressure systems for a set of utilities that did not respond to an industry survey. Estimates are based on the electric power systems from utilities for which both transmission kilometres and emissions data were available. The resulting estimates were later confirmed by more comprehensive surveys in subsequent years. Information on equipment banks, available nationally from equipment manufacturers and distributors were used to estimate emissions from sealed-pressure systems.) Transmission kilometres are likely to be a good predictor of emissions where most SF_6 is used in high voltage transmission equipment, as in the U.S. Where a high percentage of SF_6 is used in medium voltage distribution equipment or in gas-insulated substations, another type of data may be appropriate, such as the combined length of transmission and distribution lines or the number of substations. Combinations of these or other types of data may also be used although this increases the probability that one or more of the types of data will not be available for all the systems whose emissions are to be estimated.

2.2.2 Generating new data

It may be necessary to generate new data if representative emission factors, activity data or other estimation parameters do not exist, or cannot be estimated from existing sources. Generation of new data may entail measurement programmes for industrial process or energy related emissions, sampling of fuels for carbon content, land-use change and forestry sampling activities, or new census or surveys for activity data. Generation of new data is best undertaken by those with appropriate expertise (e.g., measurements carried out by competent organisations using appropriately calibrated equipment or surveys and censuses by any national statistical authority). These activities are often resource intensive and are most appropriately considered when the category is *key* and there are no other options. To optimise resource use it is recommended as far as possible to generate the required data from an extension of existing programmes rather than the initiation of very new ones. More specific details for emission factor and activity data are outlined in the respective sections of this chapter. Where guidelines exist for activities that are defined in detail by other official bodies, such as statistical offices and measurement standards committees, these are also referenced in these sections.

Generating data by measurement

Measurements should be used in the context of advice in the sectoral Volumes 2-5, for example to determine or revise emission factors, destruction/abatement efficiency factors and activity rates. Measurements can also be used to quantify greenhouse gas emissions directly or to calibrate and verify models that are used to generate data.

When considering using measurement data it is *good practice* to check whether it covers a representative sample, i.e., that is typical of a reasonable proportion of the whole category – and whether a suitable measurement method has been used. The best measurement methods are those that have been developed by official standards organisations and field-tested to determine their operational characteristics.[8] Using standardised measurement methods improves the consistency of measured data and provides the inventory compiler with additional information about the method such as statistical uncertainty levels, lower detection limits, sensitivity, and upper limits of measurement etc. The International Standards Organisation (ISO) standards, European Standards (EN) or suitable validated national standards of, e.g., U.S. Environmental Protection Agency (USEPA), or the Association of German Engineers (Verein Deutscher Ingenieure, VDI), may meet these criteria. It is *good practice* for the inventory compiler to document any measurement or quality management standards that have

[8] For example, repeatability, reproducibility, detection limit, tolerance to interference etc.

been used, and to bear in mind the data requirements of the uncertainty analysis in Chapter 3, Uncertainties, of Volume 1.

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 2.1 sets out the elements of such an approach.

TABLE 2.1
GENERIC ELEMENTS OF A MEASUREMENT PROGRAMME

Measurement objective	Clear statement of the parameter(s) to be determined, e.g., HFC-23 emissions from HCFC-22 production.
Methodology protocol	Description of the measurement methodology to be used. This should include: • The components to be measured and any associated reference conditions; • Methods to ensure that representative samples are taken that reflect the nature of the source category and the measurement objective [a]; • The identification of any standard techniques to be used; • The analytical equipment needed and its operational requirements; • Any source/sink or installation access requirements; • Any accuracy, precision or uncertainty requirements; • Data capture requirements to be met; • Calibration material and methods requirements; • QA/QC regimes to be followed.
Measurement plan with clear instructions to the measurement personnel	Measurement plan specifies for those carrying out the measurements that includes: • Number of sampling points for each parameter to be measured and how these are to be selected; • Number of individual measurements to be made for each sampling point and set of conditions; • Measurement dates and periods of the measurement campaign; • Reporting arrangements; • Additional source or process related information to be collected to enable data processing or interpretation of the results; • 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	Data processing requirements, including: • Reporting procedures that will form an account of the measurements, the description of the measurement objectives, and the measurement plan; • Documentation requirements to enable the results to be traced back through the calculations to the collected basic data and process operating conditions.

[a] When making eco-system measurements particular care is required in defining the sampling requirements – see Volume 4.

General guidance to ensure the quality of measured data to determine better emission factors and other parameters is provided in Section 2.2.2.

Relationship of data to 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 for data to drive them. 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 match, including sectors/process/abatement, location, land type, compound and years included.

Gaps in data sets

Greenhouse gas inventories require consistent estimates across time series and between categories. This section introduces approaches to fill gaps if data are missing for one or more years or the data do not represent the year or national coverage required. Examples of data gaps or inconsistencies and guidance for addressing them are presented below.

- **Filling 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 at intervals of every fifth or tenth year. Time series data may need to be inferred to compile a complete annual estimate for the years between surveys, and for fore- and back- casts (e.g., where estimates are needed for 1990 – 2004 and survey data are only available for 1995 and 2000). Chapter 5, Time Series Consistency, provides details on splicing and extrapolation methods to fill these gaps.
- **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 have been subject to further revision of historic data to correct errors or to update new methodologies. It is important that the inventory compiler look for these changes in the source data time series and integrate them into the inventory. Chapter 5 of this Volume contains more guidance on this issue.
- **Incorporating improved data:** While the ability of countries to collect data generally improves over time so they can implement higher tier methods, the data may not necessarily be suitable for earlier years for the higher tiers. For example when direct sampling and measurement programs are introduced there may be inconsistencies in the time series, as the new program cannot measure past conditions. Sometimes this can be addressed if the new data are sufficiently detailed (e.g., if emission factors for modern abated plant can be distinguished from those of older unabated plant) and the historic activity data can be stratified using expert judgement or surrogate data. Chapter 5 provides more details on methods of incorporating improved data consistently across a time series.
- **Compensating for deteriorating data:** Splicing techniques, as described in Chapter 5 on Time Series Consistency, can be used to manage data sets that have deteriorated over time. 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 these data sets may contain different definitions, classifications and levels of aggregation. The international data sources discussed in the activity data section (see Section 2.2.5) may provide another source of relevant activity data.
- **Incomplete coverage:** When data do not fully represent the whole country, e.g., measurements for three of ten plants or survey data of the agricultural activity for 80 percent of the country, then the data can still be used but needs to be combined with other data to calculate a national estimate. In these cases expert judgement (see Section 2.2 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, and that data providers be requested to compute representative yearly data with a complete coverage.

Combining data sets numerically

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 95 percent confidence limits, or by combining summary statistics using the relationships set out in statistical textbooks. It also is possible to combine measurements of a single quantity made using different methods that produce results with different underlying probability distributions. However, 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 plant 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 3 by treating each stratum in the same way as an individual category.

Inhomogeneity may be identified by specific knowledge of the circumstances of individual plants or technology types, or by a detailed data analysis, e.g., scatter plots of estimated emissions/removals 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 lying 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., plant in start-up conditions) that should really be represented separately in the inventory estimate.

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 under-estimates 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 Volume 2-5, multi-year averaging may be the best or even the only way to estimate data for a single year. In cases where the inventory compiler finds it is impossible to acquire activity data for each year it is *good practice* to use the methods given for ensuring time series consistency (Chapter 5 of this Volume). In the case of emission factors with high or uncertain annual variability then multi-year averaging can improve the quality of the overall estimate provided the underlying conditions and processes remain unchanged. For example, for the growth of various tree species in a year there may be much higher confidence in the average annual growth rate over a period of years, than in an individual year growth. Multi-year averaging applies to the emission factors or similar parameters, not activity data.

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 it is used consistently over the time series and 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 this is acceptable. 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.

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 or removals 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.

2.2.3 Emission factors and direct measurement of emissions

For categories that are NOT key categories the default values given in these guidelines can be used. For key categories it is *good practice* to develop country specific emission factors. This section provides generic advice for the derivation or review of emission factors or other estimation parameters; this includes specialised literature sources, using measured data, and further remarks on combining data sets. It is *good practice* when developing emission factors or other estimation parameters to follow the stepwise approach to data collection described above:

1. Identify emission factors (EF) which to be developed.
2. Table 2.2 lists the main parameters and effecting emission or removal rates.
3. Developing a strategy for acquiring the data.
4. Collect data for these parameters, including their uncertainty.
5. Calculate EFs to use with the methodology provided in the sectoral chapters or develop a country specific model.
6. Verify that the obtained EF is representative.
7. Prepare proper documentation on the data and methodology used for EFs estimation[9] and its uncertainty.

It is *good practice* for inventory compilers to acquire data following the suggested options:

- option 1: ask stakeholders, concerned ministries, researchers and any other data providers;
- option 2: collect data from international databases;
- option 3: consider data from other countries with similar circumstances (population density, type of industry, technologies, etc.);
- option 4: ask for expert judgments.

While collecting data for developing country-specific emission factors the main category specific parameters effecting the emissions and removal are listed in the Table 2.2. See more detailed guidance on factors influencing emission factors sector-specific guidance chapters.

