

Methodological note

Experimental statistics on short-stay accommodation offered via online collaborative economy platforms

Paper prepared for the 16 April 2021 meeting of the METAC

Background and relevance of the project

The « raison d'être » of this project is two-fold.

Firstly, the collection of data from platforms serves to improve the quality of European statistics on tourism, in particular accommodation statistics as defined in Annex I of [Regulation \(EU\) 692/211](#).

In the accommodation market, the segment of short-stay holiday rentals was traditionally under-covered. The high volume of small market players were often not included in business or tourism registers, in spite of the significant number of players, their share in the total volume of nights spent was estimated too low to justify a data collection. This was also one of the reasons to allow a data collection threshold in the Regulation: Member States can opt to limit the scope of observation to establishments having at least 10 bed places (or 20 bed places for smaller tourism countries). In practice, this means that a significant part of the holiday rentals in NACE 55.2 is currently not represented in the official statistics.

The rise of the online platforms increased the coverage problem, but also created a solution. While this share of the market is potentially growing because the existence and user-friendliness of the platforms make it easier for service providers to promote and to run their business, the platforms also capture their digital footprint (which opens perspectives for collecting statistical data...).

Secondly, the project's relevance goes beyond the concrete application for tourism statistics. It is a test case for the highly anticipated access to privately held data by statistical authorities. Therefore, the project is also part of the Trusted Smart Statistics portfolio.

Access to privately held data

One of the lessons learnt from Eurostat's work on big data¹, is that getting access to the data can be a major blocking factor. Early negotiations with Airbnb (2016-2017) did not lead to an exchange of data, partly due to a lack of understanding by statisticians of the specific problems that private companies are facing.

The project got a new start in late 2018, as a joint initiative of DG GROW and Eurostat. From the beginning, the approach was multi-disciplinary, not only having data scientists and statisticians around the table, but also involving (on both sides) legal and contracts experts, IT experts and public policy experts. Another critical success factor, was the inclusion of more than one company holding such data. The enlargement to four platforms (Airbnb, Booking, Expedia, Tripadvisor) contributed not only to a better representativeness and reduced the vulnerability of the project, but also created a fair level playing field for the different actors in this segment of the accommodation market.

¹ For instance, in the domain of tourism statistics, the [Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics](#).

Finally, along the way, Eurostat learnt that in negotiating with private entities, the statisticians can no longer decide from A to Z but need to match the expectations and statistical requirements with the margin of manoeuvring that the private entities can afford against their own internal rules, external stakeholders and investor relations.

This created a wider support base for the project and eventually led to the conclusion of non-disclosure agreements signed between the four platforms and the European Commission (Eurostat) at the end of 2019. The agreements focused on the main principles of the exchange of data and on the security measures to be taken regarding the transmission, processing, storage, sharing and dissemination of the data.

Governance in the ESS

Eurostat being the central hub for the platforms data, the project somehow inverts the governance in the ESS. While the Member States (NSIs) are traditionally in charge of the data collection and the development of national methodologies, this role is now transferred to Eurostat. However, this is in line with the subsidiarity principle: indeed, it was acknowledged that the optimal approach to getting the platforms data was a coordinated, centralised initiative by Eurostat (jointly with DG GROW).

The process takes place in close cooperation with the Member States, at technical as well as strategic level. On several occasions, the project and the progress made was presented to the Working Group on Tourism Statistics, to the Business Statistics Directors Group (BSDG), the European Statistical System Committee (ESSC) and the Partnership Group (PG). In May 2019, a dedicated half-day informal workshop at the level of directors-general of the NSIs was organised. Several webinars regularly update the Working Group on the project. In the 2019 Working Group, a Steering Group (SG) was installed, composed of delegates from five Member States². The SG assists Eurostat in all relevant negotiations with the platforms (input to documents, participation in meetings).

From the start, Eurostat convinced the platforms that it was essential for the countries to have access to the data. While the non-disclosure agreements are very strict on the further sharing of the data, an exception regime is included for the NSIs (who will receive the same granularity of data, but merged for the four platforms). The involvement of the NSIs is essential for two reasons: the national publication needs, and the validation of the data using the national knowledge of the local market. Bilateral agreements were concluded between Eurostat and the NSIs to enable an exchange of data (and to allow, in a later phase, also an exchange to Eurostat of data on nationally operating platforms).

