

Annex B: Adjustments to the Cultural and Creative Cities Monitor and year-on-year comparability

The Cultural and Creative Cities Monitor (hereinafter ‘the Monitor’) is a cross-city performance monitoring tool, compiled on a biannual basis, which continuously seeks to update and improve the way culture and creativity in cities are measured. The Monitor pays special attention to making the raw data and the final indicators and aggregated scores used accessible, providing data sources and definitions, and detailing the computation methodology (‘Annex A: The Cultural and Creative Cities Monitor methodology in ten steps’ available for download from the Cultural and Creative Cities Monitor Online¹). This annex summarises the changes made this year and provides an assessment of the impact of these changes on the comparability of rankings.

Adjustments to the Cultural and Creative Cities Monitor’s framework

Updates and improvements to the Monitor’s framework are done in a transparent manner, following the transparency principle that has inspired this work since its inception².

Table 1 provides a summary of adjustments for quick reference. A total of 10 indicators have been modified: for example, one indicator has been recomputed due to new categorisation at source, one indicator has been replaced and seven underwent methodological and/or name changes. In one case, normalised scores have been used instead of standardised scores – as was the case in 2017 – to facilitate cross-city comparability.

Table 1.
Changes to the Cultural and Creative Cities Monitor’s framework

Cultural and Creative Cities Monitor 2017	Adjustment	Cultural and Creative Cities Monitor 2017 (updated version) and 2019
1. Sights & landmarks	Recomputed due to new categorisation at source	1. Sights & landmarks
2. Museums	Name updated	2. Museums & art galleries
3. Cinema seats	Replaced with a similar indicator from a different source and name updated	3. Cinemas
4. Concerts & shows	Name updated	4. Concert & music halls
5. Theatres	Indicator changed at source	5. Theatres
13. ICT patent applications	Indicator changed at source	13. ICT patent applications
26. Passenger flights	Name and computation methodology changed at source	26. Accessibility to passenger flights
27. Potential road accessibility	Name and computation methodology changed at source	27. Accessibility by road
28. Direct trains to other cities	Name and computation methodology changed at source	28. Accessibility by train
29. Quality of governance	Normalised instead of standardised scores used	29. Quality of governance

The revision of the individual indicators is detailed below.

The data underlying Indicator 1, Sights & landmarks, have been updated due to new categorisation of culture-related facilities at the source (TripAdvisor). The 2019 categorisation identifies a higher number of sights and landmarks for nearly all the cities. However, the correlation between 2017 and 2019 data remains very high (Pearson's correlation coefficient: 0.91), showing that no substantial difference can be found between the two sets of data, which leads to no significant changes in the final normalised scores.

Indicator 2, Museums & art galleries, has changed name to better reflect what it actually measures (both museums and art galleries, although museums remain better represented in TripAdvisor).

Indicator 3, Cinemas, replaces the 2017 indicator Cinema seats due to poorer data coverage in Eurostat's Urban Audit (69% from 2011 up to 2017) compared to OpenStreetMap (98%), where the data finally included in the 2019 edition of the Monitor ultimately come from³. The high correlation coefficient between the data from the two sources suggests that the new source can be safely used, without substantial loss of information (Pearson's correlation coefficient: 0.85).

Indicator 4, Concert & music halls, has changed name to better reflect what it actually measures. Music shows were indeed featured in a previous version of TripAdvisor, while now the focus is on music facilities.

For Indicator 5, Theatres, the data source has changed due to poor data coverage in Eurostat's Urban Audit (64% from 2011 up to 2017). The new data have been retrieved from OpenStreetMap⁴ and maintain again a high correlation with the previous indicator (Pearson's correlation coefficient: 0.74).

For Indicator 13, ICT patent applications, the data source has been updated due to the lack of updated data in Eurostat's Regional Statistics (NUTS 3 level). Data collection for this variable has been interrupted since April 2019. The new total of ICT patent applications has been computed from microdata of the Organisation for Economic Co-operation and Development (OECD), available at NUTS 3 level. This maintains a very high correlation with the previous indicator (Pearson's correlation coefficient: 0.89)

Indicators 26, 27 and 28 have undergone methodological changes at source: the Directorate-General for Regional and Urban Policy has not any produced updated data on passenger flights (Indicator 26) and trains (Indicator 28) but has computed new, population-weighted indicators on accessibility by flight and by train. The underlying methodology to calculate accessibility by road (Indicator 27) has been updated as well⁵.