TABLE 2.2
MAIN PARAMETERS THAT EFFECT EMISSIONS AND REMOVALS

Sector	Category	Sub-Category	EF Sensitive parameters
Energy	Fuel combustion	Stationary	CO_2: Variable carbon content of coal and non-commercial fuels. CH_4 and N_2O: Technology type.
		Mobile	CO_2: Portion of biofuels. CH_4 and N_2O: • Vehicle type and engine technology (including domestic aircrafts); • Emission control technologies fitted to vehicle types in the fleet; • Fleet age distribution; • Maintenance effects.
	Fugitive emissions	Coal mining and handling	Characteristics of seams.
		Oil & gas	• Composition profiles of gases from particular oil and gas fields; • Efficiencies of the specific control measures used; • Maintenance - management; • Composition of reservoir.

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

Industrial Processes (many processes do not occur in all countries)			• Technology types; • Composition of feedstocks; • Leakage rates in F-gases related sectors; • N_2O atmospheric emission reduction technology; • Emission reduction technologies and their efficiencies; • GHG by-product generation rates; • Frequency and duration of process disturbances.
AFOLU	Livestock population and production system characterization	Enteric Fermentation	• Livestock species and categories; • Animal Weight; • Milk production, fat and protein content; • Quantity, quality and type of feed.
		Manure Management	• Type of AWMS (solid storage, anaerobic lagoons etc.); • Manure characteristics; • Timing of storage; • Length of storage; • Seasonal and daily temperature variation; • Climate; • Time and temperature distribution between indoor and outdoor storage.
	Agricultural soils	Direct and indirect N_2O emissions from managed soils	• Synthetic fertilizers consumption; • Animal manure applied to soils; • Production of leguminous (N-fixing) crops; • Crop residue incorporation; • Organic soil cultivation; • Soil type; • Type of AWMS (solid storage, anaerobic lagoons etc.); • Soil management practices; • Water management information; • Information about the irrigation system; • Climate.
Waste	Solid Waste		• Total municipal solid waste per year; • Waste component; • Life of product.
	Wastewater handling		• Wastewater treatment technology in the country; • Population of the country.
	Waste incineration		• Type of incinerated waste per year; • Amount of incinerated waste per year.

Volumes 2-5 provide advice on the selection and use of emission factors or other estimation parameters for specific categories.

Literature sources

Inventory compilers commonly rely on the available literature to find emission factors or other estimation parameters. Table 2.3 lists a variety of potential literature sources in order of descending likelihood of the data being representative and appropriate for national circumstances. For key categories 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 emission factors contained in the Emission Factor Database (EFDB), or other literature values can be used. These factors should reflect national circumstances. The order of presentation in Table 2.3 is indicative only, and inventory compiler should assess each data source individually to make a determination of suitability.

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 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.

Literature reviews should be fully documented so that the data used for the inventory is transparent (see Chapter 6, QA/QC and Verification). It is also helpful to record the sources not used, providing an explanation of why, to save time in later literature review activities.

TABLE 2.3
POTENTIAL SOURCES OF LITERATURE DATA

Literature type	**Where to find it**	**Comments**
IPCC Guidelines	IPCC website	Provide agreed default factors for Tier 1 methods but may not be representative of national circumstances.
IPCC Emission Factor Database (EFDB)	IPCC website	Described in more detail below. May not be representative of processes in your country or appropriate for *key category* estimates.
EMEP/EEA Emission Inventory Guidebook (EEA, 2016)	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	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	National reference libraries, environmental press, environmental news journals	Reliable if representative. Can take time to be published.
Country-specific data from international or national peer reviewed journals	National laboratories	Reliable. Need to make sure the factors are representative and that standard methods are used.
National testing facilities (e.g., road traffic testing facilities)	Industrial process regulating authority	Regularly updated and plant-specific. Quality is dependent on the regulatory requirements, which may not extend to the methods used for estimating/measuring.

Emission regulating authority records and papers, or pollution release and transfer registries	Specific trade association	Sector-specific and up-to-date. QA/QC is needed to check for bias in data and to ensure the test conditions and measurement standards are understood.
Industry, technical and trade papers	Publications, libraries, and Web search	
Other specific studies, census, survey, measurement and monitoring data	Universities (environmental, measurement and monitoring departments)	Need to make sure the factors are representative and that standard methods are used.
International Emission Factor Databases: OECD	OECD website	Useful defaults or for crosschecking. May not be representative of processes in your country or appropriate for *key category* estimates.
Emission factors or other estimation parameters for other countries	National Inventory Reports from Parties to UNFCCC, other inventory documentation, web search, national library	Appropriate for inventory use. Useful defaults 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 organization of the United Nations	(www.fertilizer.org/ifa/statistics.asp); (www.faostat.fao.org/).	Data can be used in national GHG inventories especially for developing countries.
Trade data	https://comtrade.un.org/.	Repository of official international trade statistics and relevant analytical tables

Potential data sources focusing on developing countries are the following: Harmonized world soil database (FAO, IIASA, ISRIC, ISSCAS and JRC 2012) and GlobalSoilMap.net project, WorldClim, NEO and CRU, GLCC-IGBP, MODIS, Land Use/Cover Area Survey LUCAS by Eurostat, Regional Centre for Mapping of Resources for Development in Africa RCMRD[10], International Rice Research Institute (IRRI) which provides CH_4 emission measurements from major rice producing countries including China, India, Indonesia, Thailand, Philippines, etc. These data may be available by web search and data sets can be used for preparing GHG inventories for developing countries as well as other regions.

IPCC Emission Factor Database

The Emission Factor Database (EFDB) is a continuously revised web-based information exchange forum for emission factors and other parameters relevant for the estimation of emissions or removals of greenhouse gases at national level. The database can be queried over the internet via the home pages of the IPCC, IPCC-NGGIP or directly at http://www.ipcc-nggip.iges.or.jp/EFDB/main.php.[11] The IPCC distributes a CD-ROM with a copy of the database and a query tool at regular intervals.[12] It is designed as a platform for experts and researchers to communicate new emission factors or other parameters to a worldwide audience of potential end users. The EFDB is intended to become a recognised library where users can find emission factors and other parameters with background documentation or technical references. The criteria for inclusion of data in the database (see Figure 2.3) are:

- *Robustness*: The value would be unlikely to change, within the accepted uncertainty of the methodology, if there were to be a repetition of the original measurement programme or modelling activity.

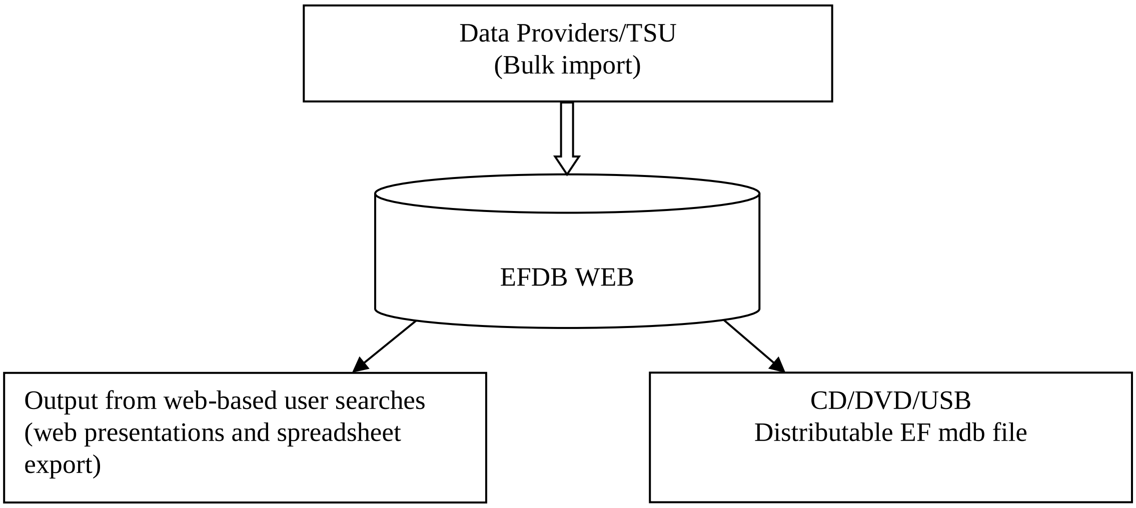
[10] www.rcmrd.org/.

[11] Information, including manuals, on how to retrieve data from or contribute new data to the EFDB can also be found at this web site: www.ipcc-nggip.iges.or.jp/EFDB/.

[12] To receive a copy of the EFDB DVD, please contact IPCC NGGIP Technical Support Unit.

- *Applicability:* An emission factor can only be applicable if the source and its mix of technology, operating and environmental conditions, abatement, and control technologies under which the emission factor was measured or modelled are clear, and allow the user to see how it can be applied.
- *Documentation:* Access information to the original technical reference is provided to evaluate the robustness and applicability as described above.

Figure 2.3 Process for including data in the EFDB



The EFDB invites experts and researchers all over the world to populate the EFDB with their data. The proposal of new emission factors (and other parameters) from data providers will be assessed by the Editorial Board of the EFDB for inclusion into the database. When the proposed new data comply with well-defined quality criteria of robustness, applicability and documentation they are included in the database. These procedures enable the user to judge the applicability of the emission factor or other parameter for use in their inventory and the responsibility of using this information appropriately however will always remain with the users.

