### Criteria for assigning points in the quality assessment framework

| **Category** | **No.** | **Subcategory** | **Subcategory explained** | **1 point** | **2 points** | **3 points** |
| --- | --- | --- | --- | --- | --- | --- |
| **Soundness of methods** | 1 | Relevance of scope | Content validity – What is measured? Which IFFs does it cover? | Single IFFs activity is covered, not clearly delineated from others.  | Single IFFs activity is covered, clearly delineated from others.  | Several IFFs activities are covered, all clearly delineated from each other and the ones potentially left outside the scope. |
|  | 2 | Clarity of concepts | Construct validity – Does it measure what it is supposed to? Is it clearly defined? Is a classification used? Is it discrete, exhaustive, and mutually exclusive (are there gaps or overlaps)? | Concept defined only partially, significant overlaps and gaps exist.  | Clearly defined concept, yet either exhaustiveness or mutual exclusiveness are not guaranteed. | Clearly defined concept, exhaustive and mutually exclusive of other (IFFs) concepts. |
|  | 3 | Robustness | How stable are the results produced by the method? Will a repetition lead to similar results? What if conditions change? | Methodology is based on several assumptions, requiring constant verification of the method to produce results.  | Several assumptions are required for the methodology to work, yet these are stable in space and time.  | Only a limited number of stable assumptions are required to produce statistics.  |
|  | 4 | Transferability | How easy it is for someone else to use the method? Availability of empirical research or application of the method | Clear methodology and concept are set, yet poorly documented and without empirical research available. | Clear methodology and concept are well documented, yet without or only little empirical research to test the methodology.  | Clear methodology and concept are well documented with abundant empirical testing available in literature.  |
|  | 5 | Equivalence | Does the method yield similar results when compared to other (sound) methods? | Both levels and dynamics estimated by this method do not reasonably match with other methods' results.  | Results differ from other methods with respect to the level of measured IFFs, whereas overall dynamics is in line with the other estimates. | Results are in line with other methods applied to the same IFFs measurement.  |
|  | 6 | Statistical alignment | Is the method similar to those applied in official statistics? Are the concepts and classifications aligned with official? | Only limited amount of method's concepts is related to the ones used in official statistics. | About half of concepts used in the methodology are aligned with those used in official statistics.  | Method's concepts are fully compatible with official statistics' frameworks.  |
|  | 7 | Capacity requirements  | How much resources and capacity are required for using the method? | Data are poorly available, or available but without proper methodological support and institutional collaboration required.  | Some data are already available, methodologically developed to a certain degree.  | Data, methodology and required collaboration are already in place in existing national statistical system. |
| **Source data**  | 8 | Timeliness | What is the delay of data becoming available after reference period? | Data are available with a time lag of more than one year. | Data are available with a time lag of a quarter to a year.  | Data are available immediately or with a time lag of less than a quarter of a year.  |
|  | 9 | Availability | How easily available are these data to statistical authorities? Are these available in many/most countries? | Data are readily available only in a few countries (issues with sensitivity, privacy, confidentiality, unwillingness to cooperate). | Data are readily available in some countries, but not in others, e.g., developing countries. | Data are readily available with little to no barriers to access. |
|  | 10 | Fit for purpose | Do these data provide information on IFFs, directly or indirectly? Which IFFs do they address? | Data are provided indirectly for only some IFFs activities.  | Data are provided only indirectly for most of IFFs activities.  | Data on IFFs are provided directly, covering all IFFs activities. |
|  | 11 | Coverage | Do the data cover the issues to be measured? Which IFFs are covered? What are the gaps and overlaps? | Data provide only limited IFFs coverage in terms of activities, breakdown, or actors.  | Data cover most of the IFFs and provide required breakdown, yet gaps and overlaps exist.  | Complete and exhaustive coverage of the IFFs activities and actors (individuals and entities).  |
|  | 12 | Granularity | How detailed are the data? Frequency. | Data are aggregated at country-level and annually.  | Data are aggregated at middle-level, e.g., product-level, available on a monthly or less frequent periodicity.  | Data are transaction-level microdata.  |
|  | 13 | Interoperability | Can the data be integrated with other data? Does the dataset include identifiers and classifiers? | Only limited integration of data with other data sources is readily available; significant resources would be needed for a full integration.  | Data are partly integrated with other data sources, or some resources are needed to bring integration at practical level.  | Data are fully integrated with other sources within national statistical system, using a full scale of identifiers.  |
| **Results** | 14 | Relevance for use | Are the results helpful for assessing IFFs or curbing different types of IFFs? How many uses are there for these results? | Results have limited relevance for policy formulation, addressing only one IFFs activity, with present overlaps or gaps, applicable only to a small circle of countries. | Results consider clearly defined IFFs activities, confounding effects may be present; applicability is limited to several (groups of) countries, but not universally.  | Results are directly applicable for policies, directed at specific and clearly delineated IFFs activities; relevance for countries is almost universal.  |
|   | 15 | Accuracy | Do the results describe what is intended? Are there large revisions? | Limited comprehensiveness of concepts and methodology, assumptions sensitive to changes.  | Limitations in comprehensiveness of defined concepts are adjusted by several assumptions.  | Clearly defined concepts, advanced statistical techniques used, limited reliance on assumption all produce results accurately representing the measured reality.  |
|   | 16 | Timeliness | How quickly will the results be available? Are they available on time to help solve problems? | Adjustments to data and/or methodology are significant enough to impair simple replication in majority of countries.  | Some data are harder to obtain, or some methodological adjustments are needed for each replication of the calculations. | Given the data, replication of the method is quick and straightforward in all, or most countries.  |
|   | 17 | Clarity | How easy are the results to use and interpret? | Significant additional effort is required by statistics compilers in most countries for users to properly use the results.  | Results are mostly simple to interpret and connect to practical concerns of IFFs, but mostly only in certain groups of countries (e.g., developed).  | Results are straightforward to interpret with respect to IFFs typologies (activities, breakdown) for most part and in most countries.  |
|   | 18 | Comparability | How comparable are the results in different conditions, across time and countries? | Significant limitations in comparability of results between various countries exist. | Limitations in methodology and/or data requires caution in direct comparison of results across time and countries of different groups (e.g., developed vs developing).  | Data and methodology are comprehensive enough to allow for straightforward comparability of results across time and countries and variations in conditions.  |
|   | 19 | Coherence | How coherent are the results internally? Can they be used together with other IFF estimates? | Using IFF estimates in combination with other estimates (categories or activities) is limited to only a narrow set of countries.  | Broad distinction and consistency between activities and categories of IFFs allows for safe comparison of results in most countries.  | Results refer to clearly delineated IFFs activities and are directly used with other IFFs categories in most countries.  |

##

**Scores of the evaluation for the suggested methods**

\* The assessment for source data and results is indicative only as it will depend on national data. Similarly, the assessment of soundness of methods may vary depending on the context of implementation.



