

# e-Commerce Statistics Workshop

ULAANBAATAR, 20-23 MARCH 2023

EMILIO BUGLI INNOCENTI

# E-Commerce Statistics Workshop Agenda

## DAY 3 – Country state of play and International Recommendations

- ❑ *Official e-Commerce Statistics – state of play in Mongolia*
- ❑ *International Recommendations on Data Collection, Processing and Disseminations*
- ❑ *Discuss Specific e-Commerce Statistics Cases*

# International Recommendations

## **Broad** definition of e-commerce transactions (OECD):

- *An electronic transaction is the sale or purchase of goods or services, whether between businesses, households, individuals, Governments, and other public or private organizations, conducted over computer-mediated networks. The goods and services are ordered over those networks, but the payment and the ultimate delivery of the good or service may be conducted on or offline.*
- As a guideline for the interpretation of the definition above, the OECD notes that the broad definition includes orders received or placed on any online application used in automated transactions, such as Internet applications, electronic data interchange (EDI) or interactive telephone systems.

# International Recommendations

## **Narrow** definition of e-commerce transactions (OECD):

- *An Internet transaction is the sale or purchase of goods or services, whether between businesses, households, individuals, Governments, and other public or private organizations, conducted over the Internet. The goods and services are ordered over the Internet, but the payment and the ultimate delivery of the good or service may be conducted on or offline.*
- As a guideline for the interpretation of the definition above, the OECD notes that the narrow definition includes orders received or placed on any Internet application used in automated transactions, such as web pages, extranets and other applications that run over the Internet such as EDI over the Internet or over any other web-enabled application regardless on how the Web is accessed (e.g., through a mobile phone or a TV set, etc.). The definition excludes orders received or placed by telephone, facsimile or conventional email.



# International Recommendations

- It should be noted that official statistics surveys are routinely carried out in most developing economies but that total resources for statistical production are often scarce.
- Even so, it is unlikely that statistics on business ICT use can be produced efficiently outside the national statistical system, especially to guarantee the production under the international standards required to achieve international comparability.
- *Unless sustained by stable economic forces and strong reputation one-off surveys by unofficial agencies are unlikely to be efficient or sustainable over time.* They should therefore be avoided and donor organizations that support the strengthening of statistical systems need to be wary of devoting resources to such surveys.
- A preferred option is to include business ICT use surveys in national statistical programmes.
- A set of general recommendations on Statistics productions on e-Commerce/DE has been issued by UNCTAD .

# International Recommendations/Data Sources

- Sources for business ICT data -1
- There are various data sources used by countries that compile data on business use of ICT. They include administrative sources, business registers, questions or modules in host survey vehicles, stand-alone data collections and big data repositories. International organizations, such as the UN, also gather and harmonize data from different countries to create publicly available registers that can be used to compute some Digital Economy indicators.
- These sources do not have equal potential for producing data on the use of ICT by businesses. The appropriateness of each source is largely determined by the balance between the type of information sought (reflecting users' needs) and available resources. The next table presents different sources, the indicators for which they are likely to be most suited, and indications of their relative cost.

# International Recommendations/Data Sources

- Sources for business ICT data -2

Type of sourceIndicators that may be collected		Indication of costs
Administrative sources	Limited number of indicators on the availability of basic ICT infrastructure (in some countries these data are available through regulators)	Not expensive (by-product of administrative activities).
Statistical business registers (Mainly business and trade)	Limited number of indicators on the availability of basic ICT infrastructure with selected breakdowns (size, sector of enterprises).	Medium cost (for establishment and maintenance).

# International Recommendations/Data Sources

## Sources for business ICT data / Administrative Sources

- In some countries, suppliers of ICT services (such as fixed and mobile telephone, as well as Internet services) compile information about their clients' businesses and supply such information to government regulatory bodies. As a by-product of this regulatory activity, it may be possible to produce at a low-cost statistical information on access to ICT by businesses.
- However, the indicators that can be produced in this way are likely to be limited in number and in other ways as well. For instance, they will usually be restricted in scope to legal entities that have contracted ICT services and such entities may not correspond exactly to a business as defined statistically. Also, unless the necessary information is required for the contracts, disaggregation by business size or industry is not possible.
- In many countries, even this source of administrative information will not be available because ICT service providers do not distinguish business subscribers from household subscribers and so cannot provide data on business access to ICT.
- In summary, administrative sources will usually be insufficient for collecting data needed to produce statistics on ICT use by the businesses sector.