Additionally, developing countries should focus mining of existing emissions data from the regional research centres conducting GHGs measurements to derive emission factors (see Table 2.3 and references). In addition to use of the EFDB to find new EF for use by inventory compilers, inventory compilers should also consider populating the EFDB with their country-specific data so that other countries with similar circumstances may consider it for use in the GHG or for assessment of their own EF. Moreover, if there is insufficient information on emissions data, then it is necessary that these countries undertake measurement programs in an effective and robust manner[13].

Data obtained by measurements

This section applies the guidance in Section 2.2.2 to assessing the quality of measurement data for determination of emissions, emission factors and abatement or destruction efficiencies. Volume 4 provides specific guidance on the use of samples and surveys in Agriculture, Forestry, and Other Land Use (AFOLU) Sector.

In this approach the emissions can be determined directly (i.e., using continuous emission monitoring systems) or calculated. Where emissions depend on variable combustion, process and operating conditions, and technologies (e.g., methane and nitrous oxide from combustion), direct monitoring is likely to be the most accurate way to determine emissions.

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 extraneous components. For example stack measurements may exclude losses to the atmosphere through evaporation or poorly burned fuel (that is emitted as volatile organic compounds (VOC); these should be included in the reported emissions totals. More details of measurement issues are included in the Industrial Processes and Product Use (IPPU) Volume.

In implementing the elements of measurement programme identified in Section 2.2.2 is *good practice* to:

[13] For example, Ganesan *et al.*, 2017; Tiwari *et al.* (2011-2014 are measurement programs of Indian sector).

- distinguish between different components in a mixed fuel/raw material 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 of exhaust gases;
- use instruments with known performance characteristics or perform relative accuracy audits against established standard reference methods.

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 greenhouse gas emission 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/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.

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 is 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 2.2.3, 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.

It is also *good practice*, as part of the measurement programme to include in the scope of a monitoring protocol how and other measurements are to be carried out, if the fuel/raw materials are not dry or there are contaminants that could adversely affect the measurement process, moisture.

Quality management is an important factor to take into account. ISO 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. It also encourages a quality management system that should cover traceable calibration artefacts; taking and storing samples; any subsequent analysis; and the reporting of results. The standards listed in Table 2.4 are relevant to greenhouse gas emissions measurement and should be used where applicable.

698

TABLE 2.4 STANDARD MEASUREMENT METHODS FOR EXHAUST GAS		
	Existing international standard methods	**Other widely used standard methods**[4]
CO_2	ISO 12039:2001 Stationary source emissions - Determination of carbon monoxide, carbon dioxide and oxygen - Performance characteristics and calibration of an automated measuring method[1] ISO 10396:2006 Stationary source emissions - Sampling for the automated determination of gas concentrations	US EPA Method 3 - Gas analysis for the determination of dry molecular weight US EPA Method 3A - Determination of oxygen and carbon dioxide concentrations in emissions from stationary sources (instrumental analyser procedure)
N_2O	ISO 11564:1998 Stationary source emissions - Determination of the mass concentration of nitrogen oxides - Naphthylethylenediamine photometric method	Standard being developed by ISO TC 264 – Air Quality
Gas velocity	ISO 10780:1994 Air Quality - Stationary source emissions - Measurement of velocity and volume flow rate of gas streams in ducts. *S-Type pitot tube* ISO 3966:1977 Measurement of fluid flow in closed conduits - velocity area method using Pitot static tubes[2]. *L-Type Pitot tube* ISO 14164:1999 Stationary source emissions. Determination of the volume flow rate of gas streams in ducts -automated method. *Dynamic pressure method for continuous, in situ/crossduct, measurements*	US EPA method 1 - Sample and velocity traverses for stationary sources US EPA Method 1A - Sample and velocity traverses for stationary sources with small stacks or ducts US EPA Method 2 - Determination of stack gas velocity and volumetric flow rate (Type S pitot tube) (or alternatively Methods 2F, 2G, 2H and CTM-041)[5]
General[3]	ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories ISO 10012:2003 Measurement management systems - Requirements for measurement processes and measuring equipment	PrEN 15259:2005 Air Quality – Measurement of stationary source emissions - measurement strategy, measurement planning and reporting, and design of measurement sites EN61207-1:1994 Expression of performance of gas analyzers - Part 1 General
	Standards under development	
CH_4	None	US EPA Method 3C - Determination of carbon dioxide, methane, nitrogen and oxygen from stationary sources (i.e., landfills) Standard being developed by ISO TC 264 - Air Quality
H_2O		EN 14790[6] US EPA Method 4 - Determination of moisture content in stack gases

PFC, SF_6, HFC, FCs	None	(N.B. Where available sector specific methodologies are referenced in the sector specific volumes)

[1] This standard describes the performance characteristics, detection principles and the calibration procedures for automated measuring systems for the determination of carbon dioxide and other substances in the flue gases emissions from stationary sources. The reported concentration range of this standard is 6 - 62500 mg m-3 with a measurement uncertainty of <10 percent of the measured value.

[2] This standard has been withdrawn pending revision; nevertheless, it is widely used in the absence of anything better.

[3] While these standards are not associated with a reference method for a specific greenhouse gas category, they have direct application to QC activities associated with estimations based on measured emission values.

[4] US EPA Methods, e.g., Method 1, 1A, 2, 3, etc., are EPA test methods that are available in Title 40 of the Code of Federal Regulations CFR) Part 60, Appendices. These test methods are developed by the Office of Air Quality Planning and Standards in the Office of Air and Radiation. 40 CFR Part 60 is published each year by the Office of the Federal Register, and is available from the U.S. Government Printing Office. Although the test methods generally do not change from year to year, users should check for the most recent version of 40 CFR Part 60, Appendices.

[5] Methods 2F and 2G correct the measured flow rates for angular (non-axial) flow. Method 2H (for circular stacks) and conditional test method CTM-041 (for rectangular stacks and ducts) are used to correct the measured flow rates for velocity decay near the stack wall, using a ‘wall effects adjustment factor’.

[6] Water measurement is needed to correct measured gas volume to standard ‘dry’ conditions.

2.2.4 Activity data

This Section provides general advice for the production or review of activity data. This includes:

- information on specialised data sources;
- conducting surveys and censuses;
- where appropriate, the use of measurement related data.

It is *good practice* when producing suitable activity data to follow the stepwise approach to set priorities for action according to the importance of the sector, putting in place a strategy for accessing the data needed, collecting the data needed, and processing it to produce the data needed for inventory purposes. This section provides generic advice relevant to the choice of activity data to use.

Volumes 2-5 provide advice on the selection and use of activity data for specific categories.

Data Sources

National and International Literature

As described in Section 2.2.1, it is preferable to use data from such bodies as National Statistical Agencies, and national regulatory authorities responsible for the permitting of industrial and other processes subject to pollution emission legislation.

There will be occasions, however, when other sources of specialised literature provide activity data i.e. UN statistics, US Geological Survey (USGS) reports on commodities, and technical reports, guidelines, standards, sectoral surveys issued by industrial trade associations.

Surveys & Census information

Survey and census information (see Box 2.2) provide the best agricultural, production and energy statistics that can be used for greenhouse gas inventories. Generally these data are compiled by national statistical agencies (NSA) or relevant ministries for national policy purposes or to comply with international demand for data, or other activities that are outside of the direct control of the inventory compiler although the needs of the inventory can sometimes trigger or influence surveys or censuses.

BOX 2.2
THE DIFFERENCE BETWEEN CENSUS AND SURVEY DATA

Survey data are derived from sampling and do not include real data for the whole population. Surveys should assess a representative sample (in the context of the survey 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 derive the national or regional total. Both the representativeness of the sample and the methods used to gross-up need careful 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. It is expensive and time consuming and this is a significant limiting factor 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 who is responsible for all national statistics, while in others the task is split among multiple agencies each of which collect official statistics related to their field, i.e., 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 define adequately the data to be collected and to have at their 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 any NSA) to provide the specialist with statistical knowledge, which is essential to 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 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 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 - for 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 - 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.

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 a NSA or ministry is that they 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. Also, 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. They will have teams of qualified and experienced statisticians, experts in sample selection, questionnaire design, data handling and verification, and the necessary software to process the data. They may also have teams of interviewers experienced in telephone or personal surveying. All of these factors contribute to the success of any survey and equally importantly, to keeping costs down.

General Guidelines for Planning Surveys & Census: It is *good practice* to plan each step with all subsequent steps in mind, from data collection, processing and analysis to dissemination of 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:

- Budget issues: 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.

- Staffing issues, including management of the interviewer workforce: Staff resources 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 collection.

- Project management and timetable issues: Good project management is essential to ensure a smooth-running collection. 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 http://www.amstat.org/sections/srms/brochures/survplan.pdf. The UN's 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 on http://unstats.un.org/unsd/HHsurveys/part1_new.htm. Another useful source is 'Basic Steps in Conducting Surveys', available at http://www.energy.ca.gov/marketinfo/documents/98-10_LANG2.PDF.

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

References to guidance for performing surveys or censuses for energy, industrial processes, agriculture, forestry and waste are given in the Annex 2A.2.

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, National Statistical Agencies (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 to make use of existing data, data gaps still remain, then approach the NSA or ministry about carrying out a new survey. Provided the financial resources are made available, the NSA or ministry will be able to provide the all-important expertise. Also explore whether other partners might be interested in sharing the work and resources needed for it.

