Annex D: Spatial distribution and accessibility of cultural venues in European cities: methodological approach

In a context of high inequalities and generalised discontent, the participation in arts and cultural activities can play an important role in the achievement of broader social policy goals, such as countering social exclusion. Culture can 'bring Europeans together to experience what connects us instead of what divides us' (European Commission, *A New European Agenda for Culture*, 2018, p. 1).

In line with this vision, the second edition of the Cultural and Creative Cities Monitor (hereinafter 'the Monitor') proposes to use some of the data collected to build the Cultural and Creative Cities (C3) Index with a view to examine how urban areas operate as opportunity structures for cultural participation in the 190 citied included in the Monitor. Geographers have extensively studied how the spatial configuration of diverse assets (e.g. health facilities, parks, etc.) shape people's lives, but this approach has rarely been applied to the field of arts and culture, let alone in a multi-country context.

A data-driven approach has been adopted to identify similarities and asymmetries between the provision of cultural infrastructures and the presence and size of the population in different areas of European cities. This was done by considering both walkable distance between the population and the venues, and by examining the venues' accessibility using public transport networks.

More specifically, we compute **three main indicators**: 1) the **proportion of inhabitants that have at least one cultural venue within 2 km**, with a view to knowing to which extent the population is more or less 'exposed' to some forms of cultural offer; 2), the **mean minimum distance to cultural facilities**, for each city; thirdly, we add a new layer of analysis (i.e. the public transport network) to have a better understanding about accessibility opportunities to cultural venues based on public transport availability within the analysed European cities. More precisely, we calculate 3) the **mean number of transport stops accessible within 500 m for all the cultural venues available in each city**.

Defining the relation between the provision of cultural infrastructure and accessibility opportunities is not an easy task. Accessibility is a very much complex and multidimensional concept related to social, psychological and financial aspects, amongst others. The intention here is thus not to give an exhaustive picture of the accessibility potential of the local cultural offer. This work is rather to be intended as first attempt to 'map' the availability of some typologies of cultural venues (subject to data availability) and assess their potential accessibility by public transport (only). This first analysis could inform future research linked to, for example, social equity and inclusion, provided that more and very detailed data are made available, notably on cultural participation by socio-demographic characteristics at neighbourhood level.

The following sections present in greater detail the methodology followed to calculate the three indicators. They are structured as follows: in sub-section 1, we explain which categories of venues have been selected for the analysis and why; in sub-sections 2 and 3, we provide more details on the way geolocalised data have been gathered and cleaned for both cultural venues and population; finally, in sub-sections 4, 5, and 6 we illustrate how the three indicators were computed.

A number of challenging aspects have been identified, ranging from data availability, to the need to reduce the bias introduced by sparsely populated areas, to the definition of different steps, with the aim of building a balanced proxy of the potential accessibility of cultural venues for the local population. These are also going to be described more in detail in the following paragraphs.

1. Selected venues and rationale

Three typologies of venues have been selected for the analysis – Museums, Theatres and Cinemas – based on the availability and usability of geolocalised data.

OpenStreetMap[1] (OSM) has been selected as data source for the georeferenced points, as it offers powerful tools to gather disaggregated data, it provides a detailed and updated definition of the venue categories and it allows for open reuse and manipulation of microdata, with higher accuracy for the georeferenced elements than other websites such as TripAdvisor. Moreover, the crowdsourced nature of OSM provides open access to data[2] provided by more than 1 million users, providing geographic information on more than 10 000 points for the cities analysed in this context[3].

However, this data source is not exhaustive as regards all existing typologies of cultural venues. Therefore, it has not been possible to gather usable data for some other relevant venues such as cultural heritage sites and music halls, but also art galleries and multidisciplinary arts centres, due to poor data coverage in these categories.

A partial correspondence ultimately exists between the categories of venues covered by the Monitor dimension D1.1, Cultural Venues & Facilities, and the ones used for this work (Table 1).

Table 1.