# International Recommendations/Data Sources

## Sources for business ICT data / Statistical Registers - 1

- Statistical registers - mainly business registers and trade registers - can play a relevant role when seeking to estimate digital economy indicators. Moreover, trade and business registers can be cross-referenced to produce breakdowns of the foreign trade ICT indicators in terms of enterprise features, such as size or activity sector.
- Business registers or directories are a key element of the statistical infrastructure of most NSOs. Their role is to maintain an updated record of a country's businesses (usually enterprises and their establishments), with information on the location, contact details and other characteristics such as industry and size. Business registers are used in the statistical process for the compilation of business demographic data and to generate population frames for business surveys. Most NSOs regularly update their business registers using administrative sources, such as tax or social security registers, and through their own statistical operations.
- The main advantages of using a business register as a source to produce statistical indicators (besides its use as a population frame from which to draw samples) are that:
  - ❖ Indicators can be quickly aggregated (no fieldwork is required); and
  - ❖ The marginal cost of statistical production is very low.

# International Recommendations/Data Sources

## Sources for business ICT data / Statistical Registers - 2

- Main disadvantage: the number of indicators on the use of ICT that can be produced from statistical business registers is generally limited to basic indicators on the presence of certain technologies such as telephone (fixed and mobile), computers or the presence of email or a website (which may be present for contact purposes). Such indicators will only be feasible, of course, if the business register contains high-quality (complete and updated) information.
- Administrative sources that are used to update an NSO's business register are usually good for identifying new businesses but less effective at detecting businesses that disappear from the population. The problem of having a high rate of "dormant businesses" (or "dead units") is generally more severe for small businesses.
- In developing economies, the importance of the informal sector implies that a big share of economic activity occurs outside the administrative framework (for example, activities of retail trade without licensing or tax control). The economic agents involved will usually be excluded from business registers.
- Coverage problems, such as those described above, can produce biases in statistical data. For example, there can be underestimation of the total number of businesses (if there is a large share of informal activity) or overestimation (if there is a large proportion of dormant businesses). Many developing economies are faced with both situations.

# International Recommendations/Data Sources

Sources for business ICT data / Statistical Registers - 3

- Trade registers are the usual source for computing ICT goods trade indicators for import and export. Trade registers are elaborated from the aggregation of the information of the Customs Authority of the country. In general, trade registers include the total volume and economic value of import and export transactions in a given period (usually one year).
- UNCOMTRADE is a DB that can be used to compute some of the ICT foreign trade indicators. It contains data on imports, exports, re-exports and re-imports, specifying annual data for each commodity by trading partner.
- challenge for the use of big data to measure (aspects of) the digital economy is that they are neither in general nor exhaustive in representing the general population. For this reason, measurements obtained from this type of source may exhibit relevant biases. The integration of big data with information obtained from registers, censuses or representative random sampling surveys is required to analyze and correct these potential biases

# International Recommendations/Data Sources

- Sources for business ICT data -3

Big data sources	Indicators on the use of ICT and e-commerce activities.	Big data on the use of ICT and online purchasing are usually owned by telecom operators and online platforms. Although such data can be obtained through strategic agreements with private partners with no cost, additional investment is required to assess the quality, clean, harmonize and integrate Big Data and other sources to cope with potential biases.
Module or questions on ICT embedded in existing sample surveys or censuses (these are often economic surveys, such as those of the manufacturing sector)	Indicators on the availability of basic ICT infrastructure with selected breakdowns (size, sector).  Generally limited number of indicators on use of ICT.	Mainly only marginal costs with respect to the cost of the survey to which it is attached.