2.2.5 Data reporting and requirements

Data reporting requirements pose a challenge to developing countries and require effective data management practices. Standard software tools should be used for data management. For example, US-EPA inventory management tools, UNFCCC software[14] and IPCC Inventory software[15].

2.3 USE OF FACILITY DATA IN INVENTORIES

[New guidance in the *2019 Refinement*].

2.3.1 Introduction

[14] http://unfccc.int/resource/cd_roms/na1/ghg_inventories/.

[15] http://www.ipcc-nggip.iges.or.jp/software/index.html.

Detailed industrial facility data, increasingly collected for various goals such as tracking the progress of emission trading programmes or climate change policies has the potential to be utilized in national inventories. When implemented appropriately, facility-specific data may be used in many possible ways from replacing a subset of data to forming a significant portion of the inventory, including emission estimates, activity data and global emission factors to better reflect country-specific industrial context while increasing the overall quality of the inventory. Facility data may also be used to evaluate, compare or support national inventory methods for a specific greenhouse gas, source, or industrial category.

Data collected directly from facilities through industry research or in the context of different directives and legislation, can also be incorporated into the national inventory. Some options for these data usage are discussed in section 2.3.4 Other Integration Approaches of Facility Data. National Inventories Experience of countries using facility data in National GHG Inventories has been noted in Box 2.3 as an information source for compilers.

It should be underlined that the purpose of data provided by facilities could be different from inventory compilation needs (see section 2.3.4). Data may be collected for air quality monitoring programmes or for regulation and compliance purposes, taking into consideration that additional reporting requirements may need to be specified for these data to be of inventory use.

It is important to note that *integrating facility-reported data (FRD) into inventory is optional and should only be considered if the information improves the quality of the inventory***.** 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. Bias is also present if measurement methods are not provided to account for facility-specific operation and processes. National inventory compilers must assess and assure that the information of interest is of high quality with low bias.

2.3.2 Designing Facility Reporting Programmes for Inventory Use

This section presents approaches for the direct integration of facility data from regulatory greenhouse gas (GHG) reporting programmes into the national inventory. To ensure high quality information is available for direct integration, it is best achieved through appropriate design of the facility-reporting programme (FRP). Inventory compilers and legislative groups developing reporting systems are encouraged to work together in establishing a streamlined and efficient system while limiting burden.

The approach outlined in this section focuses on specifically incorporating the collection of data for inventory purposes into the design of the FRP. However, this approach can also assist compilers to assess facility data, obtained from programmes primarily designed for other purposes, for its use in GHG inventories.

The quality of industrial facility-reported information is dependent on the type and availability of the information to be collected, and the design of quantification methodologies, emission monitoring methods and verification activities. Specifying reporting criteria (such as fuel quantity, carbon content, heat content, etc.) and methods appropriately, and allowing for adequate verification, will assist in achieving the quality and transparency required for the type of inventory integration intended. Elements to consider in the development of industrial facility reporting criteria and methods are discussed in the Quality Criteria and Reporting Elements in Section 2.3.2.1.

Facility-reported data (FRD) could support the development of higher tier IPCC methodologies, however, a higher tier method is not always needed. The added burden of developing a Tier 3 facility-specific approach may not be worthwhile if the same level of quality can be achieved more efficiently or cost-effectively by instead improving the primary activity dataset or updating emission factors, as two examples.

2.3.2.1 FACILITY-SPECIFIC DATA

The quality of FRD is dependent on the type of data collected as well as its accuracy, level of detail and transparency. This section is focused on elements to consider when establishing reporting requirements for attainment of high-quality industrial FRD. It highlights the importance of collaboration between relevant organisations responsible for the compilation of national statistics/activity data and the GHG inventory. To gain efficiency in the collection of data and to reduce industrial reporting burden, collaboration between organisations is needed. This should streamline reporting elements and data sharing when a decision to incorporate specific FRD into the national GHG inventory following data quality review is made.

Quality Criteria and Reporting Elements

Even with independent third party verification of FRD, the use of these data in a national GHG inventory may not be possible due to insufficient reported information or a lack of transparency, both of which can limit the ability to conduct adequate quality assessments. As a starting point in the development of facility-reporting requirements, inventory compilers should consider how to align methodologies to be applied and types of data to be reported with IPCC quality principles:

- methodologies ‐ facility GHG estimation methods and emission factors to be applied must be at least comparable with, or of a higher tier than current inventory approaches;
- activity data alignment ‐ facility fuel consumption/production data should be equivalent to a subset that is found in national datasets for compilers to properly take into account the contribution of facility relative to the total for each type of industry. Collaboration with national activity data compilers is important in ensuring that facility reported data is a subset or is used in the development of the national statistics, allowing for direct integration of reported data with completeness of coverage.

In designing methodologies and reporting requirements for potential integration of facility data (emissions and supporting data), it is *good practice* to achieve quality through the application of inventory principles. That is, design goals should be set to ensure that data from each facility would be complete, transparent, accurate, comparable, and consistent.[16] These quality goals are shown in Table 2.5.

TABLE 2.5
QUALITY GOALS FOR FACILITY DATA

Inventory Principle	Quality Goals
Comparability	• Methods align with inventory and are of equal or higher tier; • Methods across industry types are of equal or higher tier; • Facility activity data aligns with national statistics and other inventory data (e.g.: SIC, ISIC or NAICs categorization); • Facility categories align with inventory (e.g. IPCC category/subcategory).
Consistency	• Facility activity data to be reported across similar sectors/categories and by industry type should have the same reporting units of measure corrected to the same operating condition; • Time series demonstrate consistency and closely align with statistical data across facilities of similar size, type of industry and operating condition else, provisions.
Accuracy	• Activity data is accurate (e.g., fuel measurements are clearly stipulated, possibly based on standardized or regulated measurement and metering standards); • Primary emission factor are accurately determined based on accepted standardized measurement and sampling approaches (e.g. ASTM); • Emission estimates for facilities covered are at least as accurate as inventory methods.
Transparency	• Methodology applied is documented and clearly explained when different from specified approach; • Activity data and emission estimates are documented and clearly explained when different from specified measurement and sampling approaches; • Emission factors are documented and clearly explained when different from specified approach (e.g. plant/source specific emission factors determined transparently); • Reported/archived data allows for full verification.
Completeness	• Facility data (emissions, fuel quantities and emission factors) can be integrated into inventory, ensuring complete sector, category, and/or subcategory coverage; • Industrial coverage – percentage of facilities covered relative to total industry is considered (greater coverage by facility data implies higher resultant inventory quality).

As identified in Table 2.5, these quality goals cannot be pursued in isolation of one and other. For example, if a methodology is comparable with the inventory method, this does not ensure that the facility reported information will improve the quality (or reduce the uncertainty) of inventory estimates. A review of emission factors and

[16] It is also *good practice* to review of reported documentation, apply quality assurance/quality control measures to the provided data and assess the uncertainty of the data. These important inventory activities occur after the reporting program commences, but are also vital to achieving high quality.

activity data may demonstrate that plant specific operating conditions should be taken into account to compile and generate representative facility data. The following are some recommended means of obtaining high quality facility data: direct measurement of fuel, feedstocks and production quantities based on calibrated metering systems; standardized sampling methods of evaluating the carbon content of fuels and feedstock; obtaining information about on-site technologies based on age, type, efficiency or operational load changes; measurement of the leakage rates of plant equipment, including seasonal variations.

A report describing plant specific methods, measurement techniques and assumptions should be provided to allow compilers to conduct quality criteria assessments when facility specified methods or emission factors have been developed and are not consistent with specified methods. General information to be covered in the reports should also be specified.

Establishing Quality through Reporting Specifications

As implied above, in order to obtain high quality information, appropriate methods and reporting requirements must be established. Table 2.6 below is not an exhaustive list, but a starting point to identify reporting requirements in support of obtaining quality facility data. It is based on methodological and data concepts drawn mainly from the Energy and the IPPU volumes of these guidelines, but also from other volumes. Inventory compilers will need to identify their specific methodological and data requirements based on their own inventory quality requirements.

TABLE 2.6
POTENTIAL FACILITY GHG REPORTING REQUIREMENTS

Methodological Specifications	Reporting Elements
• Methods should be, at a minimum, equivalent to inventory methods or better for each source type or activity level. • Measurement or metering approaches should follow regulated standards. • Fuel/product sampling should follow regulated standards to take into account variable carbon content of fuels, process feedstock and product output. • On the basis of inventory information, country-specific default emission factors should be provided for fuels/product with low variation in properties.	• Methods for each source and emission type. • Documentation of approaches when inconsistent with specified methods. • Quantity of each fuel and process input or production output by source type. • Emission Factor by source type or fuel type. • Carbon content for each type of variable fuel and process feedstock. • Fuel properties (e.g., heat content, temperature, pressure, moisture content, etc.).
• Allow use of continuous emission monitoring (CEM) systems and specify calibration and testing requirements.	• Emissions by gases from each CEM system. • Identification of sources captured by each CEM system. • Quantity of each fuel and process feedstock, which generates emissions measured by CEM.
• Source specific methods (i.e., stationary fuel combustion, off-road combustion equipment, coke oven battery, clinker production, wastewater treatment, etc.). • GHG specific methods (i.e. CO_2, CH_4, N_2O, HFCs, PFCs, SF_6 and NF_3).	• Emission by GHG and corresponding source type.
• Specify industrial categories that align with national statistics classification (i.e. Standard Industrial Classification, International Standard Industrial Classification, North American Classification).	• Industrial classification of facility.
• Set a *de minimis* allowance[17]. It can be based on the total emissions of a facility or by source and GHG, and may be specified as a percentage of all GHGs or by type of GHG (e.g. 1% of total CO_2 from fuel combustion). The *de minimis* should in no case be set larger than the absolute value of the uncertainty.	Not applicable.