Scope

Sectoral coverage

At the technical meetings held between the platforms and Eurostat (gradually assisted by the Steering Group), it was agreed to limit the scope to accommodation facilities that fall under NACE 55.2 "holiday and other short-stay accommodation". This scope reflects the segment of accommodation where service providers in the collaborative economy are primarily active. A next stage could envisage an expansion to other types of accommodation such as hotels.

² Currently composed of tourism statistics experts from Spain, France, Latvia, Slovenia and Finland.

As is the case with administrative data, the data held by the platforms was not designed to serve statistical purposes in the first place. An important part of the technical discussions with the platforms focused on aligning their (non-harmonised) categories of accommodation with the existing concepts and definitions for tourism statistics.

The data will be based on submissions by four international platforms: Airbnb, Booking, Expedia and Tripadvisor. While these platforms most likely represent a very significant part of this segment of the market, users of the statistics will be duly informed about the coverage. It will also be clearly communicated that this data on accommodation offered via online platforms can – for the time being – not be added to the existing official statistics on occupancy of tourist accommodation establishment (compiled under Regulation (EU) No 692/2011).

Geographical coverage

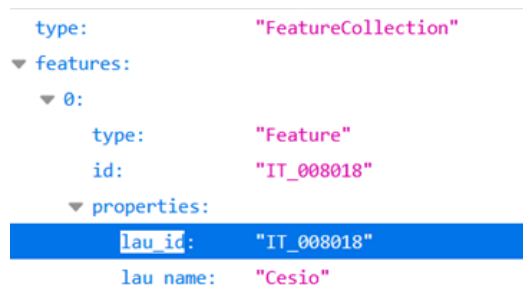
The agreements with the platforms cover the European Union and EFTA countries, but they have shown flexibility to expand to other European countries (given the same methodology and algorithms can easily be applied to other geographical areas).

The data is transmitted at the most detailed administrative granularity, namely aggregated per Local Administrative Unit (LAU) – see further.

The platforms match the listings in their databases with the LAU borders using shape files made available by Eurostat or via an [API](#) returning the LAU code based on queries of geographical coordinates. While the procedure generally performs well, a few issues remain – in particular in border areas.

Example of a query: <https://gisco-services.ec.europa.eu/id/lau?x=8.0&y=44.0>, returning:

```
{"type": "FeatureCollection", "features": [{"type": "Feature", "id": "IT_008018", "properties": {"lau_id": "IT_008018", "lau_name": "Cesio"}}]}
```



Temporal coverage

The platforms agreed to send data for the reference year 2018 and onwards. Back series before 2018 were not feasible due to changes to the information systems over time.

At the time of writing (March 2021), Eurostat had received full datasets for 2018 and 2019. In the second half of 2021, the transmission of the 2020 data can be expected. Soon after, the regular transmission at quarterly intervals of quarterly data is expected to start. The platforms agreed to transmit the datasets within 45 days after the end of the reference quarter³.

³ Note that the transmission deadline will most likely be slightly revised to align with the changed financial reporting obligations of the platforms (all four partners being publicly listed companies).

Data structure definition

Technical discussion between Eurostat and the platforms on the variables and breakdowns took place in 2019. It was agreed to create three datasets: one dataset on capacity (to be sent annually), two datasets on occupancy (to be sent quarterly). The rationale to split the occupancy data into two datasets was to limit the size of the files, by splitting into characteristics for the listings (OCCL file) and characteristics of the visitors (OCCV file). The capacity file (CAP) is already transmitted by the platforms, but it was agreed to postpone the exploitation of the file until the issue of double counting is solved – see further.