# International Recommendations/Data Sources

## Sources for business ICT data / Big Data

- Big data are provided by different types of sources, including telecom operators, payment providers, postal operators, satellite companies, online platforms, social networks, and providers of Internet of Things (IoT).
- Big data sources generated by telecom companies are of special interest for the estimation of the core ICT indicators and measurement of the digital economy in general
- Challenge for the use of big data to measure (aspects of) the digital economy is that they are neither in general nor exhaustive in representing the general population. For this reason, measurements obtained from this type of source may exhibit relevant biases. The integration of big data with information obtained from registers, censuses or representative randomsampling surveys is required to analyze and correct these potential biases

# International Recommendations/Data Sources

## Sources for business ICT data / Modules and stand-alone surveys on the use of ICT by businesses

- Collecting data from businesses is usually a costly task that must be undertaken to the highest technical standards to ensure good quality of collected data (and resulting aggregates). The effectiveness and efficiency of a data collection system can be assessed from the viewpoints of different actors in the statistical system, namely data producers, data providers and data users:
  - ❖ **Data producers** are interested in obtaining high quality data at the lowest cost and in the shortest possible time. Data collection costs include outlays for preparation of collection instruments (usually, questionnaires), training interviewers and other staff, and the costs of collecting and capturing data; beyond the data collection phases, there are additional costs for data processing and dissemination.
  - ❖ **Data providers** (respondents) wish to minimize the burden of data collection in terms of their costs and time (for gathering requested data and completing questionnaires,
  - ❖ From the data users' point of view, a data collection system will be satisfactory if it ultimately provides relevant and reliable information, in an accessible way and in a timely manner.

# International Recommendations/Data Sources

## Sources for business ICT data / Modules and stand-alone surveys on the use of ICT by businesses

- The majority of OECD countries, as well as many developing countries, have collected data from businesses on the use of ICT through questions included in current business surveys or through stand-alone surveys on the topic. The choice of one or another approach is related to several factors, including policy needs for information on use of ICT by businesses, and the resources and organization of the statistical system.
- Developing economies may not be able to afford stand-alone surveys on ICT and instead may prefer to include questions in existing surveys that also include background information such as on employment and industry.
- Stand-alone surveys on ICT use are generally necessary for countries that are interested in investigating more sophisticated ICT applications (such as e-business, e-commerce and IT security measures).

# International Recommendations/Data Sources

## Sources for business ICT data / Modules on the use of ICT

- The decision to include a module on the use of ICT in an existing survey must consider:
  - ❖ The information needs that can be satisfied with such a module (variables that can be included and variables in the survey vehicle that can be used for analyzing ICT data). If the survey vehicle records classification variables such as industry and size, it will be possible to break down ICT indicators by these classificatory variables and
  - ❖ The methodology of the candidate survey vehicle in terms of population scope and coverage, periodicity, sample design, sample size and distribution. Data collected via the module will reflect the statistical characteristics of the survey vehicle (population frame used, sampling method, factors to weight the observations, collection and processing methods, level of detail for economic activities, etc.). This may be a significant disadvantage if, for instance, the scope of the survey vehicle is narrower than that recommended for the ICT use core indicators, or the sample size is too small, leading to large sampling errors.
- Some countries combine results from multiple surveys to publish ICT data . Candidate survey vehicles that have been selected in different countries include economy-wide business surveys, surveys on the manufacturing or primary sectors, surveys in the service sector and R&D



# International Recommendations/Data Sources

Sources for business ICT data / Modules on the use of ICT

- The cost of collecting ICT data via modules is generally marginal to that of the survey vehicle, since the most important part of the survey cost is usually related to data collection (including fieldwork, where relevant). In addition, staff involved in data collection and processing are already trained and will require only complementary training on ICT questions. For countries with severe budgetary restrictions, the inclusion of a module should be more cost-efficient than the implementation of a stand-alone ICT survey of businesses (though this will depend on factors such as the size of the survey vehicle and the complexity of the ICT module).
- The design of the module itself consists of choosing a limited number of relevant questions that can be easily interpreted by respondents.