[17] A quantity of emissions, below which they are not required to be reported.

2.3.2.2 COLLABORATION WITH NATIONAL STATISTICS DATA AGENCIES

For the majority, if not all of developed countries, national statistics datasets represent complete coverage of supply and disposition of materials produced or consumed by economic sectors and are developed based on accepted statistical methods. For these reasons, the majority of national inventories use activity data obtained from these datasets to estimate fuel combustion and industrial process emissions. These national datasets provide complete coverage of a given industrial sector, category or subcategory type. On the other hand, information obtained from FRP in the same subsector will likely not provide complete coverage of emissions for that subsector where aggregated statistics totals can be used to calculate the residual value for an industrial sector. This is a result of the reporting thresholds and *de minimis* allowances used by these programs, in addition to potential differences in sector definitions.

To ensure that emissions are neither over- nor underestimated, collaboration with national statistics data agencies is an essential step in integrating facility reported data into an inventory. Assuming that the information provided by facilities to a GHG reporting program includes activity data (which is fundamental to transparency), it is possible to fill the gaps in facility reported emissions for a given subsector. This can be done by taking advantage of the linkage between activity data reported to the GHG reporting program and that reported to the national data agency. Emissions reported for the portion of the subsector covered by the GHG reporting program can be directly used in the inventory, while emissions for the remainder of the subsector can be estimated using the remaining activity data from the national statistics dataset. Techniques for this are discussed in detail below, but in order for it to be accomplished, some general guidance can be provided.

Aligning facility reported data with national statistics datasets is essential to ensure accuracy, completeness and coverage. Working jointly with the national data organisation can improve issues related to sector allocation and over- or underreporting of consumption or production information organisations, with QA/QC performed by both organisations. This collaborative effort will produce a more consistent dataset, increase data collection efficiency and reduce reporting burden.

To facilitate direct inventory integration of FRD, it is suggested that a working group consisting of experts from the GHG facilities reporting programs, inventory compilers, national statistics data compilers and other relevant ministries/partners work jointly in establishing data reporting elements. Determining reporting elements should start with the identification of data needs by relevant agencies and how best to address gaps and align facility data with both national statistics and emissions reporting. This will ensure that all necessary data is collected while industrial reporting burden. Elements to consider may include:

- industrial classification system such as Standard Industrial Classification [SIC] or North American Industrial Classification System [NAICS];
- data type and units;
- fuel properties – calorific values;
- reporting period;
- statistical sample size/weighting adjustment relative to facilities in the GHG reporting program.

Statistical agencies may be able review the information available from FRP to identify areas of overlap as this is an opportunity to gain efficiencies and reduce data collecting costs. Should cost reduction not be a primary concern, this may be an opportunity to improve national activity data statistics instead. If this is the goal, the agency may choose to: increase the sample size of industries with low coverage; increase survey response rate through improved respondent support; or allow for additional QA/QC and follow up to improve the overall quality and representativeness of the statistical dataset.

It is to be noted that facility reporting deadlines and survey collection timelines should be coordinated between the FRP, inventory institution and the statistics agency. This will allow for the effective scheduling of processes and quality checks, leading to improved compilations of FRD, GHG inventories and national energy and production statistics.

2.3.3 Approaches for Use of Reported Facility Data

The challenge for inventory compilers is assessing how best to integrate facility reported data to achieve improvements, especially if there are some outstanding coverage and completeness issues. Based on the available information and the results of quality evaluations, use of the data may range from simply updating inventory emission factors or activity data to directly integrating the emissions into the inventory. Some examples of integration options, which may be used separately or together (while keeping in mind the need for sectoral/category/industrial completeness), are listed below:

- obtain carbon content of key parameters/activity data to develop improved industry or process-specific emission factors (possibly accounting for technology and process efficiencies);
- improve industrial disaggregation level;
- improve trends analysis and assessment by capturing operation changes including technology or process change, fuel switching, production shut down etc;
- direct incorporation of collected activity data;
- direct incorporation of emissions, usually along with improved emission factors, and resulting in higher quality estimate(s) and method(s).

When an industrial facility provides information to a GHG reporting system, these reported emissions will be in the form of Equation 2.1 (for a given facility in an industry category):

EQUATION 2.1
TOTAL FACILITY EMISSIONS FROM ALL SOURCES

$$E_{Fa} = \left(\sum_{i=1}^{n1} E_{f_i} + \sum_{j=1}^{n2} E_{fug_j} + \sum_{k=1}^{n3} E_{p_k} + \sum_{l=1}^{n4} E_{other_l} + \sum_{m=1}^{n5} E_{other_m} \right)$$

Where:

E_{Fa} = total reported facility emissions from all sources (excluding the *de minimus*, when applicable);

E_{f_i} = sum of all combustion emissions in the facility by type of fuel *i*, with n_1 fuels;

E_{fug_j} = sum of all fugitive emissions in the facility by type of activity *j* with n_2 fugitive sources;

E_{p_k} = sum of all industrial process emissions in the facility by type of process *k*, with n_3 processes;

E_{other_l}, E_{other_m} = sum of all other emission source type *l*, *m…* (i.e. wastewater treatment, product use…), with n_4 and n_5 source types, respectively.

2.3.3.1 COLLABORATION WITH NATIONAL STATISTICS DATA AGENCIES

A rigorous bottom up approach involves using a complete set of facility data for each type of industrial category and takes into account operating conditions of each facility. If all the quality goals are met and the statistical data is aligned with the FDR and if coverage is complete, the compiler can directly integrate the reported facility emission estimates for the industrial category. The facility estimates are summed by combustion, fugitive, appropriate industrial process and other (such as wastewater treatments and product use) source categories, and incorporated into the applicable IPCC categories and subcategories. The reported emission data then supplants the former estimates for this industrial category.

Full coverage, however, is difficult to achieve. It is even less likely that completeness will be assured for a category (e.g. 1.A.2. Manufacturing Industries and Construction), rather than at the subcategory level/industrial category (1.A.2.a. Iron and Steel). Furthermore, coverage of FRD for a specific industrial category (e.g. iron & steel, aluminium, cement, etc.) is only fully complete if there are no significant emissions less than the allowable *de minimis* level. This may well not be the case given various reporting specification and allowances such as set thresholds.

As stated earlier, a key way to ensure completeness under these circumstances is by taking advantage of the linkage between activity data reported to the FRP and, by survey, to the national statistics agency. This linkage exists because many of the facilities in a subsector would be required to report to both statistical survey and the GHG reporting program. By identifying these facilities and the industrial categories in which they are categorized (e.g. by NAICS, SIC or ISIC code), they can be related.

For industrial categories, where reported information provides insufficient coverage, national statistics activity data can be combined with facility reported data to obtain more accurate inventory estimates. The industrial categories can be matched using the appropriate industrial classification, while improved emission factors for that industry can be obtained from the each FRD. Establishing data sharing agreements between statistics agencies, the FRP, and GHG inventory organisations will allow access to facility-level information, by industrial classification, increasing transparency and possible use. To meet confidentiality agreements, data can be presented by total, aggregated industrial categories.

Equation 2.2 is applicable when integrating industry or facility-specific emission factors developed based on reported data. For example, this equation can be used for combustion-based emissions, fugitive emissions or process specific emissions such as coke oven gas batteries or blast furnaces at integrated iron and steel plants. Documentation of methods, assumptions and approaches to develop facility specific emission factors is necessary to meet quality goals.

Equation 2.2 makes use of national activity data for an industrial category (following a standard classification system) to calculate emissions by source or fuel, based on relevant facility-specific information:

EQUATION 2.2
EMISSIONS CALCULATED BY FACILITY-SPECIFIC EMISSION FACTORS

$$E_{IC_s} = \sum_{s=1} \left(AD_{NS_s} \right) \bullet EF_s$$

Where:

IC = industrial classification by IPCC and corresponding industrial classification system such as SIC, ISIC or NAICS;

E_{IC_s} = emissions total for a specific industry classification and IPCC category for source s (e.g. fuel consumption quantity by type of fuel, each type of process input, each type of production output quantity);

AD_{NS_s} = activity data applicable to source s (e.g. fuel consumption quantity by type of fuel) from national statistics agency by industrial classification;

EF_s = source s specific emission factor.

For reference or comparative purposes, the grand total of emissions for a specific industry classification can be calculated by summing all emission sources (e.g. from fuel combustion, fugitive, industrial processes, wastewater, and product use).

Integration approaches will differ across all inventories due to differences in national circumstances and available data. The Facility Data Integration equation (Equation 2.3) takes into account completeness issues and provides flexibility to compilers in adapting the equation to meet their integration objectives and inventory methods. Compilers will also need to reassess integration approaches as reporting requirements are refined and reported information improves in quality over time.