Variables and breakdowns

| Label | Description | Number of categories (est.) |
|---|---|--|
| Study variables for capacity and occupancy | | |
| cap_host | Number of hosts (number of hosts renting out one or more listings) | 1 |
| cap_list | Number of listings | 1 |
| cap_bedp | Number of bed places | 1 |
| occ_stay | Number of stays (number of rentals of each listing during the reference period) | 1 |
| occ_ni | Number of nights rented out (number of nights each listing was rented out during the reference period) | 1 |
| occ_gni | Number of overnight stays (number of guest nights spent at each listing during the reference period) | 1 |
| Explanatory variables and breakdowns | | |
| TIME | The reference year & month (format YYYYMM, e.g. 201801) | 12 months |
| GEO_CC | Country code | (31: EU + IS+LI+NO+CH) |
| GEO_LIST | The LAU code where the listings are located (e.g. BE12345) | +/- 97.600 |
| GEO_GSTF | (only for OCCV) Country of origin of the guests (at country level for EU and EFTA, grouped by regions for the rest of the world, with some important partner countries reported separately) – full version [see Regulation (EU) 692/2011, Annex I, Section 3.E] | 50 [34 countries + 1 region for Europe; and 8 other countries + 5 regions for the rest of the world + 'not allocated' + world total] |
| GEO_GST2 | (only for OCCL) Origin of the guest – reduced version with 2 categories (residents versus non-residents) | 2 |
| BREAKDOWN | Defined by combining categories of type of accommodation (first phase: only NACE 552) x type of host (peer vs. pro vs. unknown) x type of listing (shared vs. entire) x size class (less than 10 bedplaces vs. 10 or more). Eg.: 552_PEER_SHAR_LT10 refers to holiday rentals in listings by peers, with shared facilities and less than 10 bedplaces in total). | 8 [1 (nace) x 3 (hosttype) ⁴ x 2 (listtype) x 2 (sizecl)] |
| VALUE | The actual value for the corresponding variable (e.g. number of overnight stays) | - |

⁴ The aim is to have only two categories, namely peers versus professionals. However, given the known issues in implementing this breakdown in the short term, a third category “Unknown” is temporarily added, until re-evaluation of the issue.

CAP file format

One single file, including the 12 months, including in the column 'Value' the figures for the three variables mentioned in the first column 'Variable', for each breakdown defined by the other columns.

| VARIABLE | TIME | GEO_CC | GEO_LAU | Breakdown | Value | Flag | Comment |
|----------|--------|--------|---------|--------------------|-------|------|---------|
| cap_host | YYYYMM | CC | LAU-1 | ACC_HOST_LIST_SIZE | | | |
| cap_list | | CC | LAU-x | | | | |
| cap_bedp | | CC | LAU-n | | | | |

TIME = year + month

GEO_CC = the country code (for the purpose of the test files: LV, ES or FI)

GEO_LAU = the LAU code

Breakdown: defines the combination of categories of the explanatory variables

- **ACC** ~~551~~, 552 or ~~553~~ (currently, only 552 is considered)
- **HOST** PEER, PROF, UNKN (for the first deliveries, a category Unknown is allowed)
- **LIST** SHAR (shared), ENTR (entire)
- **SIZE** LT10 (less than 10 bed places), GE10 (greater or equal to 10).

Note: the capacity refers to the maximum capacity during the reference period (month), namely all listings that were active (available or booked) during the given month, excluding those listings that were never rented in the course of the reference year.

OCCL file format

This file collapses the information regarding the origin of the guest (only keeping a dichotomous split between residents vs. non-residents).

| VARIABLE | TIME | GEO_CC | GEO_LIST | GEO_GST2 | Breakdown | Value | Flag | Comment |
|----------|--------|--------|----------|----------|--------------------|-------|------|---------|
| occ_stay | YYYYMM | CC | LAU-1 | RES | ACC_HOST_LIST_SIZE | | | |
| occ_ni | | CC | LAU-x | NRES | | | | |
| occ_gni | | CC | LAU-n | | | | | |

TIME = year + month

GEO_CC = the country code (for the purpose of the test files: LV, ES or FI)

GEO_LAU = the LAU code

GEO_GST2 Origin of the guest: resident / non-resident (of the country)

Breakdown: defines the combination of categories of the other explanatory variables

- **ACC** ~~551~~, 552 or ~~553~~ (currently, only 552 is considered)
- **HOST** PEER, PROF, UNKN (for the test files, a category 'UNKN' or Unknown is included)
- **LIST** SHAR (shared), ENTR (entire)
- **SIZE** LT10 (less than 10 bed places), GE10 (greater or equal to 10)

OCCV file format

This file collapses the information regarding the listings (type of host, type of listing, size class).

| VARIABLE | TIME | GEO_Count | GEO_LIST | GEO_GSTF | Breakdown | Value | Flag | Comment |
|----------|--------|-----------|----------|----------|-----------|-------|------|---------|
| occ_stay | YYYYMM | CC | LAU-1 | 1 ... 48 | ACC | | | |
| occ_ni | | CC | LAU-x | | | | | |
| occ_gni | | CC | LAU-n | | | | | |

TIME = year + month

GEO_CC = the country code (for the purpose of the test files: LV, ES or FI)

GEO_LAU = the LAU code

GEO_GSTF = Origin of the guest: full breakdown (50 categories)

Breakdown: defines the combination of categories of the other explanatory variable(s)

- **ACC** ~~551~~, 552 or ~~553~~ (currently, only 552 is considered)

Transmission and storage

The data is transmitted using EDAMIS. This necessitates the creation of a new EDAMIS domain ("CETOUR") and EDAMIS user accounts for the four private companies (where EDAMIS is normally used for transmission by NSIs or ONAs).