# International Recommendations/Data Sources

## Sources for business ICT data / Economic census

- The Economic censuses are exhaustive surveys of the business sector (or parts of it), with the objective of collecting statistical information from all in-scope businesses in a country. In some countries, they include also the economic activities carried out by households. They are often used to construct population frames for sample surveys.
- Because of exhaustiveness, economic censuses are expensive to conduct, due to the cost of collection (which may require a detailed cartography of the country), and therefore, their periodicity is usually very low (5 or 10 years). This may make them inadequate for continuous monitoring of fast-growing phenomena such as Internet access

# International Recommendations/Data Sources

- Sources for business ICT data -4

Stand-alone ICT survey	Indicators on the availability of basic ICT infrastructure with selected breakdowns (size, sector). Indicators on use of ICT. Indicators on barriers to the use of ICT. Indicators on costs, value of investments, etc.	High cost for design, data collection (which could include fieldwork) and processing.
------------------------	--	---

# International Recommendations/Data Sources

## Sources for business ICT data / Stand alone surveys

- The Stand-alone surveys on ICT allow for the collection of significantly more information than modules in other surveys. Stand-alone surveys are especially designed to collect information on different topics of interest in the measure of ICT use by businesses, such as access to ICT, purposes of use, e-commerce, security, skills, etc. that usually go beyond the possibilities of a module embedded in an existing business survey.
- Several developing economies have implemented stand-alone ICT surveys of the business sector. These countries have relatively widespread access to ICT in comparison with other countries in their region or have an economy increasingly based on ICT. In general, in countries that have carried out stand-alone ICT surveys, there is a high demand for business ICT indicators from policymakers, the market and society.
- If done outside the statistical system, coordination must be ensured since external actors have no access to statistical infrastructure such as business registers, data collection systems (call centres, trained interviewers in regional or municipal offices, etc.), sampling methodology and the possibility that methodologies are not harmonized with those of the national statistical system (which are usually linked to international statistical standards).



# International Recommendations/Data Sources

## e-Commerce measurements

- **Adding questions to existing household and enterprise surveys:** relatively easy and cheap to implement but faces limitations in terms of the number of questions that could be included
- **Stand-alone surveys:** more costly and time consuming to implement but could provide more detailed information
- However, usually some info are not usually collected such as , MSMEs (particularly relevant to may countries), C2C (a growing segment of e-commerce), etc.
- Other sources can be big data owned by platforms

# International Recommendations/Data Collection

## Data Collection Methods - 1

Method	Main advantages	Main disadvantages
Face-to-face personal interview	<p>This is the most direct method of collecting information. It facilitates direct interaction between the interviewer and the interviewee, allowing for checking and follow-up questions. An interviewer can also assist respondents in answering complex questions and can clarify concepts such as definitions of ICTs. Because the interviewer is in view, s/he can use visual prompts such as prompt cards.</p> <p>In addition, face-to-face interviews are especially useful for questions about opinions or impressions, and for surveys that take a long time to complete.</p> <p>The technique usually produces lower non-response rates. Data collection can be managed efficiently with specific software (Computer Assisted Personal Interviewing – CAPI, see below).</p>	<p>Interviewers are part of the measurement tool and they can induce important biases if they have not received suitable training.</p> <p>High personnel costs may be incurred (for hiring and training interviewers). However, this could be a minor issue in developing economies where salaries of interviewers are low, or agreements are reached with certain institutions to provide part-time interviewers (such as university students).</p> <p>In developing economies with poor quality transport infrastructure, reaching businesses located in some country areas may prove difficult.</p>

# International Recommendations/Data Collection

## Data Collection Methods - 2

<b>Telephone personal interview</b>	<p>Although to a lesser extent than the face-to-face personal interview, telephone interviewing allows direct interaction between the interviewer and interviewee.</p> <p>It is a fast and relatively inexpensive way to collect information, since a small number of interviewers from a single call centre can carry out a great number of interviews.</p> <p>The data collection can be managed efficiently with specific software (Computer Assisted Telephoning Interviewing – CATI, see below).</p>	<p>Correct and comprehensive telephone numbers may not be available, particularly in developing economies where mobile telephony may be more common than fixed telephone.</p> <p>Interviews must be relatively short since a long telephone conversation can be perceived as an annoyance. Some people also feel that it is intrusive to be interviewed by telephone.</p> <p>Telephone interviews may not be suitable for collecting quantitative information, for which the interviewee may have to check business records.</p> <p>The non-response rate is usually larger than for face-to-face interviews (but lower than for mail-based surveys).</p>
-------------------------------------	---	---