In this equation, estimates computed by national activity data and facility specific emission factors are combined with direct integration of facility emission estimates, allowing for the full use of reported data. To fully utilize this approach, the importance of tracking facility data by standard industrial classification systems (NAICS, SIC or ISIC) is emphasized.

For any category (e.g. iron and steel by fuel combustion, blast furnaces and coke oven battery), total emissions for an industrial classification can in general be determined as a combination of calculated and facility reported estimates. This Facility Data Integration equation accounts for the residual emission value (from non-reporting facilities or missing emission sources) with the use of a top down and bottom up approach to integrate both statistical and facility data and to directly incorporate reported emissions.

EQUATION 2.3
FACILITY DATA INTEGRATION BY EMISSION SOURCE

$$E_{IC_s} = \left(AD_{NS_s} - \sum_{Fac_{s=1}}^{n} AD_{Fac_s} \right) \bullet EF_s + \sum_{Fa_{s=1}}^{m} E_{Fa_s}$$

Where:

IC = industrial classification by IPCC and corresponding industrial classification system such as SIC, ISIC or NAICS;

E_{IC_s} = emissions total for a specific industry classification and IPCC category for source *s* (e.g. fuel consumption quantity by type of fuel, each type of process input, each type of production output quantity);

EF_s = fuel-, source-, or process specific emission factor by source *s*;

AD_{NS_s} = activity data from national statistics organisation applicable for source *s* by standard industrial classification systems/IPCC categories;

AD_{Fac_s} = facility reported activity data applicable to source *s* for a specific standard industrial classification category, with *n* facilities reporting for a specific industrial classification;

E_{Fa_s} = reported emissions applicable for each source *s* from each reporting facility, with *m* facilities reporting for a specific industry classification.

The grand total of emissions for a specific industry classification (NAICS, SIC or ISIC) can be calculated by summing the results of Equation 2.3 for all sources and activities as shown in Equation 2.4.

EQUATION 2.4
TOTAL FACILITY EMISSIONS BY INDUSTRIAL CLASSIFICATION

$$E_{IC} = \sum_{s=1} E_{IC_s}$$

Where:

E_{IC} = emissions total for a specific industry classification by IPCC and corresponding industrial classification system such as SIC, ISIC or NAICS;

E_{IC_s} = emissions total for a specific industry classification and IPCC category for source *s* (e.g. fuel combustion, fugitive, industrial processes…).

Note that if no facility data is directly incorporated, Equation 2.3 reduces to 2.2. Furthermore, if only the facility-reported emissions are to be considered, Equation 2.3 reduces to a summary of Equation 2.1.

By accessing detailed national activity data it is possible, in collaboration with the national statistical agency, to make use of the facility GHG emission reporting program data in the most comprehensive and robust manner. Assuming that the activity data has been reported to the FRP (and this should be the case for transparency and verification purposes), by identifying what portion of the national activity data has been utilized by facilities to estimate their emissions, it is possible to calculate the remaining emissions subtracting that portion of the data from the national activity data. As national data agencies generally determine industrial allocations by ISIC, SIC or NAICS codes, it is strongly recommended that emissions reporting facilities be tracked by these same codes. In this way, the emissions can be combined, ensuring that there is no overlap or gaps and with all relevant facility reported emissions appearing in the final inventory results.

If appropriate, update inventory emission factors based on industry or facility specific operating conditions or fuel properties should be applied when FRD meets all quality requirements.

Depending on the approach to integrate facility data, the amount of information available over time and the level of transparency and quality, there may or may not be a break in the time series consistency. When such break occurs and it may be justifiable, such as when a new technology or an alternative energy or fuel has been implemented. In these cases, explanatory documentation should be required. It is *good practice* to document methods and provide an explanation for the use of facility reported data, including a discussion on time series consistency, in the inventory. Refer to Volume 1, Chapter 5 of this *2019 Refinement*.

2.3.4 Uses of Facility Data not Originally Designed for Inventory Application

As already noted, data from facilities could be collected primarily for other purposes than for inventory compilation. Nonetheless, these data could be used in the inventory estimation process provided an accurate examination of the purpose and information collected is made.

As for basic data and parameters published in official statistics and documentation, the quality of the data collected at facility level should be assessed. This means that the inventory compiler needs to know:

- the calculation and measurement methods used by the facility;
- if those methods are in line with the IPCC methodologies to estimate emissions from the specific category;
- which QA/QC and verification activities have been undertaken and which level of uncertainty is associated to the data collected;
- if there are confidential data an explanation of the reason;
- the representativeness of the data over the category to be estimated and at national level (e.g., for combustion emissions: How data compare with national energy statistics? For production activities, in general: Is the sample size adequate compared with the number of national facilities in the IPCC category?).

In fact, these data do not always cover inventory categories entirely whereas national statistics, at least for activity data, may provide the complete basic data for inventory compilation (e.g. energy categories, such as the national energy balance and the energy production and consumption statistics). Therefore, the first step is to check for the coverage of data reported.

When the sectoral coverage of the category to be estimated is known to be complete, data collected from facilities in the context of different directives and legislation, e.g. Emissions Trading Schemes, can be incorporated into the national inventory (IPCC, 2011 [a], [b]).

Even in case of partial coverage, data may be used to derive country-specific emission factors, sum up to other activity data and as a verification for emission factors and activity data levels.

There are other situations where industries communicate figures only if they exceed specific thresholds. Furthermore, basic data such as fuel consumptions are not supplied and production data are not always split by product but reported as an overall figure. Notwithstanding, the inventory compiler may use these data for verification checks and take the opportunity to have contacts with industries which may supply additional information as necessary to carry out sectoral emission estimates.

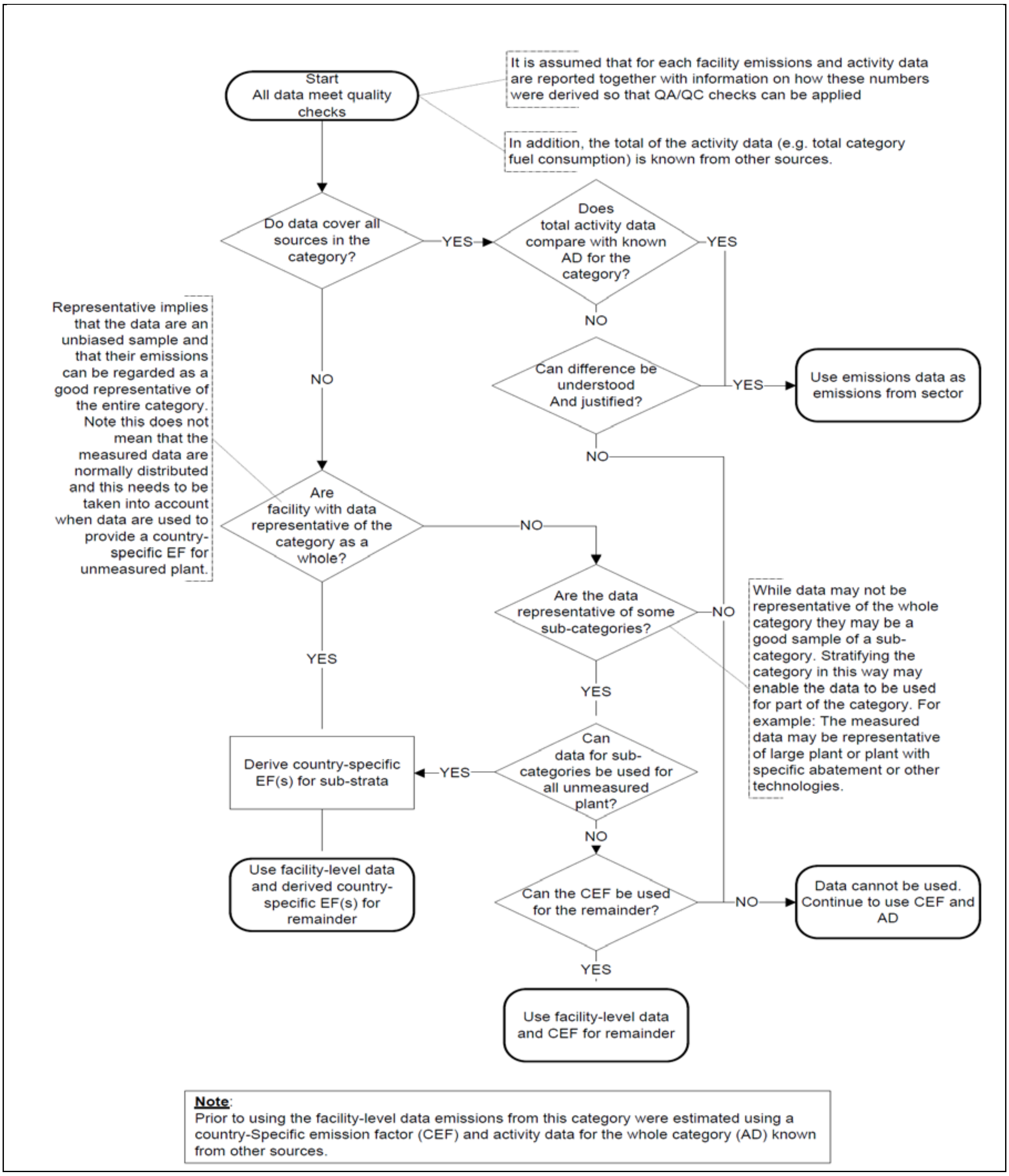
When a category is not completely covered by facility-level data, an important decision is whether to use the data integrating them in the emission inventory, and how to completely estimate the emissions from the category.