The transmission of the data, and the quality of the files (format, structure, contents) was extensively tested based on test files containing real data for selected regions and cities in three countries, in the first half of 2020.

At Eurostat, the data is stored on the secure servers. In line with the non-disclosure agreements, only a limited number of staff can access the data (including a few external consultants working on IT solutions, in particular development in SAS).

Transmission of the data to the NSIs will also take place via EDAMIS. The bilateral agreements signed between Eurostat and the NSIs explicitly stipulates the secure storage and access in the NSIs.

Metadata

The non-disclosure agreements with the platforms also mention providing metadata that can be publicly shared with the data releases. The platforms completed the metadata files and returned the documents by mail (a switch to the ESS-MH will be considered later on).

The template can be found in Annex I.

It still needs to be discussed internally and with the platforms if the metadata will be published together with the first data release.

Dissemination and release policy

Subjects and characteristics

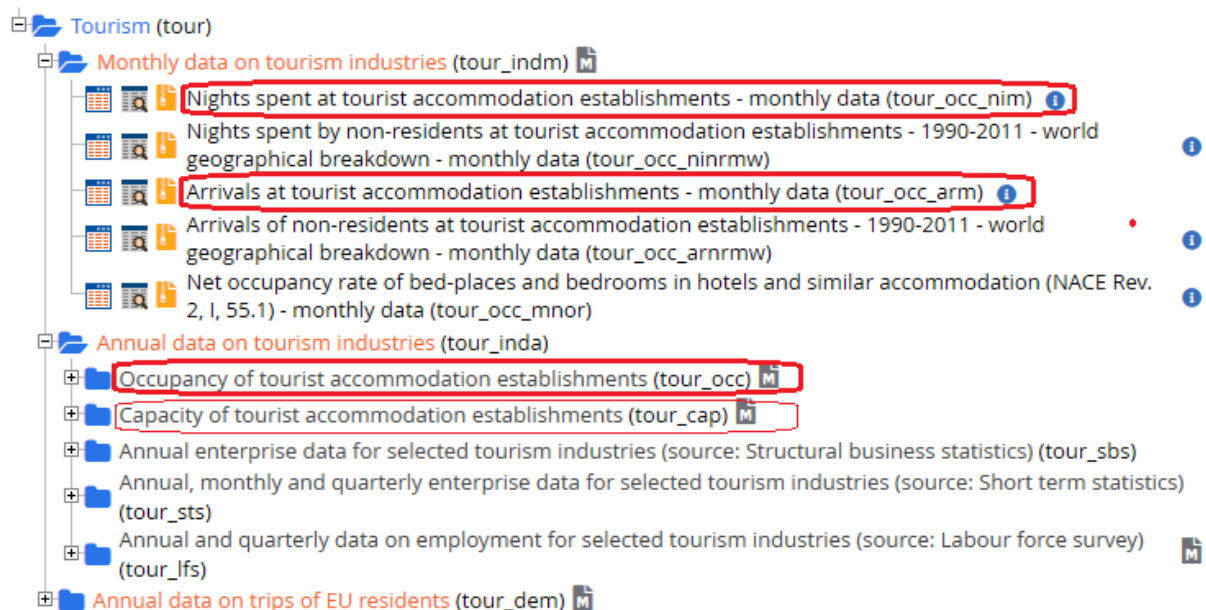
The data output will be similar to the current traditional accommodation statistics, for two reasons. Firstly, to offer users data comparable with the data they are familiar with (clarity, coherence, consistency). Secondly, to guarantee a level playing field in terms of type and detail of data released for the traditional accommodation segment and the platforms based segment.

The aggregate tables to be included in the first release are largely inspired by the data on tourism industries (in particular the accommodation sector covered by NACE 55.1, 55.2 and 55.3) – see Eurobase data tree below. In the context of the platforms data, the “nights spent at tourist accommodation establishments” are obtained by using the variable “number of overnights stays (occ_gni)” (= the number of nights the accommodation was rented out, multiplied by the number of guests staying at the accommodation). The “arrival at tourist accommodation establishments” are not directly available from the platforms data, but the “number of stays (occ_stay)” is used as a proxy for the data on arrivals (note that a closer proxy to the “arrivals” can be calculated by multiplying occ_stay with the ratio occ_gni/occ_ni (\approx the average number of guests)).