As regards Indicator 29, standardised scores (with a mean of zero) have been replaced with normalised scores on a 0-100 scale, as calculated by the developers of the regional Quality of Governance Index⁶. This represents a cosmetic change that allows for clearer cross-city comparability across years, but does not substantially change the index's measurement approach.

Revision of the gross domestic product, employment and population groups

The gross domestic product (GDP) and employment groups have been adjusted to allow for year-on-year comparability. In 2017, data were retrieved from the Joint Research Centre's (JRC's) LUISA platform⁷ for metropolitan regions, from Eurostat's Regional Statistics for NUTS 3 regions and from Eurostat's Urban Audit for cities. Due to the lack of updated data from LUISA at the time of data collection, data on GDP and employment rates for metropolitan

regions have been replaced with data from Eurostat's statistics on metropolitan regions, both for the 2017 and 2019 editions.

The calculation method for the group thresholds has been updated as well, passing from a percentile approach to a natural breaks approach that allows groups of similar cities to be identified in a more precise way⁸. More specifically, the latter is a data clustering method designed to determine the best arrangement of values into different classes. This is done by seeking to minimise each class's average deviation from the class mean, while maximising each class's deviation from the means of the other groups. In other words, the method seeks to reduce the variance within classes and maximise the variance between classes⁹.

Table 2 and Table 3 summarise the resulting changes for the GDP and employment rate groups due to the updated data source and percentiles. The new composition of the employment rate groups probably also reflects the recovery of jobs after the years of deep recession.

Table 2.

Cultural and Creative Cities Monitor's GDP groups for 2017 and 2019 editions

Groups	2017 (first version)			2017 (updated version)			2019		
	Threshold	Total cities per group	Cities per group (%)	Threshold	Total cities per group	Cities per group (%)	Threshold	Total cities per group	Cities per group (%)
1	> 35 000	40	24%	> 45 000	20	10%	> 45 000	20	11%
2	30 000-35 000	26	15%	35 000-45 000	37	19%	35 000-45 000	37	19%
3	25 000-30 000	38	23%	27 000-35 000	57	30%	27 000-35 000	57	30%
4	20 000-25 000	27	16%	19 000-27 000	48	25%	19 000-27 000	50	26%
5	< 20 000	37	22%	≤ 19 000	30	16%	≤ 19 000	26	14%
Total		168			190			190	

NB: percentages may not exactly sum up to 100 due to rounding.

Table 3.

Cultural and Creative Cities Monitor's employment rate groups for 2017 and 2019 editions

Groups	2017 (first version)			2017 (updated version)			2019		
	Threshold	Total cities per group	Cities per group (%)	Threshold	Total cities per group	Cities per group (%)	Threshold	Total cities per group	Cities per group (%)
1	> 74%	47	28%	> 78%	28	15%	>79%	33	17%
2	71-74%	22	13%	74-78%	34	18%	75-79%	38	20%
3	68-71%	24	14%	69-74%	47	24%	70-75%	50	26%
4	65-68%	31	18%	60-69%	59	31%	63-70%	33	17%
5	< 65%	44	26%	≤ 60%	24	13%	≤ 63%	36	19%
Total		168			190			190	

NB: percentages may not exactly sum up to 100 due to rounding.

As regards population, the groups' thresholds have remained stable as they follow the OECD–JRC definition of cities¹⁰. However, due to demographic changes, the composition of the groups has slightly changed.

Table 4.
Cultural and Creative Cities Monitor's population groups for 2017 and 2019 editions

Groups	Threshold	2017 (updated version)*		2019	
		Total cities per group	Cities per group (%)	Total cities per group	Cities per group (%)
1	> 1 million	19	10%	20	11%
2	500 000-1 000 000	39	21%	41	21%
3	250 000-500 000	40	21%	42	21%
4	100 000-250 000	60	32%	65	30%
5	50 000-100 000	21	11%	22	12%
Total		190		190	

*The new 22 selected cities have been added – otherwise, the groups' thresholds are the same as in 2017.