In fact, the inventory compilers could:

- not use the facility-level data set in the estimation of emissions but the data could still be used for QA/QC and verification activities;
- devise a method to combine facility-level data with other data and information available, where the overall accuracy of the sectoral emissions would be increased by integrating the facility-level data into the inventory;
- use a proxy measure to derive the data necessary to be used for inventory purposes (e.g. facility-level energy consumption to split the total emissions into combustion and processes emissions, if the disaggregation is not available);
- assume the facility-level data are representative of the whole population so the average emission rate of the facility-level data applies to any missing data.

It is *good practice* to follow decision tree provided in Figure 2.4 on how to use data collected from facilities (IPCC, 2011).

Figure 2.4 Example of Decision Tree for Selecting Facility-Level Data



Source: IPCC, 2011

Another aspect to consider is that, in general, facility-level data are not available for all inventory years and that methods and parameters that facilities use to monitor or estimate their emissions are likely to improve over time.

Given this situation, it is important that inventory compilers integrate facility-specific estimates in the national emission inventories with particular attention to ensure the consistency of the time series and splice facility-level data with traditional category level estimates (see chapter 5 of Volume 1, Box 5.2).

On one hand, it is important to consider the temporal period data are collected. If data are collected annually, or if the collection occurs every 5-8 years, compilers should consider the benefits to use the data as representative of a specific length of the period (e.g. for landfills, the collection of data on waste composition may occur every 5-8 years and the composition of waste should be used as representative of the other years). Another issue to consider when recalculating data for earlier years is the changes in technology and practices that may have occurred.

When the details of collected information improve during the time, a way to ensure time series consistency is that inventory compilers estimate emissions using both the old and new methods (less and more advanced tiers) for one or more years and use splicing techniques to achieve time series consistency. In some cases, the reporting facility can be requested to provide estimates using both methods for one or more years or check the results estimated by the inventory compiler. For further examples, please refer to Volume 1, Chapter 5 of this *2019 Refinement*.

Whatever the use of facility data, it is important to provide sufficient data and documentation such that any external reviewer or user of the data can understand how the data have been collected and make a determination whether the methods are consistent with IPCC Guidelines. The information can be incorporated in the current inventory reports, as part of a description of the institutional arrangements, or of source category specific discussions. In addition, information on regulations should be referenced, even not fully reproduced in the inventory report.

Some guidance in this regard may be found in the Report of the IPCC Expert Meeting on Use Models and Measurements in GHG Inventories (IPCC, 2011).

In summary, it is *good practice* for an inventory compiler to report:

- which facility-level data have been used (AD, EF or emissions);
- what were the specific methods followed;
- cases where emission estimates were used directly;
- what is the source specific coverage by facility-level data (if not complete, how have emissions been estimated from non-reporting facilities);
- cases where data set used was determined to be more appropriate than the facility-level data reported under regulations;
- the level of uncertainty of facility data;
- QA/QC activities (e.g., comparison of implied EF with IPCC default EF);
- treatment of time-series consistency where facility-level data only available for part of the time-series or methods changed over the time-series;
- cases where emissions are aggregated to protect confidential data, identify what data could be published (i.e. EFs for individual categories).

In addition, it is *good practice* for Inventory compilers to explain any differences between national inventory data and data collected at facility level. For instance, a relevant article of the European Union Regulation for data consistency, on emissions reported under the European Emissions Trading Scheme, is Article 10. A detailed table with a comparison between reported GHG emissions under UNFCCC and data from the European emissions trading system by category is included in the National Inventory reports (NIRs) of many European countries.

BOX 2.3
FACILITY DATA CONSIDERATION AND USE

Additional resources for research purposes are available to inventory compilers when considering use of facility data. These resources can help guide compilers in gaining knowledge and assessing how best to use their data. These references can range from reporting design consideration, QA/QC checks, implementation and direct usage of facility data.

It is *good practice* to use the following references as starting point:

IPCC (2011). Use of Models and Facility-Level Data in Greenhouse Gas Inventories (Report of IPCC Expert Meeting on Use of Models and Measurements in Greenhouse Gas Inventories 9-11 August 2010, Sydney, Australia). eds: Eggleston H.S., Srivastava N., Tanabe K., Basssansuren J., Fukuda M., Pub. IGES, Japan 2011.

WRI (2014) Working Paper - Exploring Linkages Between National and Corporate/Facility Greenhouse Gas Inventories, World Resources Institute. N. Singh, T. Damassa, S. Alarcon-Diaz, M. Soto. March 2014.

Compilers may also learn from experiences of national inventory programmes such as the US EPA, Australia and European members on how they have integrated facility data into their national GHG inventory along with information on facility reporting requirements and systems.

References

Cullen A.C. and Frey H.C. (1999). *The Use of Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum: New York, 335 pages.

Frey H.C. and Burmaster D.E. (1999). 'Methods for characterizing variability and uncertainty: comparison of bootstrap simulation and likelihood-based approaches,' *Risk Analysis*, 19(1):109-130, February 1999.

Frey H.C. and Rhodes D.S. (1996). 'Characterizing, simulating, and analyzing variability and uncertainty: an illustration of methods using an air toxics emissions example', *Human and Ecological Risk Assessment*: an International Journal, 2(4):762-797, December 1996.

EEA (2016) EMEP/EEA air pollutant emission inventory guidebook 2016. EEA Report No 21/2016.

Fadnavis, S., Ravi Kumar, K., Tiwari, Y. K. & Pozzoli, L. (2016) Atmospheric CO_2 source and sink patterns over the Indian region. *Annales Geophysicae* **34**(2): 279-291.

Ganesan, A. L., Rigby, M., Lunt, M. F., Parker, R. J., Boesch, H., Goulding, N., Umezawa, T., Zahn, A., Chatterjee, A., Prinn, R. G., Tiwari, Y. K., van der Schoot, M. & Krummel, P. B. (2017) Atmospheric observations show accurate reporting and little growth in India's methane emissions. *Nature Communication* **8**(1): 836. DOI:10.1038/s41467-017-00994-7.

IPCC (2011) [a]. Use of Facility-Specific Data in national Greenhouse Gas Inventories. Task Force on National Greenouse Gas Inventories (TFI). TFI Technical Bulletin 1. Hayama Japan. Accesible at: http://www.ipcc-nggip.iges.or.jp/public/tb/TFI_Technical_Bulletin_1.pdf.

IPCC (2011) [b]. Use of Models and Facility-Level Data in Greenhouse Gas Inventories (Report of IPCC Expert Meeting on Use of Models and Measurements in Greenhouse Gas Inventories 9-11 August 2010, Sydney, Australia). eds: Eggleston H.S., Srivastava N., Tanabe K., Bassansuren J., Fukuda M., Pub. IGES, Japan 2011.

Ogle, S. M., Buendia, L., Butterbach-Bahl, K., Breidt, F. J., Hartman, M., Yagi, K., Nayamuth, R., Spencer, S., Wirth, T. & Smith, P. (2013) Advancing national greenhouse gas inventories for agriculture in developing countries: improving activity data, emission factors and software technology. *Environmental Research Letters* **8**(1): 8.

Tiwari, Y. K., Ramesh K. Vellore, K. Ravi Kumar, Marcel van der Schoot, Chun-Ho Cho (2014) Influence of monsoons on atmospheric CO_2 spatial variability and ground-based monitoring over India, *Science of The Total Environment* **490**: Pages 570–578, ISSN: 0048-9697. http://dx.doi.org/10.1016/j.scitotenv.2014.05.045.

Tiwari, Y. K., Revadekar, J. V. & Ravi Kumar, K. (2013) Variations in atmospheric Carbon Dioxide and its association with rainfall and vegetation over India. *Atmospheric Environment* **68**: 45-51.

Tiwari, Y. K., Valsala, V., Vellore, R. K. & Kunchala, R. K. (2013) Effectiveness of surface monitoring stations in representing regional CO_2 emissions over India. *Climate Research* **56**(2): 121-129.

Yogesh K Tiwari, Prabir K. Patra, Frédéric Chevallier, Roger J Francey, *et al.* (2011) CO_2 observations at Cape Rama, India for the period of 1993-2002: implications for constraining Indian emissions. *Current Science* **101**(12):7.

Annex 2A.1 A protocol for expert elicitation

No refinement.

Annex 2A.2 General guidance on performing surveys

[Elaboration of Annex 2A.2 of the *2006 IPCC Guidelines*].

Survey data are often compiled using financial/fiscal incentives for reporting. This may introduce possible bias if the incentives favour a certain bias in reporting. For example, 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.

The collection of information for an inventory use is the most critical and often resource demanding aspect of the process since most inventories rely on data collected for other reasons, especially socio economic reasons, or such as fuel consumption data that are collected for tax reasons.

To minimize the cost of new data collection campaigns, collaboration with existing agencies and programs that have an interest in collecting related data from the same subjects should be put in place.