Note that data on capacity of tourist accommodation establishments is not included in the first release of the experimental statistics, as agreed with the platforms. Data on capacity will only be published after having dealt with the issue of double counting (see further).

The first data release of the experimental statistics will already incorporate the forthcoming changes to the traditional output, recently introduced by Commission Delegated Regulation (EU) 2019/1681: monthly data at NUTS 2 regional level, annual data at NUTS 3 regional level and for selected cities (capital cities, cities with 200.000 or more inhabitants, and other selected cities with relevance for tourism⁵).

It is important to stress that the experimental statistics concern stand-alone data on accommodation booked via four online platforms, covering a very significant part of this market segment (but not the entire online segment).



Draft lay-out of the first tables

- Monthly data
 - T1a. Number of stays (occ_stay), by country; broken down by resident/non-residents (GST2) and by month
 - T1b. Number of nights spent (occ_ni), by country; broken down by resident/non-residents (GST2) and by month

⁵ Defined as those cities that, jointly, contribute to the top 90% cumulative nights spent in cities in the country.

- T1c. Number of guest nights spent (occ_gni), by country; broken down by resident/non-residents (GST2) and by month
 - T2a. Number of stays (occ_stay), by NUTS1/NUTS2 region; broken down by resident/non-residents (GST2) and by month
 - T2b. Number of nights spent (occ_ni), by NUTS1/NUTS2 region; broken down by resident/non-residents (GST2) and by month
 - T2c. Number of guest nights spent (occ_gni), by NUTS1/NUTS2 region; broken down by resident/non-residents (GST2) and by month
 - T3a. Number of stays (occ_stay), by city*; broken down by resident/non-residents (GST2) and by month
 - T3b. Number of nights spent (occ_ni), by city*; broken down by resident/non-residents (GST2) and by month
 - T3c. Number of guest nights spent (occ_gni), by city*; broken down by resident/non-residents (GST2) and by month
(* capital cities (CPTL), top50 cities in terms of annual guest nights spent (TP50), top100 cities that are not in top50 (T100))
- Annual data
 - T4. Number of guest nights spent (occ_gni), by NUTS3 region; broken down by residents/non-residents (GST2)
 - T5. Number of guest nights spent (occ_gni), by city**; broken down by residents/non-residents (GST2)
(** capital cities (CPTL), cities with at least 200.000 inhabitants (200K), other cities that contribute to the top 90% cumulative nights spent in cities in the country)
 - T6. Number of guest nights spent (occ_gni), by country; broken down by country of origin of the guest (GSTF)
 - T7. Number of guest nights spent (occ_gni), by NUTS1/NUTS2 region; broken down by country of origin of the guest (GSTF, fewer categories)
 - T8. Number of guest nights spent (occ_gni), by country; broken down by type of accommodation
 - T9. Number of guest nights spent (occ_gni), by country; broken down by size of the accommodation

Release policy and disclosure control

In the negotiations with the platforms and the drafting of the non-disclosure agreements, the data protection was an essential element.

Provisions were agreed to avoid the identification of individual platforms data, at least during a certain period following the reference quarter.

The initial agreements laid down a 13 months embargo period for sensitive data. However, to avoid a potentially complex and opaque dissemination strategy, the process was streamlined in a recent joint meeting with the platforms (February 2021). It was agreed to apply a two-stage dissemination [tentative approach, still to be confirmed via formal amendments to the non-disclosure agreements]. Data at country level and for capital cities can be released as soon as possible after the validation of the data (expected to be around three months after the end of the reference period), and the remainder of the data six months after the end of the reference period.

Since the data to be published in the first release of the experimental statistics refers to 2018 and 2019, both the initial disclosure rules with an embargo period of 13 months, as well as the

forthcoming rules with a deferred publication until six months after the reference period are being observed.

Outstanding methodological issues

Eurostat Unit G-2, jointly with Unit A-3, has implemented validation rules in SAS. The validation of the data is very satisfactory, thanks to the good cooperation and responsiveness of the data providers. For the current scope and ambitions of the project (namely stand-alone information on occupancy booked via online platforms) leading to the first release of experimental statistics, no significant methodological issues remain.