# International Recommendations/Data Collection

## Data Collection Methods - 3

Method	Main advantages	Main disadvantages
Interview assisted by computer (CAPI/CATI)	<p>CAPI and CATI systems can eliminate errors of flow and data consistency and can thus improve input data quality and reduce the time for data capture and validation.</p> <p>Questionnaires can possibly be customized based on available information about the business.</p> <p>Modern IT equipment such as PDAs or smartphones may present a cheap and comfortable tool for data collection.</p>	<p>CAPI and CATI techniques require interviewers with some technical skills.</p> <p>CAPI and CATI systems are usually based on commercial software that may be costly. Skilled staff are required to adapt the software to the questionnaire.</p> <p>CAPI requires that interviewers carry costly IT equipment, which can be damaged, stolen, etc. during field operations.</p> <p>In developing economies with poor road networks, there is a risk of damaging the equipment.</p>

# International Recommendations/Data Collection

## Data Collection Methods - 4

<b>Mail survey</b>	<p>This method is relatively inexpensive, and the statistical office can send the same measurement instrument (questionnaire) to many businesses.</p> <p>It allows the respondent to complete the questionnaire at his or her convenience.</p> <p>It eliminates the problem of interviewer bias though note that interviewer follow-up (e.g., for non-response or inconsistent answers) can potentially introduce bias if not managed properly, and if questionnaires are not properly designed and tested, they can introduce bias to the survey results.</p>	<p>Requires separate data entry unless advanced Optical Character Recognition (OCR) tools are available. It usually suffers from high non-response rates.</p> <p>It is not designed for detailed written responses, but for numerical questions or those that can be answered by selecting a limited list of choices (including yes/no responses).</p> <p>The lack of help from an interviewer can produce information of low quality. It therefore requires clear questions and instructions.</p> <p>Delays in mailing back questionnaires can induce delays in the survey. In developing economies with a low-quality postal system, such delays may be prohibitive.</p> <p>Some of the problems inherent to a postal survey can be partially solved by, for instance, use of written or telephone reminders to reduce non-response rates. In addition, data quality can often be improved if a telephone helpline is available.</p>
--------------------	--	--



# International Recommendations/Data Collection

## Data Collection Methods - 5

<b>Electronic survey</b>	<p>With the growth in ICT skills and availability of ICT, possibilities for electronic surveys have increased. Data collection mechanisms can be of several types, but the most common are electronic questionnaires sent by e-mail or posted on web pages that respondents can access.</p> <p>This method has most the advantages of mail surveys but is usually faster and cheaper. Because respondents complete the questionnaire electronically, manual data entry is not required, and edits can be applied at the time of data entry (and resolved by the respondent).</p>	<p>Businesses that can be surveyed this way do not cover the entire business population, in particular in developing economies with low ICT penetration. This will cause either bias in the data, or the necessity of using another method as well (e.g., mail out questionnaires) for data capture.</p> <p>There is an extra need for technology to ensure security and confidentiality of data and for staff with the training and skills to handle the data collection tools. Costs associated with this expertise can offset savings offered by electronic data capture.</p> <p>The technique is in general not suitable to be used as the only channel for collecting data. However, if complemented with other methods, it can be a useful tool.</p>
--------------------------	--	--

# International Recommendations/Data Collection

## Data Collection Methods - 6

Method	Main advantages	Main disadvantages
Big data acquisition	<p>Establishing strategic agreements with private owners of big data sources could facilitate the acquisition of relevant data covering actual company behaviour at a reduced cost.</p> <p>Publicly available big data sources could be acquired by applying web scraping methods and APIs, with no cost to the respondent and avoiding hypothetical response biases.</p>	<p>Big data sources are not in general representative of the corresponding populations and the estimations from these sources need to be corrected by applying results from representative samples or official sources.</p>

# International Recommendations/Data Collection

## Data Collection Methods - 7

Behavioural economic experiments (BEEs)	BEEs are flexible methods to observe actual digital behaviour, allowing the control of the conditions under which these observations are undertaken.	BEEs provide formation on company's digital behaviour in gamified environments. The validity of the conclusion for real decision-making (ecological validity) needs to be supported by a sound experimental design.
---	--	---