Sources of changes in the rankings and comparability of results

The Monitor compares the performance of local cultural ecosystems, and it also presents changes in performance and rankings over time. Importantly, all scores and rankings from 2017 have been recomputed so that results from one year to the next are **directly comparable**.

To allow for year-on-year comparability, data adjustments introduced for the 2019 edition have been applied to the 2017 edition as well.

More specifically:

- the 22 new cities have also been added to the 2017 dataset;
- data have been updated retroactively to take into account possible corrections¹¹ and updates¹²;
- data sources – both for the Monitor's indicators and for GDP and employment groups – have been updated as well, wherever updates from the same source were not available;
- indicators' names have been adjusted to be consistent with the 2019 updated names;
- the thresholds of population, GDP and/or employment groups have been recalculated for the 2017 edition, based on the updated data source and recalculated groups, as explained above;
- the imputation method for missing data has been simultaneously refined and simplified, based on national averages and the *k*-nearest neighbor (*k*-NN) method, using the average of the values of the three nearest (or statistically closest) cities having the most similar population, GDP and employment rate (see more details in 'Annex A: The Cultural and Creative Cities Monitor methodology in ten steps', available for download from the Cultural and Creative Cities Monitor Online¹³).

Making inferences about absolute or relative performance on the basis of year-on-year differences in rankings can, however, be misleading. Each ranking reflects the relative positioning of a particular city on the basis of the conceptual framework, the treatment of outliers and missing values, and the sample of cities in the given year, also reflecting changes in the underlying indicators at the source and data availability.

In addition, a few particular factors influence the year-on-year ranking of a city and complicate the time-series analysis simply based on the Monitor's scores or rankings, as follows.

- **Missing data and imputations on outliers.** It may happen that a city that is an 'outlier' in the value of a certain indicator had no data available on that particular indicator in 2017 and that such data are now available due to updates at the data source. This is for instance the case of Venice (Italy), which this year takes the maximum score (100) on Tourist overnight stays, while the score for this same indicator had to be imputed in 2017. As long as regularly updated data become available, the final scores are likely to become even more accurate, especially for those cities whose performance is difficult to estimate due to their 'exceptional' character.
- **Reference population-GDP-employment groups.** Due to demographic changes and socioeconomic recovery after the crisis, some cities have been assigned to a different population and/or GDP and/or employment group, compared to the 2017 grouping. This means that the rankings by group, and therefore the group of peer cities that a city can be compared with, may have changed. For instance, in the 2017 edition of the Monitor the city of Eindhoven (Netherlands) ranked first on the Cultural and Creative Cities Index (hereinafter 'C3 Index') in the smallest group (S-M) of cities having between 50 000 and 250 000 inhabitants. However, due to a reclassification of administrative boundaries of Dutch cities, this year Eindhoven moved to the large group (L) of cities, with inhabitants from 250 000 to 500 000, where it ranks 7th on the C3 Index, albeit maintaining a very similar score (33.19 in 2019 vs. 35.50 in 2017).
- **Reference year.** The data underlying the Monitor do not refer to a single year but to several years, depending on the latest available year for any given variable. In addition, the reference years for different variables are not the same for each city and may change due to retroactive updates. The motivation for this approach is that it widens the set of data points for cross-city comparability. At the same time, it is helpful to identify the reference years used to correctly interpret the change. On dimensions D2.1, Creative & Knowledge-based Jobs, and D2.3, New Jobs in Creative Sectors, for instance, the evolutions registered by cities like Budapest (Hungary) and Kaunas (Lithuania) happened between 2013 (i.e. year of the data included in the 2017 edition) and 2015 (i.e. year of the data included in the 2019 edition). For Kraków (Poland), Tallinn (Estonia) and Wrocław (Poland), however, the time span considered for the indicators underlying these same dimensions is much larger (2010-2016). For Vilnius (Lithuania), the time period considered is also of 6 years but with 2011 as baseline year (2011-2017) (see also 'Chapter 3 – 2019 Scores and rankings' in the main report).
- **Standardisation factor.** Most of the Monitor's variables are standardised using population size. This approach is also intended to enable cross-city comparability. However, year-on-year changes in individual variables may be driven either by the variable's numerator or, usually to a lesser extent, by its denominator.
- **Consistent data collection.** Finally, measuring year-on-year performance changes relies on the consistent collection of data over time. Changes in the definition of variables or in the data collection process could create movements in the rankings that are unrelated to true performance.