Before the dissemination can be extended to including total occupancy figures for the accommodation sector or capacity figures for the platforms, two ways of double counting need to be solved.

Firstly, service providers often list their property on more than one platform. Therefore, data on the number of listings (= capacity data) is not additive. In the course of 2021-2022 methodological work will be undertaken to find methods for deduplication. These methods can range from scraped data (for example, Finland has experimented with text matching to find duplicates, Spain has experimented with picture matching across platforms) to register level matching processes (e.g. using secure multi-party computation techniques). In the case of occupancy data obtained from the four platforms, it is assumed that there is no double counting because only one single platform can record the transaction for a given stay and the corresponding nights spent.

Secondly, service providers that are captured via the footprint they leave on the platform on which they advertise their property, may also already be included in the sample of the NSIs for the surveys on tourist accommodation. Here too, a method to deal with the double counting is needed before the data from the two sources (surveys, platforms) can be added to obtain a total (or at least more complete) estimate for occupancy of rented accommodation. The methodological work for deduplication of occupancy data will also start in the course of 2021.

Contacts

Eurostat Unit G-2:

- Christophe Demunter
- Simon Bley
- Carsten Olsson (head of unit)

Annex

Metadata structure for the agreement on the Exchange of data from online platforms regarding short-stay accommodation⁶

Metadata report (data reference year: xxxx)

Description of the metadata structure

Eurostat (and its users) need methodological information on how the data submitted by the platforms is produced (delineation of scope, definition and estimation/approximation of variables and categories, known shortcomings in the data, etc.).

The template below intends to gather this information in a standardised way, which is user-friendly to complete and to consult.

Please note that information already known by Eurostat, has been pre-filled in the template, but needs to be checked and confirmed.

In some cases, several options are proposed as example, but you will need either to select those applicable or add new ones, depending on your specific situation. When filling in, please keep in mind that the metadata will also be shared with users of the data.

⁶ For indicative purposes, based on the 2018/2019 metadata template used for the first exchange of data with the international platforms.

1. CONTACT

| | |
|---|--|
| Organisation / company | |
| Contact name(s) | |
| Contact e-mail address(es) | |
| Last update of this metadata file (date) | |

2. COVERAGE

| | |
|---|--|
| Economic activities covered (NACE) | [pre-filled] NACE 55.2 |
| Geographical coverage | [pre-filled] All local administrative units in the EU + IS + NO + CH + LI + UK <i>(please report any deviation, e.g. French overseas territories)</i> |
| Time: available reference periods | <i>(please fill in the reference years (YYYY), or quarters for years that are not entirely available (YYYY/QQ))</i> |
| Other relevant info relating to coverage | <i>(please report on coverage issues that may affect comparability, e.g. built-in pre-selection of certain types of accommodation)</i> |
| Reference period | [pre-filled] Data on occupancy (files OCCL and OCCV): month Data on capacity (files CAP): month |
| Transmission frequency | [pre-filled] Data on occupancy (files OCCL and OCCV): quarterly Data on capacity (files CAP): annual |

3. METADATA FOR VARIABLES

| VARIABLE | STANDARD DESCRIPTION | METADATA (TO BE FILLED OUT BY PLATFORM) |
|----------|--|--|
| cap_host | Number of hosts (renting out one or more listings) | <p><i>Please explain how you determine the number of hosts:</i></p> <ul style="list-style-type: none"> - Are all hosts with active listings during the reference month included (at least one active listing)? [Yes / No] - Are all host <i>without</i> active listings during the reference month excluded? [Yes / No] - How do you define "active"? ... - How are property managers offering properties for several hosts treated? [counted as several hosts (each renting out 1 or more properties) / counted as one 'host' / other (specify)] - Is capacity data obtained from snapshots in different moments (i.e. end of the month)? [Yes / No] - Is capacity data obtained from occupancy records? [Yes / No] - Other comments (if any): ... |
| cap_list | Number of listings (e.g. host offers three rooms in the same apartment = 3 listings ; host offers three (separate) apartments in the same block of apartments = 3 listings) | <p><i>Please explain how you determine the number of listings:</i></p> <ul style="list-style-type: none"> - Treatment of properties with no booking in the reference period (where reference period is 'month', while transmission frequency is 'annual')? [excluded / included / other threshold (please specify)] - Are shared rooms in the same property counted as one listing? [yes / no (counted as separate listings)] - Can entire properties be counted as separate listings (e.g. an apartment listed simultaneously as an entire apartment as well as separate rooms)? [Yes / No] - The agreed concept for "capacity" refers to "the maximum capacity during the reference period (month), i.e. all listings that were active (available or booked) during the given month, excluding those listings that were never rented in the course of the reference year". Are there deviations from the standard concept? [No / Yes (please specify)] - Are listings integrated from other platforms or from independent property managers excluded from the reported listings? [Yes / No] - Other comments (if any): ... |
| cap_bedp | Number of bed places (best possible approximation based on e.g. the maximum capacity shown for each listing) | <p><i>Please explain how you determined the number of bed places:</i></p> <ul style="list-style-type: none"> - Self-declared maximum number of bed places for each active listing (where active listings mean at least one listing during the reference period, where reference period is 'month')? [Yes / No (please specify)] - Treatment of permanent beds: single bed = one bed place, double bed = two bed places; convertible sofa: two bed places? [Yes / No (please specify)] - Treatment of non-permanent / extra beds: are non-convertible sofas or foldable beds included in the number of bed places? [Yes / No] - Other comments (if any): ... |