Correspondence table between Cultural and Creative Cities Monitor indicators and OSM venues used for the spatial analysis

Included in the spatial analysis			
Monitor indicator (D1.1)	**Monitor source**	**Source used in this work**	**OSM Tag used**
Museums	TripAdvisor	OSM	tourism=museum
Theatres	OSM	OSM	amenity=theatre
Cinemas	OSM	OSM	amenity=cinema
Not included in the spatial analysis			
Monitor indicator (D 1.1)	**Monitor source**	**Reason for non-inclusion**	
Sights & landmarks	TripAdvisor	Number of points in OSM too low for some cities ([4])	
Music venues	TripAdvisor	Number of points in OSM too low ([5])	

2. Data collection and treatment

2.1. Venue data

Two different procedures have been used to extract the relevant **georeferenced data from OSM**, both relying on Overpass Turbo API[6], either through the browser-based interface or through a script in R 3.6.0 extracting only data for the relevant cities.

The output generated has then been cleaned based on the textual attributes' basic accuracy; namely, records with no information contained in the 'name' field have been removed.

2.2. Population data

The population data included in this work have been developed by the Joint Research Centre[7] , and are provided as a regular 250 metre grid[8]. The steps to prepare the population data for our analysis were the following (all done with QGIS 2.18.14):

1. creation of a regular 250 m point vector grid;
2. clipping of the vector grid on the reference areas[9];
3. overlay of the newly created point vector grid on the initial Global Human Settlement Layer (GHS-POP) grid[10];
4. sampling of the values of the raster grid in the point vector grid (using the Point Sampling Tool).

The **output of this process** is a **vector grid with the count of population in each 250 m cell**. It has full coverage of the cities being analysed.

3. Indicators computation

3.1. Percentage of population within 1 and 2 km from cultural venues

As a preliminary task to compute this indicator, buffers at 1 km and 2 km have been created, around each point in the three categories of venues, inspired by the work of Martínez & Viegas (2013, p. 90). [11] These distances are considered as maximum.

'Dissolved' buffers have been created, so as to avoid double-counting the population which is close to more than one venue during the calculation of the points in each generated buffer. In Figure 1, it is possible to see an overlap of the buffers in the left map: in these cases the population contained in each buffer would be counted multiple times. In the map on the right it can be seen that the buffers have been merged so as to avoid overlaps.

Figure 1.
An example of the outputs with undissolved buffers (left) and dissolved buffers (right) for the city of Stuttgart.

NB: purple circles represent the buffers around each cultural venue.

Then, an output shapefile has been created by calculating the number of population cells contained in each dissolved buffer, and the sum of the population contained in all the intersecting population cells. This step has been performed with the 'Point statistics for polygons' processing tool in QGIS.

The last step of data preparation was done using Excel and consisted of the calculation of:

1. the number of people who can access each type of venue within 1 km and 2 km;
2. the percentage of people having access to cultural venues in the thresholds defined above over the total population of the city.

3.2. Mean minimum distance from cultural venues

As a next step, the linear distance was calculated between each population cell and the closest cultural venue: the difference is that with this type of processing, only the distance from the first available venue is calculated, instead of the average of the distance from all the venues. This provides a snapshot of the minimum **effort** that residents need to make in order to reach the closest cultural venue.

In order to reduce the bias introduced by wide and sparsely populated administrative areas, the following two steps were performed.

- **Deletion of the grid points with no inhabitants.** This step aimed at reducing the noise introduced in the denominator of the 'mean minimum distance indicator' by grid cells with no population. Otherwise, the risk is that areas with no population negatively affect the final mean value, an issue that is particularly relevant for cities having wide and sparsely populated urban areas.
- **Deletion of grid points included in anomalous morphologies.** Some urban areas showed unrealistically high minimum distance averages. Therefore, the following manual corrections have been made:

City	Urban code 2018	Issue	Correction
Hamburg	DE002C1	Island located away from the main city boundaries (Insel Neuwerk)	Removal of the population grid points in the affected area
Nuremberg	DE014C1	Portion of the administrative area located far away from the main city boundaries	
Tallinn	EE001C1	Island located away from the main city boundaries (Aegna Saar)	
San Sebastián-Donostia	ES510C1	Portion of the administrative area located far away from the main city boundaries	
Norwich	UK566C1	River areas outside the city included in the Urban Audit administrative border	

3.3. Accessibility to cultural venues by public transport

An additional layer of analysis has been developed as a proxy for transport network connectivity. This has been considered relevant as the 'reachability' of a given space in the city cannot be measured only based on the linear **walking** distance (500 m) the population has to cover, given the fact that public transport networks exist and represent a relevant factor in the accessibility of specific areas of cities. This is particularly true for bigger cities.

Therefore, a basic accessibility measure based on the presence of public transport stations and stops near cultural venues has been computed.

This has been done using data on **bus stops**, **bus stations** and **public transport stations** ([12]) from OSM. Additional data had to be retrieved for the area of Venice, which is a unique case: Venice has no consistent road infrastructure on its islands, and it relies on water buses to provide public transport within the central part of the city and the nearby island in the Laguna. Therefore, GTFS (General Transit Feed Specification) data has been retrieved to also include the water bus stops, on top of the regular bus stops for the mainland.

The aforementioned typologies of public transport stops and stations have then been overlaid to the three categories of venues, in this case with buffers of 500 m. The output measures the number of transport services available within 500 m from each cultural venue, and provide a set of ranges of accessibility (no accessibility to high accessibility) based on the number of public transport stops available in each 500 m proximity area. The ranges are then grouped to provide a snapshot for each city, defining the percentage of venues with low/high accessibility from the transport network.

KJ-04-19-004-EN-N

Endnotes

1 OpenStreetMap is a community-developed street and data map formally operated by OpenStreetMap Foundation (https://www.openstreetmap.org/about).

2 Map data ©OpenStreetMap contributors available under the Open Database Licence (https://www.openstreetmap.org/copyright).

3 Namely, this corresponds to 1 868 cinemas, 5 519 museums and 4 075 theatres for the 190 cities analysed. All data used in this report are openly accessible via OpenStreetMap.

4 In this case the tag *historic=monuments* was tested but the number of points identified was unrealistically low for some cities, for example Brussels.

5 In this case the tag *amenity=music_venue* was tested but the number of points identified was unrealistically low (https://taginfo.openstreetmap.org/search?q=amenity%3Dmusic_venue).

6 The Overpass API is an Application Programming Interface (API) that serves to select parts of the OSM map data. It acts as a database over the web: the client sends a query to the API and gets back the dataset that corresponds to the query (source: Wikipedia).

7 European Commission, Joint Research Centre; & Columbia University, Center for International Earth Science Information Network (2015): GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a

8 The original data provided by GHS-POP is in raster format (for a technical definition of raster data please refer to: https://docs.qgis.org/2.8/en/docs/user_manual/working_with_raster/supported_data.html).

9 This step consisted of the overlay of the population grid point to the urban areas being analysed (Urban Audit 2018 areas provided by Eurostat), to ensure that no points outside the reference areas were mistakenly included in the analysis. The Eurostat Urban Audit 2018 areas have undergone a preliminary step of preparation: as the population point grid is not continuous, but instead has a sampling frequency of 250 metres, the Urban Audit areas polygons have been extended in a uniform way so as to include 360 metres more around the original shapes. This distance has been defined based on the maximum distance between each population grid point (diagonal of a 250 m-by-250 m square).

10 This step aims at creating a vector layer which contains the same data as the initial GHS-POP raster data. This is important because most of the other processing steps of this analysis are simplified when working on vector data.

11 Martínez, L. M., & Viegas J. M. (2013). A New approach to modelling distance-decay functions for accessibility assessment in transport studies, 26, 87-96.

12 This includes underground stations and train stations.

ISBN 978-92-76-08815-8
doi:10.2760/173720