A detailed study based on the Monitor's database and city performance over time, coupled with analytical work on grounds that include local stakeholders and decision-makers, yields the best results in terms of grasping local performance over time as well as possible avenues for improvement.

Endnotes

- 1** The Cultural and Creative Cities Monitor Online: <https://composite-indicators.jrc.ec.europa.eu/cultural-creative-cities-monitor/>
- 2** See also Montalto, V., Jorge Tacao Moura, C., Langedijk, S., & Saisana, M., *The Cultural and Creative Cities Monitor: 2017 edition*. doi:10.2760/58643, <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC107331/kj0218783enn.pdf>
- 3** As the '0' values for the cities of Kalamata (EL) and Limerick (IE) from OpenStreetMap are supposed to significantly underestimate the presence of cinemas in the two cities (as confirmed by a quick web search), these have been replaced with a missing value (NA) and therefore estimated.
- 4** As the '0' value for the city of Matera is supposed to significantly underestimate the presence of theatres in the city, this has been replaced with a missing value (NA) and therefore estimated.
- 5** These indicators, available at various geographical levels (including NUTS 3 and NUTS 2 regions), assess the level of daily accessibility. They measure accessibility within an appropriate maximum travel time, suitable for daily travel (for instance for commuting purposes). A high level of accessibility can be the result of a high-performance transport network, but also of the presence of many densely populated areas close the place of departure. Hence, the absolute level of accessibility, measured in inhabitants that can be reached, does not allow the influence of both factors to be distinguished. For this reason, the population accessible within 1 hour 30 minutes is expressed as a percentage of the population that lives in a 120 km radius around the place of departure. Consequently, the resulting indicator reflects the relative accessibility levels, regardless of the population concentrations found in the neighbourhood of the place of departure.
- 6** https://ec.europa.eu/regional_policy/en/information/maps/quality_of_governance/
- 7** <https://data.jrc.ec.europa.eu/collection/luisa>
- 8** For GDP groups, the most recent data covering the period 2014-2016 for the 2019 edition and the period 2013-2015 for the 2017 edition were used, at metropolitan level. The value for Luxembourg was replaced with the second highest value to avoid the influence of this outlier in the threshold calculation. For the employment rate groups, it has not been necessary to remove any outliers. The employment rate classes have been calculated based on 2017 data for all available metropolitan regions for the 2019 edition and on 2015 data for the 2017 edition, except for Basel, for which the latest available value dates from 2010.
- 9** Jenks, G. F. (1967), 'The data model concept in statistical mapping', *International Yearbook of Cartography*, Vol. 7, pp. 186-190.
- 10** Dijkstra, L., & Poelman, H. (2012). *Cities in Europe: the new OECD-EC definition*. https://ec.europa.eu/regional_policy/sources/docgener/focus/2012_01_city.pdf
- 11** As the '0' value for the city of Liverpool is supposed to significantly underestimate the number of tourist nights per capita in the city, the value has been replaced with a missing value (NA) and therefore estimated.
- 12** In particular, Eurostat (Urban Audit and Regional Statistics datasets) and European Tertiary Education Register data relative to the 2017 edition have been re-downloaded, respecting the most recent data available principle, within the 2010-2015 time range. This means that while for some cities the data had to be imputed in the first edition, this year might instead be available or correspond to a more recent year due to updates/corrections at the source. This is mostly the case for variables coming from Eurostat's Urban Audit and the European Tertiary Education Register. The updated Urban Audit's data also reflect the change of administrative borders in French cities due to a recent reform (i.e. data for French cities have been updated retroactively by Eurostat). In addition to that, Paris's population value from Eurostat was manually corrected to represent the city's value (at the time of data collection, Eurostat still only reported the Great City value due to a lack of data updates since the administrative reform). Consequently, the indicator Tourist overnight stays per capita was also calculated manually for Paris, using the correct population size. For the recalculated 2017 edition, the population for the cities of Lund (SE), Norrköping (SE), Pula (HR) and Veszprém (HU) were added manually from Urbistat, due to missing data from Eurostat.
- 13** <https://composite-indicators.jrc.ec.europa.eu/cultural-creative-cities-monitor/>