| VARIABLE | STANDARD DESCRIPTION | METADATA (TO BE FILLED OUT BY PLATFORM) |
|----------|---|--|
| occ_stay | Number of stays (number of rentals of each listing during the reference period) | <p>Please explain how you determined the number of stays:</p> <ul style="list-style-type: none"> - Count of the number of booked stays for each active listing in the reference period (where reference period is 'month')? [Yes / No (please specify)] - Booked stays that were cancelled (or postponed to another reference period) are excluded? [Yes / No] - Other comments (if any): ... |
| occ_ni | Number of nights rented out (number of nights each listing was rented out during the reference period) | <p>Please explain how you determined the number of nights:</p> <ul style="list-style-type: none"> - Count of the number of nights each property was rented out? [Yes / No (please specify)] - Other comments (if any): ... |
| occ_gni | Number of overnight stays (number of guest nights spent at each listing during the reference period) | <p>Please explain how you determined the number of overnight stays:</p> <ul style="list-style-type: none"> - Count the number of nights the property was rented out multiplied by the maximum capacity listed for that property? <i>or</i> - Count the number of nights the property was rented out multiplied by the number of people who actually stayed (estimated on the number of guests indicated at the time of booking)? <i>or</i> - Combined or other estimation methods (e.g. based on local tourist taxes per night)? [please specify] - Other comments (if any): ... |

4. METADATA FOR BREAKDOWNS AND ATTRIBUTES

| BREAKDOWN/ ATTRIBUTE | BREAKDOWN NAME | POSSIBLE CATEGORIES | Explanation Eurostat | METADATA (TO BE FILLED OUT BY PLATFORM) |
|-------------------------|---|------------------------|----------------------|--|
| TIME | The reference year and month (format YYYYMM, e.g. 201801) | | | <p>Please explain how you allocated stays / nights to a reference month?</p> <ul style="list-style-type: none"> - Stays? [month of arrival / month of departure] - Nights during one stay? [split over the different months (= recommended) / all allocated to month of arrival / all allocated to month of departure] - Other comments (if any): ... |

| BREAKDOWN/ ATTRIBUTE | BREAKDOWN NAME | POSSIBLE CATEGORIES | Explanation Eurostat | METADATA (TO BE FILLED OUT BY PLATFORM) |
|-------------------------|---|------------------------------|--|---|
| GEO_CC | Country code of the listing | ISO 2 letter code | | <p><i>Please specify information (field) used to define the country of the listing:</i></p> <ul style="list-style-type: none"> - Used geolocation transformed into country code? <i>or</i> - Used self-declared country from profile? <i>or</i> - Other (specify) |
| GEO_LIST | The LAU code where the listings are located | ~98.500 LAUs | | <p><i>Please specify information (field) used for transforming into LAUs and how the transformation took place:</i></p> <ul style="list-style-type: none"> - Allocation of listings' geolocation to LAUs? (using Eurostat API / using Eurostat public shapefiles / other (please specify)) - Known issues in matching listings with LAUs (e.g. number or percentage of non-classified listings) - Addresses declared for listings are checked against addresses declared in the owner or property manager's profile (to avoid that addresses declared are indeed those of the actual listing instead of those of the owner or property manager)? |
| GEO_GSTF | Country of origin of the guest [full version] | 50 [see list GEO_GSTF] | (Only for OCC) At country level for EU and EFTA, grouped by regions for the rest of the world, with some important partner countries reported separately) | <p><i>1. Please specify information used:</i></p> <ul style="list-style-type: none"> - Self-declared country of residence in the guests' user profile? <i>or</i> - Self-declared nationality in the guests' user profile? <i>or</i> - Information requested for finalising the booking? <i>or</i> - Country of the issuing bank of the payment card used for the booking? <i>or</i> - Other? (please specify) <p><i>2. Please specify if non-metropolitan territories are considered as part of that country or as individual countries (give examples if needed).</i></p> |
| GEO_GST2 | Country of origin of the guest [reduced version with 2 categories (resident/non-resident)] | RES NRES | | <p><i>Please report any deviations from the definition:</i></p> <p><i>GEO_GSTF = GEO_CC => RES</i></p> <p><i>GEO_GSTF <> GEO_CC => NRES</i></p> |