It is clear that the establishment of collection procedures should, over time, improve the quality of the inventory available data sets and consequently emission estimates. The collection of data or parameters should be carried out at a level of detail appropriate to the methods used for each category and the focus starting from the most relevant emission categories "*key categories*".

As a first step, an inventory compiler should review what data are likely to be available through existing national statistical data system or administrative sources of data should be explored.

Once the collection program is in place, activities should be revised on a regular basis as part of an ongoing improvement of data and inventory quality and the possibility to incorporate new questions or module in the questionnaires of existing questionnaires should be explored. In addition, agreements with data suppliers should guarantee a consistent and continuing information flow

Where there is no information available at national level, international bodies, such as the United Nations and International Energy Agency, should be considered. In addition, sectoral groups such as international trade and industrial associations may have country-specific data sets for industries or other economic sectors.

International organizations publishing statistics include:

- the United Nations Statistics Division (http://unstats.un.org/unsd/databases.htm);
- Eurostat (http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes);
- International Energy Agency (http://www.iea.org/stats/index.asp);
- Organization for Economic Cooperation and Development (http://www.oecd.org/statistics/);
- International Monetary Fund (http://www.imf.org/external/data.htm);
- FAO (http://faostat.fao.org/).

Problems with international statistics and their use in the inventory compilation are mainly related to the level of disaggregation reported of parameter/data, their definitions, the use of units of measures and conversion factors used. Therefore, the inventory compiler, when considering international sources, should pay attention to all these factors.

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 international level. The United Nations publication Energy Statistics: A Manual for Developing Countries (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 https://unstats.un.org/unsd/publication/SeriesF/SeriesF_56E.pdf This publication should be used in conjunction with two other UN publications:

- An updated document on methods is the 'International Recommendations for Energy Statistics (IRES), United Nations (2016)'[18]. Concepts and Methods in Energy Statistics, with Special Reference to Energy Accounts and Balances (1982) which 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 (1987), which 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.

Also IEA has published an Energy Statistics Manual providing useful background information for collecting, reporting and understanding energy statistics[19].

As for energy statistics, there is also a database by the IEA documenting the worldwide energy statistics (2017 edition)[20].

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. Aggregated data on annual production, trade and apparent consumption of primary energy products are also published in the UNSD Statistical Yearbook. UNSD is involved with the Asia Pacific Energy Research Center (APEC), EUROSTAT, the International Energy Agency (IEA/OECD), 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)[21].

Enerdata and Eurostat also provide additional data sets on energy and other statistics[22].

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[23].

Usually not all the information is available from international sources and international questionnaires should be analyzed in order to understand where are the main issues to be solved to improve emission estimates. A discussion among national inventory experts, Ministries of Energy and international organizations and national statistical offices should be engaged.

[18] https://unstats.un.org/unsd/energy/ires/IRES_Whitecover.pdf.

[19] http://www.iea.org/Textbase/publications/free_new_Desc.asp?PUBS_ID=1461; https://www.iea.org/publications/freepublications/publication/energy-statistics-manual.html.

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

[21] http://www.jodidata.org/.

[22] http://ec.europa.eu/eurostat/web/energy/data/database; http://ec.europa.eu/eurostat/web/energy/data/main-tables; http://ec.europa.eu/eurostat/statistics-explained/index.php/Category:Energy.

[23] www.irena.org/.../Publications/IRENA_RE_Capacity_Statistics_2017.pdf.PDF file.

Additional explanations are usually necessary on the type of fuels to understand what specific figures they refer to. In case 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 of potential country-specific caloric values and the underlying calorific values used are not always transparent.

In addition, timeliness of the reporting of data should be considered. International organizations may use preliminary data for reporting of recent years, and the updating of historic data that usually are not always 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 the IPCC guidelines.

Key publications on International Recommendations for Energy Statistics (IRES), United Nations (2016), Studies in Methods - Series M No.93. Sales no. E.14.XVII.11.

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. The revised list will be based on the Central Product Classification (CPC) and will be fully compatible with the International Standard Industrial Classification (ISIC), the European Union's PRODCOM commodities list and the Harmonised Nomenclature System (HS) used for foreign trade statistics. The revised industrial commodities list and guidelines for countries will be 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/unsd/cr/registry/regct.asp?Lg=1. Detailed chemical industry data (production per country of many products and process data) can acquired from SRI Consulting (www.sriconsulting.com): Process data on aluminium production can be obtained from Aluminium Verlag (www.aluverlag.de); Steel process information can be acquired from IISI (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 field they will have an insider's knowledge of the most common processes used, and may even be willing to survey their members at regular intervals to assess penetration of new processes. In the 1990's Eurostat produced the NOSE-p list – a Nomenclature of Sources of Emissions that links processes to industries. This needs to be revised, but remains a useful starting point for countries starting work in this area.

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 from 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 import or exports are thought to be significant but not quantifiable. However, since there is a possibility of incompleteness or overestimation due to underreporting 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 represents gross or net production (i.e., with or without internal recycling). For some categories these figures could differ by 5 to 10 percent 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[24] and the UN[25].

In addition, the WTO 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 (NTMs) and trade in value added.

AGRICULTURAL SURVEYS AND CENSUSES

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

[25] https://comtrade.un.org/ SITC Rev.3 three-digit level.

Since its establishment, the Food and Agriculture Organization of the United Nations, (FAO) has promoted national censuses of agriculture through its Programme for the World Census of Agriculture; see http://www.fao.org/es/ess/census/default.asp; the Programme is prepared by the Statistics Division of 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/wca2020/en/.

The Programme is complemented by practical information on the steps involved in actually conducting an agricultural census.

Other guidance from FAO on conducting agricultural surveys includes:

- '*Sampling Methods for Agricultural Surveys*', FAO Statistical Development Series No. 3 (1989); which 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[26].

- '*Collecting Data on Livestock*', FAO Statistical Development Series No. 4 (1992); which 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 cost of production and feed/fodder27.

- '*Multiple Frame Agricultural Surveys: Volume 1&2*', FAO Statistical Development Series No. 7 and 10 (1996&1998). 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 which 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 to 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 with an indication of their respective 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.[28]

Volume 2 presents the area and multiple frame survey methods for Agricultural Survey Programmes currently used in a wide range of countries. It provides actual examples of the application of the survey methods presented in the first volume[29].

FOREST SURVEYS

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

[26] http://www.fao.org/fileadmin/templates/ess/ess_test_folder/Publications/SDS/3_sampling_method_for_agricultural_survey.pdf.

[27] http://www.fao.org/fileadmin/user_upload/esa/docs/1_Collection%20data%20on%20livestock.pdf.

[28] http://www.fao.org/docrep/009/a0135e/A0135E00.htm.

[29] http://www.fao.org/fileadmin/templates/ess/ess_test_folder/Publications/SDS/10_multiple_frame_agricultural_surveys.pdf.

The FAO has also produced the following relevant publications:

- Knowledge Reference for forest resource assessments[30] and Voluntary Guidelines on National Forest Monitoring[31];
- Manual of forest inventory FAO Forestry Paper 27 (FAO, 1981)[32];
- Forest volume estimation and yield prediction. FAO Forestry Paper 22/1 and 22/2 (FAO, 1980)[33];
- Methods & Guidance Documentation of the Global Forest Observation Initiative (that contains methodological advice for countries wishing to make use of remotely sensed and ground-based data for forest monitoring and reporting)[34].

Further FAO has developed a set of tools that support data collection for forestry as:

- Collect, survey design and data management tool (http://www.openforis.org/tools/collect.html);
- Collect Mobile for data collection and validation in the field (http://www.openforis.org/tools/collect-mobile.html);
- Calc, a tool for data analysis and results calculation (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 to industry should result in reliable data on waste generated and its composition. 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. Much 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 European Environment Agency's report 'Development and application of waste factors: an overview' see https://www.eea.europa.eu/publications/technical_report_37 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, nor 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 and rural, wealthy and 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 monitor trends. Because the panel will need to be very actively involved in weighing and analyzing the contents of their 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.

In other cases, e.g. for landfills, the amount of waste disposed of in controlled landfills should be monitored together with the amount of population or area served. Therefore, the amount of waste generated per inhabitant

[30] http://www.fao.org/3/a-i4822e.pdf.

[31] http://www.fao.org/3/a-I6767e.pdf.

[32] www.fao.org/docrep/016/ap358e/ap358e00.pdf.

[33] http://www.fao.org/docrep/016/ap353e/ap353e00.pdf & http://www.fao.org/docrep/016/ap354e/ap354e00.pdfhttp://www.fao.org/icatalog/inter-e.htm.

[34] http://www.gfoi.org/methods-guidance/.

could be approximately derived. In addition, some information on the amount of waste disposed of in uncontrolled landfills could be available. For incinerators, which capacity is usually known, the same method could be applied.

International waste statistics may be found in Eurostat's website[35] where data are provided by the national statistical authorities.

Other statistics on international waste are generated by different sources:

- European Environmental Agency[36];
- Organization for Economic Co-operation and Development (OECD)[37];
- the United Nations[38];
- the World Bank[39].

[35] http://www.municipalwasteeurope.eu/waste-statistics.

[36] https://www.eea.europa.eu/data-and-maps/data/external/generation-of-waste-statistics.

[37] https://data.oecd.org/waste/municipal-waste.htm.

[38] http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27.

[39] https://openknowledge.worldbank.org/handle/10986/17388.