# International Recommendations/Quality Control

## Quality control good practices -1

- Use in-scope businesses and correct addresses
- Provide suitable training to interviewers on the questionnaire contents (especially required for complex technical concepts), and on dealing with respondents;
- Prepare questionnaires so that questions are worded in a correct, clear and unambiguous manner, and respondents can perceive a logical flow in the order of questions (especially in self-administered questionnaires);

# International Recommendations/Quality Control

## Quality control good practices - 2

- Filter the collected data by a series of controls that are applied at the moment of data capture and in the data entry process;
- Give wide publicity before the start of the survey, highlighting the relevance of data collection for national policies and therefore the need for respondents' collaboration. This may include a mention of the compulsory response required in countries where the statistical law grants this;
- Establish a policy of incentives and sanctions that encourage the provision of good answers to interviews or questionnaires



# International Recommendations/Data Processing

## Data Editing

- *Micro-editing refers to controls, validations and modifications applied to the data of a given business. The process includes the treatment of incomplete or missing data and the detection and treatment of answers that are inconsistent with other questions, and*
- *Macro-editing refers to controls, validations and modifications of whole datasets by means of the analysis of aggregations. The aim of the process is to check whether certain estimates are jointly compatible and are consistent with other knowledge. A sophisticated macro-editing procedure consists of readjusting sample weights according to frame errors detected during the survey. This point is elaborated later in this chapter.*

-

# International Recommendations/Data Processing

## Data Editing- 1

- **Treatment of internal inconsistencies and errors.** It is very important that practices be established that decrease the incidence of incomplete or inconsistent data, so that the impact of data editing is minimized. Quality controls already embedded in data collection instruments or at the data entry stage will directly improve the quality of raw data and reduce the task of data processing.
- **Treatment of missing data.** To reduce the extent of non-response, by means of advertising the survey to data providers through the best available media, improved questionnaires, high-quality fieldwork (if used) and good non-response follow up.

# International Recommendations/Data Processing

## Data Editing- 2

### Weighting procedures.

- Surveys on ICT use by businesses are generally based on a stratified random sample design, with strata defined by industry and size (at least). Businesses should be selected at random from strata (except for those that are completely enumerated, e.g., large business strata).
- It is assumed that the sample design is based on random selection, without replacement, within strata. Stratum estimates are therefore calculated based on a simple expansion (weighting) to the total number of businesses in the stratum. The method also applies if the selection is systematic with a random starting point in each stratum. The method explained below can be applied to both qualitative variables (such as the presence of a website) and quantitative variables (such as the number of employees who used the Internet).

# International Recommendations / Dissemination

- Agencies that produce and disseminate ICT indicators should enhance their usability by routinely disclosing associated metadata. It is strongly recommended that production of these reports is integrated into the statistical production process and not undertaken as a separate activity.
- The Eurostat Code of Practice is well documented and can be used to specify appropriate quality information about ICT use indicators according to six quality dimensions: relevance, accuracy, timeliness and punctuality, accessibility and clarity, comparability, and coherence. Between them, the six dimensions cover the range of metadata that result from statistical collection work.

# International Recommendations / Technology Trends

- **Big Data.** THE UNECE Big Data project showed the feasibility of the production of official statistics based on Big Data and the EU Essnet Big Data demonstrated the feasibility of the integration of Big Data in the regular production. Use of mobile phone data source is reported for the measurement of tourism, population, and commuting in as well as satellite imagery for agricultural statistics.
- **Web scraping.** It is the process of automatically collecting information from the World Wide Web, based on tools (called scrapers, internet robots, crawlers, spiders etc.) that navigate, extract the content of websites and store scraped data in local data bases for subsequent elaboration purposes. Many examples of the use of Internet data sources can be reported. For instance:
  - ❖ Internet queries: the use of Google Trends has been evaluated in order to produce now-casting estimates of unemployment indicators;
  - ❖ Web prices: Web scraping is already in use in order to collect prices related to goods and services for the construction of Consumer Prices Indexes;
  - ❖ Social media: posts in social media, like Twitter or Facebook, can be used in order to support the production of traditional Official Statistics indexes like, e.g., the Consumer Confidence Index



# International Recommendations

## Discussion, Questions / Answers

- State of play in Mongolia