| BREAKDOWN/ ATTRIBUTE | BREAKDOWN NAME | POSSIBLE CATEGORIES | Explanation Eurostat | METADATA (TO BE FILLED OUT BY PLATFORM) |
|-------------------------|-----------------------|------------------------|--|--|
| ACC | Type of accommodation | 552 | Only NACE 55.2 “holiday and other short-stay accommodation” is covered in the first transmission | <p><i>Please list all type of properties included:</i> (please specify) (examples: <i>Holiday homes, Holiday parks, Chalets, Homestays Country houses, Hostels, Farm stay, Lodges, Gites, Riads, Condos, Student accommodation, Apartments, Villas, etc.)</i></p> <p><i>Types of properties not included, but possible borderline (not sure if they had to be included or not)</i> (please specify)</p> <p><i>Is the delineation based on auxiliary information about services offered to the guest (e.g. cleaning service, breakfast, reception)?</i> [Yes (please specify) / No]</p> |
| HOST | Type of host | PEER | Host renting out on an occasional basis and in a private capacity | <p><i>Please explain internal rules used to determine a ‘peer’:</i> (examples)</p> <ul style="list-style-type: none"> - <i>Single individual managing one single unit property</i> - <i>Single individual managing multiple units in one property</i> - <i>Single individual managing less than X properties (please specify X)</i> - <i>Self-declaration by the host as ‘peer’</i> - <i>Other (please specify)</i> |
| | | PROF | Host renting out as a trader or professional | <p><i>Please explain internal rules used to determine a ‘professional’:</i> (examples)</p> <ul style="list-style-type: none"> - <i>Company managing multiple properties</i> - <i>Company managing multiple properties across different countries</i> - <i>Self-declaration by the host as ‘professional’</i> - <i>Individual host managing more than X properties (please specify X)</i> - <i>Property managers are considered by default as professional</i> - <i>Other (specify)</i> |

| BREAKDOWN/ ATTRIBUTE | BREAKDOWN NAME | POSSIBLE CATEGORIES | Explanation Eurostat | METADATA (TO BE FILLED OUT BY PLATFORM) |
|-------------------------|-------------------|---|---|--|
| | | UNKN | Category temporarily added until revaluation of the issue | <p><i>Please specify why you used this category:</i> (examples)</p> <ul style="list-style-type: none"> - Cannot decide if underlying host is 'PEER' or 'PROF' in the case of property managers - Other (specify) |
| LIST | Type of listing | SHAR (shared) | Individual or shared room in a shared property | <p><i>Please specify internal rules used to determine a 'shared' property:</i> (examples)</p> <ul style="list-style-type: none"> - Self-declared information by the host that the listing is a shared property - Same entrance but different rooms (with or without private bathroom or kitchen) - Shared bathroom, shared kitchen - Other (specify) |
| | | ENTR (entire) | Entire property | <p><i>Please specify internal rules used to determine an 'entire' property:</i> (examples)</p> <ul style="list-style-type: none"> - All properties except those considered as 'shared' - The listing includes a private bathroom - Other (specify) |
| SIZE | Size class | LT10 (less than 10 bed places) GE10 (greater or equal to 10) | | <p><i>Please specify internal rules used to determine the size class:</i> (examples)</p> <ul style="list-style-type: none"> - Maximum self-declared capacity for a listing - Single bed: one place, double bed: two places, convertible sofa: two places - Other (specify) |